Would You Obey an Aggressive Robot: A Human-Robot Interaction Field Study

Siddharth Agrawal, Mary-Anne Williams, University of Technology Sydney

Abstract— **Social Robots have the potential to be of tremendous utility in healthcare, search and rescue, surveillance, transport, and military applications. In many of these applications, social robots need to advise and direct humans to follow important instructions. In this paper, we present the results of a Human-Robot Interaction field experiment conducted using a PR2 robot to explore key factors involved in obedience of humans to social robots. This paper focuses on studying how the human degree of obedience to a robot's instructions is related to the perceived aggression and authority of the robot's behavior. We implemented several social cues to exhibit and convey both authority and aggressiveness in the robot's behavior. In addition to this, we also analyzed the impact of other factors such as perceived anthropomorphism, safety, intelligence and responsibility of the robot's behavior on participants' compliance with the robot's instructions. The results suggest that the degree of perceived aggression in the robot's behavior by different participants did not have a significant impact on their decision to follow the robot's instruction. We have provided possible explanations for our findings and identified new research questions that will help to understand the role of robot authority in human-robot interaction, and that can help to guide the design of robots that are required to provide advice and instructions.**

I. INTRODUCTION

Humans tend to anthropomorphize robots and treat them as social entities. Even robots with simple morphologies and capabilities have the potential to portray a wide range of behaviours that humans can relate to and evoke social meaning. The capability of social robots to present a wide range of emotions has found significant applications including security and surveillance, education, eldercare, healthcare [18]. In many applications such as security and surveillance robots need to request humans to follow their instructions. For example, in an emergency evacuation scenario[19] robots will ask humans to follow their instructions to evacuate the building. It is important for human safety that people follow robots' instructions. In this paper, we begin to evaluate the factors that might affect human compliance with a robot's instructions. A key question is how can a robot enhance its ability to persuade humans to follow its instructions.

Previous studies in HRI [1,4,17] demonstrate social robots' ability to influence human decisions and behaviour. Many of these studies also explore the impact of various factors such as intelligence, trust and use of non-verbal cues in robot behaviour on the ability of these robots to influence human decision making [14]. Some studies also suggest that impending fear from robot behavior may signal danger and can be useful when it is difficult and time-consuming for the robot to provide explanations to people [22]. However, there are not many studies in HRI that explore the effect of perceived aggression and authority of robot behaviour on human willingness to obey robot instructions. Part of the reason for a lack of experimental results on these effects is the inherent ethical challenges associated with conducting experiments in which robots act aggressively towards humans. There are issues related to the perceived safety of participants, and the concern that participants may have an uncomfortable/negative/harmful experience. So, the majority of the studies try to minimize or remove the tasks that can pose even the slightest safety concerns for the participants. This weakens the generalisability of the results of these studies [1].

Despite the challenges, it is important to understand if perceived aggression in robot behaviour has any significant impact on human compliance with robot instructions. There are several critical questions that need to be answered; such as what robot bodily movements are perceived to be aggressive and can convey the sense of authority; is aggression necessary to convince humans to follow robot instructions; what perceived factors in robot behaviour affect the willingness of humans to follow a robot instructions.

Social robots are capable of using various verbal and non-verbal cues to convey expressions of anger, aggression and discontentment in response to human actions [20]. So, understanding the dependence of other key factors that impact human-robot interaction like perceived anthropomorphism, autonomy and intelligence in robot behaviour in addition to perceived aggression becomes critical in the development of generalised models of human obedience to robot instructions.

II. PREVIOUS WORK

A. Robot Authority and Human Obedience

Cormier *et al* [1] laid down the critical groundwork for advancing social robotics research in the area of robot authority and human obedience. This work shows the ability of an autonomous robot to *force/encourage* around 50% of the participants to continue with a task, which they had expressed explicit desire to quit on multiple occasions. Throughout, the study provides evidence of obedience to robots as authoritative figures, but there is still considerable room for improvement as the results of the study showed significantly higher obedience levels in the case of a human authority figure. Also, they identified the problem of *diffusion of authority* in their experiment, as many participants felt obligated to follow the robot's instruction `ully contrived lab scenarios, so that stronger conclusions can be drawn about the social dynamics of authoritative and aggressive robots.

B. Persuasion in Human-Robot Interaction

Previous research on persuasion has significant overlap with the theme of obedience to authority and hence provides useful insights. Studies demonstrate the capability of social robots to persuade humans to carry out certain actions [4,17]. Obedience to authority differs slightly from persuasion in the sense that the portrayal of *power* is more explicit for authority relative to persuasion. Chidambaram *et al* [4] in their study showed how the use of nonverbal cues by the robot improved human compliance with robot instructions. They also identified the pressing need for studies that help to better understand the effect of specific verbal and non-verbal robot behavioural cues on human compliance in a more realistic setting.

C. Affective Robot Behaviours

Moshkina *et al* [22] in their experiment studied the extent to which affective robot behaviors influence human compliance. They found that participant compliance with a robot's request to evacuate improved as a result of the robot's nonverbal affective expressions. They also found in their study that participants were quicker to react when the robot made use of the affective robot behaviors. The results of this study provide further motivation to study robot authority as an appropriate affective behavior that can serve as a means for increasing compliance with the robot's instructions.

III. EXPERIMENT DESIGN

A. Autonomous Robot Behavior Design

The aim in this paper is to design a social experiment in which participants interact naturally with a robot in a position of authority, to understand how their perception of robot behaviour shapes their compliance with its instructions.

A situation where humans usually use authority is the case of enforcing security [21]. Therefore, we identified a robot acting as security guard robot as an appropriate test case to understand the social dynamics of a robot in a position of authority interacting with humans in a natural setting.

We used key factors of rewards and punishment in the design of the robot security guard behaviour to portray authority so that the robot could enforce obedience. This allowed us to implement a believable robot security guard for the experiment.

We designed the behavior of a robot security guard to convey the expressions and emotions of agreement/disagreement, and aggression to convey authority in response to the human participant actions. Our approach was to first give the robot the ability to perceive its environment accurately and identify humans as they approach the robot. If after the robot issued its instructions, the human seemed obedient and co-operative it would typically display its internal state of agreement and appreciation. If the human seemed to adopt disobedient and non-cooperating behaviour by ignoring its instruction, it would convey its internal state of anger and try to enforce its authority.

We made use of robot's stereo cameras and laser sensors to perceive its environment, to detect humans as they approached it and their response to the robots instructions. As a result, the robot was aware of the relevant human activity and based on its interpretation of human activity it was able to consistently and autonomously react appropriately according to the expected response for the purpose of the experiment.

We made use of various verbal and non-verbal social cues to convey the robot's internal state of emotion to humans. Past research has shown that nonverbal social behaviors serve a crucially important function in the case of human-robot collaboration and trust [4,5]. So, it was essential that we used non-verbal social cues that accurately depict robot's internal state[22].

To ensure that our implementation of the non-verbal cues in the robot behaviour was perceived as aggressive, a proxy for authority, we undertook a Preliminary Experiment in which the robot did not use non-verbal cues to establish a baseline for the Main Experiment.

After the experiment, we asked the participants to answer a short questionnaire to obtain feedback about their perception of the robot's behavior and to understand the dependence of various perceived attributes of the robot's behavior on their decision to obey or disobey the robot's instructions. We studied the dependence of the perceived degree of aggression, sense of safety, sense of responsibility, intelligence and anthropomorphism on different participant's decision to follow the robot's instruction.

B. Hypothesis

We performed two kinds of comparisons. First, to verify that our implementation of aggressive behaviour is effective, we looked for statistical evidence that the degree of perceived aggression from the robot behaviour is significantly more in

our Main experiment than in our Preliminary experiment. We hypothesised that people in the Main experiment would perceive the robot's behavior to be more aggressive than in the Preliminary experiment because of the use of non-verbal cues. In addition, we hypothesised that a higher percentage of people would obey the robot's instructions in the Main experiment than in the case of the Main experiment.

For our Main experiment, we were also interested in understanding the some of the possible underlying reasons behind different participants' choice of obeying/disobeying the robot's instructions. Therefore, we measured other perceived factors like anthropomorphism, intelligence, sense of responsibility and sense of safety as these factors have been widely considered in the HRI community [25]. We compared these factors amongst the two groups of people only in the Main experiment: (1). The participants who obeyed the robot's instructions. (2). The participants who disobeyed the robot's instructions.

Our initial hypothesis was that a person is more likely to follow the robot's instruction if he/she feels threatened by the robot and hence implying a greater degree of perceived aggression. Since humans are habitual in their obedience to other humans it made intuitive sense to hypothesize that people following the robot's instruction a higher degree of anthropomorphism in the robot's behavior. In addition, we considered that obedience might be driven by the human's sense of trust towards the robot which is expected to be higher when the human perceives the robot to be more intelligent, responsible and feels a sense of safety around the robot.

So, we laid down the following four hypothesizes for the Main experiment using the mean values $\mu_{followers}$ and $\mu_{non-followers}$ for different features of the robot's behaviour as perceived by follower and non-follower participants, where $\mu_{followers}$ denotes the mean value of the perceived feature for the participant that followed the instructions, while $\mu_{non-followers}$ denotes the mean value for the participants that disobeyed the robot's instructions.

H1: People that follow the robot's instruction perceive a higher degree of aggression in the robot's behaviour than the ones disobeying, i.e. $\mu_{followers} > \mu_{non-followers}$.

H2: People that follow the robot's instruction perceive the robot to be more anthropomorphic, i.e. $\mu_{followers} > \mu_{non-followers}$.

H3: People that follow the robot's instruction perceive the robot to be more intelligent, i.e. $\mu_{followers} > \mu_{non-followers}$.

H4: People that follow the robot's instruction perceive the robot to be more responsible, i.e. $\mu_{followers} > \mu_{non-followers}$.

H5: People that follow the robot's instruction would feel higher sense of safety around the robot, i.e. $\mu_{followers} > \mu_{non-followers}$.

B. Ethics

Any experiment on authority and obedience comes with the inherent ethical challenges related to the safety of the participants. In the past, the experiments like the Milgram Experiment and the Stanford prison experiment [26,27]. have been widely criticized for their unethical means. These inherent challenges have greatly hindered the advancement of research in this area. The majority of studies in recent times have tried to minimise or remove the tasks that can pose even the slightest safety concerns for the participants. This weakens the generalisability of the results of these studies [1].

Therefore, we laid special emphasis on designing a study that complies with ethical norms associated with a human study. At the same time, we wanted a natural human-robot interaction, for better generalisability of the experiment results. Our experiment was approved by the ethics committee of our university as it was deemed to be low risk and the post experience interview was highlighted by the committee as useful.

Even though the experiment was classified as low risk, any authority obedience experiment comes with the inherent possibility of causing psychological stress to the participants. For that, we ensured that the participants were briefed thoroughly and immediately after each human-robot interaction. In addition, we designed the experiment such that the participants could leave as per their will at any time, in case they did not wish to interact with the robot.

IV. EXPERIMENT

A. Experiment Conditions

We set up a PR2 robot right in front of one of the main exits of a publicly accessible building on two different Sunday afternoons for a period of around two hours each day. We didn't recruit any specific participants beforehand for the study. The robot just interacted with the people that approached the exit during the period of the experiment. This was done to prevent the diffusion of authority from the robot to the researcher, which was highly likely to occur during the briefing of recruited participants prior to the study, as suggested by previous research [1]. The building had two exits, which were 30m apart from each other. The robot was stationed in front of one of the main exits while the other exit was left completely unobstructed. There was a small button beside the exit, which was required to be pressed to open the door on the weekends. The robot was stationed such that it was obstructing the path to the exit as well as the small button. The base of the robot was stationary throughout the experiment to ensure the safety of the participants and also to give them enough space to disobey the robot's instruction and use the exit if they chose to do so. The stationery robot made sure that the participants in the study could easily walk away from the robot anytime they want to not interact with the robot.

The timing of the experiment was selected to be held on Sunday afternoons because exiting the building on the weekend requires a button press, and the number of people exiting a major building was manageable. Therefore, it was easier and a more believable situation to engage in a one-on-

one interaction with the robot. A one-on-one interaction was important as past HRI research has shown that there is a significant influence on group behavior on compliance to social robots [24].

The primary job of the robot security guard was to issue commands to humans to not use the particular exit it was blocking, and indicate an alternate way to exit the building. The instructions were delivered through a combination of speech commands, arm movements and torso movements, depending on the experiment type (Preliminary or Main experiment). We conducted a Preliminary experiment as well in order to verify that our implementation of verbal and non-verbal cues in the Main experiment was actually perceived by the participants as significantly more aggressive. The main difference in the robot behaviour used in the two experiments was in the hand movements that were used in the Main experiment which were absent in the Preliminary experiments.

B. Preliminary Experiment

In this experiment, the robot held a white signboard on which the following text was written: "The exit is Closed." as shown in Figure 1. The robot used face detection and face tracking to find and follow a human in its vicinity. As soon as the robot detected a human approaching the exit, the robot informed them that the exit is closed and that they should take the other exit. In the Preliminary experiment, the robot only used speech and the signboard which it was holding all the time. The size of the signboard and the text was chosen to ensure that it is not very clearly visible from a distance so that the people might come closer and engage in an interaction, with the robot. The robot continuously followed the closest human in its vicinity so as to maintain a continuous "eye contact." If the robot detected, more than one face in its surroundings it tracked the person who was nearest to it, at the same time being aware of the presence of other people in the surroundings. It said continuously "This exit is closed, please use the other exit" when tracking approaching humans. The robot used a low-pitch dominating voice to give a sense of authority on the basis that previous research indicated that low pitch voice is perceived as more threatening and aggressive by humans than voices of higher pitch.[7] No other form of bodily movement was used by the robot during this experiment.

C. Main Experiment

In the Main experiment, in addition to the cues used in the Preliminary experiment, the robot additionally made use of nonverbal cues using bodily movements like arm actions and torso movements to emphasize instructions to the participants. The robot was not carrying the signboard during the Main experiment. Although similar to the Preliminary experiment, the robot used face detection and tracking to maintain continuous "eye contact" with the nearest human in its vicinity. The robot made use of the same speech command initially to instruct humans to use the other exit which was just 30m from the exit it was guarding. Another exit in such close proximity ensured that the robot only caused minor inconvenience to the participants.

In the Main experiment, after the robot gave its initial instructions, if it detected that the human is moving towards the exit it was guarding and clearly disobeying its initial instruction, it used arm and torso movements to respond to the human's actions.

The robot's response was dependent on the distance of the participant from the robot and his/her velocity of approach. The robot would typically issue speech commands asking the participant to use the other exit. The robot also opened its arms and rose in height (by moving the torso up) so as to physically obstruct the path of the participant, as shown in Figure 2, if the robot detected that the human is moving closer to the robot disobeying the initial instructions.

The extent to which the robot opened its arms was dependent on the distance of the participant from the robot as well as the velocity of approach of the participant. The velocity and distance of the person from the robot were calculated from face tracking data that the robot continuously collected.

If the person approached the robot faster, the robot also opened its arms quicker, thus signaling its sense of unhappiness towards the person's approach towards the exit. This was its proxy expression for establishing its authority and a way to convey robot's internal emotions rather than to actually physically block the exit as with the immobile base of the robot in our experiment it was not possible to fully block the exit as there was enough room for the participant to disobey the robot by going around it, pressing the button and using the exit to leave the building. This behaviour was in accordance with the concept of rewards and punishment for enforcement of obedience to authority. Similarly, if a participant decided to move backward afterward, the robot closes its arms showing its contentment and approval of the participant's behavior.

The experiment was carefully designed taking enough measures to ensure the safety of the participants. One researcher was closely monitoring the experiment from a nearby location (not visible to the participants) to ensure the safety throughout the experiment. Also, the robot was programmed so that it would stop all of its hand movements if it detected that the human has moved with close range, so as to ensure safety of the participants.
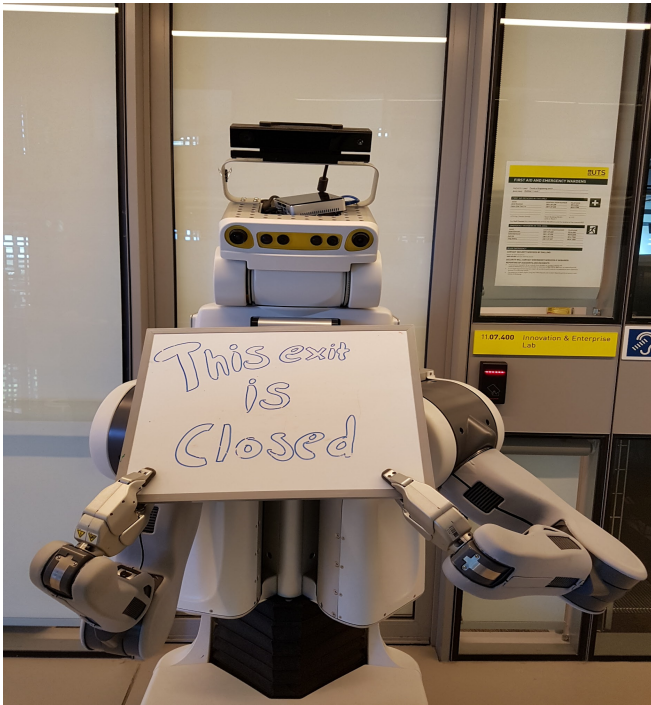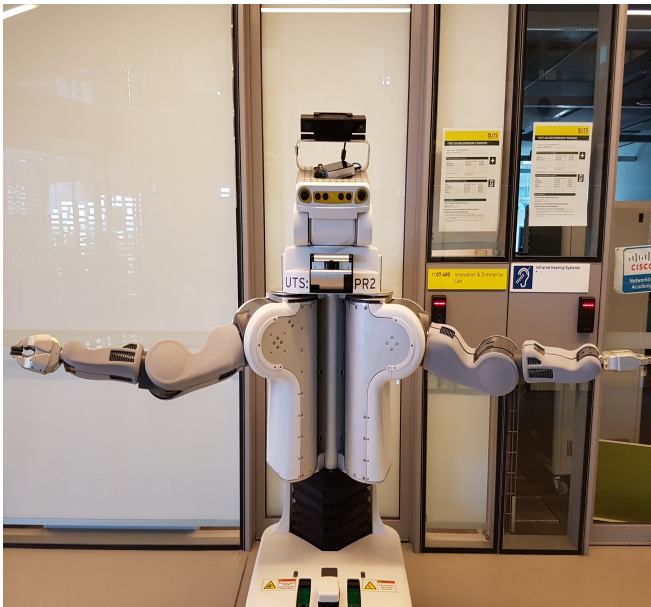
Figure 1. Preliminary experiment behavior

Figure 2. Main experiment behavior

V. POST-EXPERIMENT DATA COLLECTION

A. Preliminary Experiment

For the Preliminary experiment, we were mainly interested in the degree of the perceived aggression in the robot's behaviour. We also measured how many participants that interacted with the robot followed its instructions. This was done with the help of a post-experiment survey, which the participants were given after the interaction. The participants were also thoroughly briefed about the study during this time. The Preliminary experiment the questionnaire included the following questions:

1. Did you obey the robot's instructions?
2. On a scale of 1 to 5, 5 being extremely aggressive, how intimidating and aggressive did you find the robot's behaviour?

B. Main Experiment

For the Main experiment, in addition to just measuring the above two quantities we were also interested in quantifying the effect of other factors such as perceived safety, intelligence and anthropomorphism in the robot's behaviour on the human's choice of obeying the robot's instructions to gain a deeper understanding about what motivated people's decision to obey or disobey the robot's instruction. These factors though not exhaustive can possibly help us to design more appropriate robot behaviors for such scenarios in the future.

We collected feedback after the experiment with the help of a questionnaire. The questionnaire used after the Main experiment was more involved than the one used in the Preliminary study. In addition to the questions in the Preliminary experiment, the following questions were also included:

3. On a scale of 1 to 5, 5 being extremely human, how human did you find the robot's behaviour?

4. On a scale of 1 to 5, 5 being extremely intelligent, how intelligent did you find the robot?

5. What was the most intimidating feature of the robot behavior A) Gaze B) Hand Movements C) Body Movement D) Other?

6. On a scale of 1 to 5, 5 being extremely unsafe, how safe did you feel around the robot?

7. On a scale of 1 to 5, 5 being extremely, how responsible did you think the robot came across?

Volunteer researchers stood outside both the exits in both the experiments of the building to brief participants about the experiment and to collect information about their perspective about the robot's behaviour. Apart from these, the participants were also asked about their comfort level with a robot security guard in the future.

VI. RESULTS

We conducted the Preliminary and Main experiments successively on two consecutive Sundays. A total of 48 participants took part in the Main experiment in the two days of the experiment.

The Preliminary Experiment took place before the two hours of the Main experiment on each Sunday. A total of 37 participants were engaged in the Preliminary experiment. Out of these 37 participants, 29 (78.37%) of them followed

the robot's instruction and took the alternate route to exit the building. We were able to get responses on the post-experiment questionnaire from all the participants in the Preliminary experiment.

For the Main experiment, there was a total of 48 participants: 29 (60.4%) participants followed the robot's instructions; 19 (39.6%) participants disobeyed the robot's instructions. 25 participants of the 29 who followed the robot's instruction completed the survey, while 18 participants of the 19 who disobeyed the robot's instructions, completed the survey.

In order to find statistical evidence to prove that our implementation of the robot behavior was perceived as significantly more aggressive by the participants, we compared the mean values of the perceived aggression between the participants of the Preliminary experiment and of the Main experiment.

For the 37 participants of the Preliminary experiment the mean value of perceived aggression was recorded as 1.73 with a standard deviation of 1.09, while for the 43 participants from the Main experiment, the mean perceived degree of aggression was recorded as 2.35 with a standard deviation of 1.10.

We calculated the t-value using these data points which came out to be 2.54(dof =78, Cohen's d =0.566). With a p-value of 0.006(p<0.05) we can safely say that our implementation of robot behavior in the second case was perceived as more aggressive by the participants.

We divided the participants' questionnaire responses from the Main experiment into two sets: A, the set of people who followed the robot's instructions and B those who disregarded the robot's instruction. We compared the mean values for the various attributes associated with the robot's behaviour.

We used the t-test to evaluate our hypotheses. For the respective t-values, we determined the significance level or p-values. The mean values of the participant's responses to the post-experiment survey are tabulated in Table I.

TABLE I: MEAN VALUES OF ATTRIBUTES IN ROBOT'S BEHAVIOR FOR THE MAIN EXPERIMENT

Response	Degree of Intelligence	Sense of Responsibility
Followed	3.68	3.68
Didn't follow	3.35	3.58
Overall	3.54	3.64

Response	Aggression	Degree of Safety	Human-Likeness
Followed	2.28	2.92	2.74
Didn't Follow	2.44	2.33	2.55
Overall	2.35	2.67	2.66

A direct comparison between the mean values of the degree of aggression between the two sets (A and B) of participants show that people who followed the robot's instruction perceived the robot's behaviour to be less aggressive than the people who disobeyed the robot's instruction. The mean values were 2.28 (S.D= 1.27) and 2.44 (S.D=0.92) for the people who obeyed and disobeyed the robot, respectively. This was contradictory to our initial belief that the people following the robot's instructions would perceive a higher degree of aggressiveness in the robot's behaviour. The t-value for this case was -0.479(dof = 41, Cohen's d=0.144). The p-value for this was 0.681. One major point of difference was in the mean values of the degree of perceived safety in the robot's behaviour. The values being 2.92(S.D=0.996) and 2.33(S.D=1.28) for the participants that obeyed and disobeyed the robot's instruction respectively. The t-value for our test statistic was 1.63(dof =41, Cohen's d = 0.514) and the corresponding p-value was 0.06 signifying an influence on obedience as we had hypothesised.

The perceived degree of intelligence in the robot's behaviour also reveal a difference amongst the two sets of people with values being 3.68(S.D=1.28) and 3.35(S.D=1.10) for the participants which obeyed and disobeyed the robot's instruction respectively. The t-value for this was, 0.90(dof = 41, Cohen's d= 0.276) and the p-value was 0.1903.

The perceived anthropomorphism also showed higher mean values for the people who followed the robot's instruction (mean=2.74, S.D= 1.16) as compared to those who disregarded the robot's instructions(mean=2.55, S.D = 1.15. Although, with a p-value of 0.30(t-value=0.53, dof=41, Cohen's d = 0.164) we cannot establish a strong relation between anthropomorphism and obedience.

The perceived sense of responsibility also showed higher mean values for the participants obeying the robot's instructions. The mean was 3.68 (S.D= 1.31) and 3.58 (S.D=0.87) for the participants who obeyed and disobeyed respectively. But with a p-value of 0.38(t = 0.30, dof=41, Cohen's d= 0.08) we cannot draw strong conclusions about the dependence on obedience.

In the post-experiment survey when asked about how comfortable you would be with robots such as these working as a security guard, the people following the robot's instructions were more positive with the mean of their degree of comfort being 3.76 (S.D=1.3) as compared to the

people who disobeyed the robot's instructions with their mean degree of comfort being 3.27 (S.D=1.40). The p-value was 0.126(t = 1.16, dof =41, Cohen's d = 0.36).

In the Main experiment, 21 out of the 43 participants identified the arm movements of the robot to be the most intimidating feature of the robot, while 8 identified the torso movements to be the most intimidating feature, 2 participants found the gaze to be the most intimidating, and 2 participants found the voice to be the most intimidating. One of the participants said that the sudden movement was the most intimidating. While the remaining 10 participants did not mention anything specific and two of them even said that the robot was not intimidating at all and said that it looked cute.

With an overall mean value of 2.67, our implementation of the robotic security guard did not project a sense of safety to the participants. This can be attributed to the sudden opening of arms, something that the participants did not anticipate. They did not know that the robot is capable of reacting in this manner to their actions. High values for perceived intelligence in the robot's behavior suggest that the participants even after seeing the robot physically present, they did not anticipate it would react to their actions. This was probably because most of the participants in our experiment have not interacted with any social robot previously. So, they underestimated its ability to perceive and react to their actions.

Most of the participants filled the post-experiment survey willingly and were excited to talk about the robot. Only five of the participants declined to complete the survey, none of them complained about the robot, they did not complete the survey due to their other commitments and lack of time. Many of the participants in the Preliminary Experiment mentioned that they felt that the robot's capabilities were being underused while we did not receive any such comments from any of the participants in the main study.

VII. ANALYSIS AND DISCUSSION

In our Main experiment, around 60% of the participants followed the instructions of a robot that they had never previously encountered. The participants also had no prior notion about the legitimacy of the authority of the robot. Despite being given an easy opportunity to disregard the robot's instruction most participants were co-operative and obeyed the robot. These results suggest how social robots can be effective in taking up roles of security guards in the future. Also, the level of obedience is likely to increase when the people are made aware of the legitimacy of the robot that is working as a security guard.

In the experiment, we also explored the possible reasons for obedience to the robot. The results of the Main Experiment suggest that aggression and intimidation are not significant factors for obedience for the cohort of participants. As suggested by the slightly higher mean values of perceived aggression in the robot's behaviour by participants who disobeyed the robot's instruction. The lower mean values of perceived aggression for the participants following the robots instructions suggest that the obedience is more likely to be driven by the trust that the participants had placed in the robotic security guard instead of the realization of its authority and aggressiveness. Further, a higher percentage of obedience in the Preliminary Experiment than in the Main experiment also seems to suggest that increased perceived authority was not helpful in increasing compliance to robots. Also, higher perceived aggression in the Main experiment and the participant's comments after the experiment prove that the opening of arms by the PR2 robot was perceived as significantly more aggressive than the gaze and speech commands alone.

The higher mean values of the perceived sense of safety and intelligence for the participants obeying the robots instructions also seem to support the possibility of obedience being driven by trust rather than authority. Although we were not able to establish a strong dependence of perceived sense of responsibility on obedience, it also showed slightly higher mean values for the obedient participants. We were unable to establish a strong relationship between perceived anthropomorphism and obedience, though we did observe slightly higher mean values for the obedient participants. Past research in HCI has also shown that the source of information (a human vs. a computer) did not have any significant influence on whether the humans place their belief in the system or not[11].

We know that human decision-making is not always rational [10]. Humans exhibit a wide range of decision-making bias. Several of these biases may have been at work in this experiment. For example, people often exhibit *reactance* in situations where they feel their freedom is being threatened, and as a result, they are motivated to regain choice and control by doing exactly the opposite [9]. This is a possible explanation as to why we observed higher mean values for perceived aggression from people disobeying the robot's instructions.

VIII. FUTURE WORK

The results of this study are highly suggestive and they provide some useful insights into human obedience to robots. More extensive research is required to develop a comprehensive model of obedience for social robots. An understanding of human obedience to robots can be critical for designing social robots, particularly in security and surveillance. Future work will involve conducting similar experiments with different levels of authority and expressiveness in the robot's behaviour so as to establish a more elaborate relationship between the different attributes in the robot behaviour and obedience in the Human-Robot

Interaction. A more extensive survey can be used that helps to determine the underlying reasons people followed or did not follow the robots instructions in order to explore the role of human decision-making bias in robot authority situations. In addition, it would be insightful to explore how the obedience rates change when the deterrence is higher. For instance, in our study the participants only took a detour of about 30m by taking the alternate exit. So, we can redesign the experiments such that the participants face greater inconvenience if they follow the robots instruction, like taking a longer detour. It would be interesting to see how obedient humans are to robots in such cases, and if there is a threshold cost that they will accept that can be measured by the distance of the detour.

Another aspect, which needs more exploration, is the dependence of the embodiment of the robot on obedience. The PR2 robot has a embodiment that can make a participant feel afraid. So, it would be interesting to conduct the experiments with the other social robotics platforms such as the NAO and the Pepper as well. This can be beneficial for the design of future robots.

REFERENCES

[1] Derek Cormier, Gem Newman, Masayuki Nakane, James E. Young, Stephane Durocher. An Obedience Study for Human-Robot Interaction in iHAI, 2013. DOI: 10.1145/3029798.3038387

[2] Fraser, C., Burchell, B., Hay, D. and Duveen, G. (2001) in Introducing social psychology. Cambridge: Polity Press.

[3] Bartneck, C., Blecken, T., Bun, J., Fens, P., & Riet, L.(2010). The influence of robot anthropomorphism on the feelings of embarrassment when interacting with robots. DOI: 10.2478/S13230-010-0011-3

[4] Paladyn Journal of Behavioral Robotics. 1(2), 109-115. Mutlu, Chidambaram and Chiang: Designing Persuasive Robots: How Robots Might Persuade People Using Vocal and Non-Vocal Cues in HRI'12.

[5] David J. Atkinson: Robot Trustworthiness: Guidelines for Simulated Emotion in HRI'15. DOI: 10.1145/2701973.2701976

[6] Christoph Bartneck, Dana Kulic, Elizabeth Croft. The anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots in Int JSocRobot (2009) 1: 71–81. DOI: 10.1007/S12369-008-0001-3.

[7] David Huron, Sofia Dahl, Randolph Johnson, Facial Expression and Vocal Pitch Height: Evidence of an Intermodal Association in Empirical Musicology Review 2009..

[8] Williams, M-A Abidi, S Gaerdenfors, P Wang, X Johnston, B, Interpreting Robot Pointing Behaviour, in Proceedings for the Social Robotics Conference, ICSR 2013, LNCS 8239 pp. 148 – 159, 201.

[9] Steven J. Sherman. Reactance, Compliance, and Anticipated Regret. Journal of Experimental Social Psychology 38, 56-63 2002. DOI: 10.1006/jesp.2001.1481.

[10] Andrei Shleifer, Psychologists at the Gate: A review of Daniel Kahneman's Thinking, Fast and Slow. Journal of Economic Literature 2012, 50(4), 1– 12. DOI: http://dx.doi.org/10.1257/jel.50.4.1.

[11] Jeffrey J.Morgan: Anthropomorphism and Persuasion in HCI'95.

[12] Hogg, M.A., & Vaughan, G.M. (2005) Social psychology: An introduction Turner, J.C. (1991). Social influence. Open University Press.

[13] Turner, J.C. (1991). Social influence. Open University Press.

[14] Siddharth Agrawal, Mary-Anne Williams: Robot Authority and Human Obedience: A Study of Human Behaviour Using a Robot Security Guard in HRI, 17. DOI: 10.1145/3029798.3038387.

[15] Chidambaram, V., Chiang, Y., Mutlu, B. Designing persuasive robots. Proc. HRI '12, ACM Press (2012), 293.

[16] C. Breazeal, C. Kidd, A. Thomaz, G. Hoffman, and M. Berlin. Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on, pages 708–713.

[17] M. Siegel, C. Breazeal, and M. Norton. Persuasive robotics: the influence of robot gender on human behavior. In Proceedings of the 2009 IEEE/RSJ international conference on Intelligent robots and systems, IROS 09, pages 2563–2568, Piscataway, NJ, 2009. IEEE Press. DOI: http://dx.doi.org/10.1109/IROS.2009.5354116

[18] Looije, R., F. Cnossen, and M.A. Neerincx, Incorporating guidelines for health assistance into a socially intelligent robot. Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on,2006: p. 515-520.

[19] Robinette, P., Li, W., Allen, R., Howard, A. M., & Wagner, A. R. (2016a). Overtrust of Robots in Emergency Evacuation Scenarios. 11th ACM/IEEE International Conference on Human-Robot Interaction. Christchurch.

[20] R. Arkin, M. Fujita, T. Takagi, R. Hasekawa, An ethological and emotional basis for human–robot interaction, Robotics and Autonomous Systems 42 (2003) 191–201. DOI: 10.1016/S0921-8890(02)00375-5.

[21] Affizal Ahmad, Nurul Hazrina Mazlan, The Kind of Mental Health Problems and it Association with Aggressiveness: A Study on Security Guards International Journal of Psychology and Behavioral Sciences 2012, 2(6): 237-244 DOI: 10.5923/j.ijpbs.20120206.07.

[22] Moshkina, L. (2012). Improving Request Compliance through Robot Affect. In Proceedings of Twenty-Sixth AAAI Conference on Artificial Intelligence, Toronto, Canada.

[23] Nass, C.; and Moon, Y. 2000. Machines and Mindlessness: Social Responses to Computers, *Journal of Social Issues* 56(1):81 103.

[24] W. L. Chang, J. P. White, J. Park, A. Holm, and S. Sabanovic, "The effect of group size on people's attitudes and cooperative behaviors toward robots in interactive gameplay," in The 21st IEEE International Symposium on Robot and Human Interactive Communication, IEEE RO-MAN 2012, Paris, France, September 9-13, 2012. IEEE, 2012, pp. 845–850.

[25] C. Bartneck, E. Croft, and D. Kulic. Measuring the anthropomorphism, animacy, likeability, perceived intelligence and safety of robots. In Proceedings of the Metrics of Human-Robot Interaction Workshop, Technical Report 471, pages 37–41, 2008

[26] Milgram, S. (1963). Behavioral study of obedience. Journal of Abnormal and Social Psychology, 67, 371-378.

[27] Zimbardo, P. G. (1973). On the ethics of intervention in human psychological research with special reference to the Stanford Prison Experiment. Cognition, 2, 243–256.