BO THE BOX: AN EXPERIMENTAL SOCIAL GAMING ROBOT

Panteleimon Manouselis, p.manouselis@student.utwente.nl, s3084493, M-ROB

***Abstract* — This report presents both the fundamental theory needed to design a multimodal social robot and an experimental design for a robot that can play a rock/paper/scissors game. The key elements that need to be considered are the motivation for the application as well as the behavioral goals of the robot, namely, its purpose, communication modes and the emotions one wants to elicit from the human participants. A qualitative study with three participants was performed in order to evaluate our design. Lastly, future improvements of the robot are proposed regarding its software, hardware, interaction and inclusivity.**

***Keywords* — Social robots, speech recognition, human robot interaction/communication, design in human terms, multimodal communication.**

I. INTRODUCTION

The final challenge of the course Human-Robot Communication was the development of a robot that could play the popular "*rock, papers, scissors*" game with any random person while simultaneously conveying the various game states and emotions using both verbal and non-verbal communication. Implementing this interaction proved to be a difficult task for four main reasons. Firstly, we had limited time to develop the software and create the robot's exterior. Secondly, the provided hardware had significant physical limitations (e.g., the servo motor could not rotate more than 180°). Thirdly, each person who interacted with the robot asked similar questions by using different terminology, syntax and accents. Thus, accounting for different speech acts as well as redundancy had to be considered during the design and development of our robot. Finally, and most importantly, the robot would have to communicate its "*intentions and states*" [1] at multiple stages of the game (e.g., end of the game, "emotion" based on outcome of game).

According to *Bonarini*, "*in social robots the quality of communication between robots and humans should always be evaluated in order to support the effectiveness of the robot to perform tasks where communication is needed*" [2]. Therefore, after developing our robot, we proceeded to evaluate our design by interviewing three random people who interacted with it.

II. BEHAVIOURAL DESIGN & MOTIVATION

In the final project, our robot could communicate using all available modalities. Thus, the *physical design*, the possible *conversation flows*, and the *behavior* of the machine were examined in detail prior to development. As research posits, "*an agent's appearance should be consistent with the agent's behavior and also with the task*" [3]. We therefore created a new cuboidal and cartoonish physical design, which has asymmetrical dimensions (i.e., large head to body ratio compared with a human). Moreover, we laser-cut part of the robot's chest and added a servo-powered arrow, whose purpose was to perform the game's countdown and indicate the robot's choice in a timely manner. The rationale behind those decisions was that the appearance of the robot should be consistent with both the task (playing a game) and the behavior (funny, sassy and cartoonish). Accordingly, we investigated the effect of our robot's appearance in interaction with the other variables. The results of this design are discussed in section III.

"*Typically, the primary assumption in social robotics is that robots should adhere as much as possible to the modalities and ways humans communicate*" [1]. Therefore, initially, we tried to design a robot that utilizes all available modalities for "Human-Like" communication. This requires two main actions from the designer. Firstly, creating the means to deal with natural language generation, dialogue management, speech and language recognition [4]. Secondly, utilizing nonverbal communication channels to mimic the ones of humans, namely, gaze, eye movement, gestures, posture and bodily movements [4]. Our initial inclination to design using a "*Human-Like*" approach was based on recent research which theorizes that "*Human-Like communicative behavior have positive effects on relational outcomes*" [5].

Although this approach is widely adapted in HRI, for a variety of reasons we instead chose to design our robot in "*Human-Terms*". Firstly, the central idea behind our robot is the chest movement that indicates both the in-game choice of the robot and the countdown to the game's start. Therefore, a necessary condition for a successful interaction with our robot is that the robot's movement is perceived as intentional by the human participants [4]. Secondly, our robot was not humanoid and consequently a design in "*Human-Terms*" was favored. Additionally, and most importantly due to time constraints and hardware limitations, mimicking human communication channels like posture, gestures and eye movement would have been impossible. Consequently, instead of designing in "*Human-Like"* ways, we chose to design in "*Human-Terms"* [1].

Both Malle Folk's Theory of Action that "*humans interpret and explain agents' actions by providing reasons for their behaviors*" and the theory that humans "*tend to animate and parse the intentions of a robot*" [1] were integrated in our robot's movements and expressions. An example of this integration is the game's countdown, where the three consecutive rotations of the robot's chest arrow were interpreted by the human as a countdown (i.e., all three human participants considered the arrow as part of the robot and parsed the intention of the robot "to begin the game"). Another example of people "*interpreting and explaining the actions of robots by providing reasons for their behaviors*" is through the changes of colors and shapes of the robot's "eyes" (the LED lights). All three participants interpreted the changing color and shape of the robot's eyes as communicating 'feelings' and concluded that these 'feelings' were based on whether the robot won, lost, or tied.

The three pillars of non-verbal action-based communication, which are described by *Tomasello* [6] (cited in *Zaga* [4]), were considered when developing the non-verbal communication modalities of our robot (i.e., the "eyes" and the rotational chest). Firstly, the communication should take place in the joint attentional space to be perceived – our robot could recognize whether a human face was directly in front of it and would only interact with participants when a face has been detected in the past few seconds. Secondly, actions should clarify who is the recipient of the communication and the object of the communication – in our case the human understood through the robot's actions and interaction that communication was bidirectional and at most cases simultaneous (as one can see from our evaluation results). Lastly, the action itself (namely the different states and movements of the robot during the rock, paper, scissors game) embodied the goal of communication.

According to Kahn et al. [7] design patterns can benefit applications of human-robot communication. Therefore, based on Kahn's work three of them were created prior and during the development of the robot, to implement a socially compelling human-robot interaction.

Initial Introduction: An essential aspect of social interaction involves an initial introduction. This Design Pattern uses a scripted and conventionally established verbal and behavioral repertoire (a) to recognize the other, and (b) to inquire politely about the other. Instantiations of this Design Pattern in the robot: If the participant appears for the first time in front of the robot or if the participant has said some form of greeting (*"hi", "hello", etc.*) then the robot proceeds to greet the participant (*"hello there"*). If the participant then asks for information regarding Bo's capabilities *("What can you do"* or similar), it will provide a short presentation of himself. During the Initial Introduction, the human participant could use a variety of different terminology or syntax. Thus, we used Google's Dialogflow to take into account different speech acts as well as redundancy.

Simultaneous Engagement in Game Context: A lot of social games, no matter their variation, involve choosing your action without knowledge of the actions chosen by other players (simultaneous or static games). Simultaneous Engagement in Game Context is a design pattern for sociality that may easily set into motion claims of unfairness (pattern described below). Instantiations of this Design Pattern: After the countdown is over the human participant should shout its choice whilst the robot is turning the arrow to show its choice. Both the human participant's selection and the robot's selection should occur in the same time frame as otherwise the player that was late to make a move can observe the opponent's choice and counter it with the winning move. In this case a claim of unfairness pattern is set into motion.

Claiming Unfair Treatment: In moral psychology, a central mode has been used experimentally to set up claims about immoral treatment, based on deontological justifications of fairness, justice, and rights [7]. Claiming Unfair Treatment or Wrongful Harms is a design pattern that allows one to make claim to its moral standing. Instantiations of this Design Pattern: During a game and after the countdown is complete the robot shows its choice (e.g., "*Rock*") whereas the player does not say anything or says something irrelevant (e.g., "*I choose banana*"). Thereafter, the robot shows through its eyes that it got mad and says ("*Omg, like, are you even listening to me?*").

III. RESULTS OF THE EVALUATION AND DISCUSSION

In order to evaluate our robot, the three random people who interacted with it completed a survey immediately after the interaction and 18 open-ended questions were answered. To account for the observer effect (i.e., the interaction between the participants and the robot is different if the interviewer is not present) the participants were left alone in a room with the robot during the interaction and the completion of the survey. Both the questions and the interpretation of those questions were based on two evaluation methods presented in the article "Evaluating Human-Robot Interaction" [3]: *Task Completion and Efficiency* and *Emotion and Active Computing*.

Task Completion and Efficiency: Task Completion and Efficiency were evaluated through responses to questions *D, E, L, M* and *N*. Overall, 2 out of 3 participants felt that the robot managed to show its choice (between rock, papers and scissors) clearly and unambiguously, while one participant was confused about the rules of the game as no instructions have been given by the interviewer to the participants before the interaction with the robot: "*I didn't know I had to say what I choose. I thought there was a camera and that I had to turn the arrow on the chest*". In addition, all three participants noted that the robot clearly expressed its "emotions" based on the context in which they occurred (e.g., they observed that the robot was sad when it lost a game). Furthermore, they were able to correctly identify most of these emotions, namely, sadness, happiness, and surprise. Regarding the robot's ability to introduce itself appropriately and describe its purpose, all participants indicated that it managed to perform its goal. Moreover, 2 out of 3 participants considered our robot's introduction as "*useful*", "*humane*" and "*welcoming*". Regarding the robot's comprehension skills, one interview participant noted that he never had to correct the robot during their interaction whereas two of the participants stated that they had to correct the robot either "*once*" or "*often*".

Emotion and Affective Computing: Emotion and Affective Computing were evaluated through responses to questions *B, C, F, G, H, I, J* and *O*. Regarding first impressions the feedback was overwhelmingly positive as the participants described the robot as "*cute*", "*fun*" and "*kind*". Moreover, the physical design of the machine received positive reviews: The chest of the robot was described as "*cute*" and "*likeable*" with a "*very clear*" functionality. Similarly, the LED eyes were described as "*nice*" and "*fitting*". Furthermore, the participants thought that the eyes "*added expression*" to the robot and that the emotions were clearly displayed. When asked about the intelligence of the robot and whether they would interact with the robot in the future, the participant's answers differed, as 2 of them found the robot "*pretty intelligent*" and would like to interact with it again, while 1 of them stated that he/she thought that the robot was "*Not (smart) at all. I thought it chose randomly*" and would not seek further interaction with the robot giving the following explanation: "*There are no mind games, and rock-paper-scissors is a mind game with your opponent*". Lastly, all participants concluded that they felt "*(very) comfortable*" with the robot.

Technical aspects: Technical aspects of the robot were evaluated through responses to questions *P, Q, R* and *S*. The participants considered the machine's accent as "*good*" or "*unnoticeable*" and the voice was characterized as "*understandable*" and "*nice*" Additionally, the speech rate was deemed "*good*". However, the robot's volume was described as "soft" or inadequately high by two of the participants.

Regarding task completion, our robot performed admirably. Given the time constraints of the project, we focused on the central functions of the robot and as a result it performed successfully all five major tasks, namely, making a choice and presenting it, understanding the participants choice, displaying an "emotion" related to the outcome of the game or the course of the conversation, recognizing and greeting the human and conversing with the human about the rules of the game or the capabilities of the robot. Furthermore, according to the evaluation results extracted from the Emotion and Affective Computing method, the evocation of positive emotions was more successful than in previous projects.

Although the feedback we received was mostly positive, some limitations of the robot's natural language understanding abilities are observed. The robot's shortcomings not only affected its efficiency, but also the user's experience and the positive emotions felt by the participants. Although this problem may be partly due to an inadequate microphone setup, it is possible that either the dialogflow conversation agent created, was not complex enough or that the interface between the robot, the Arduino code, the Python code and the cloud agent was problematic - different programming languages and functions were running in parallel and consequently a small glitch (e.g., internet speed drops momentarily) may be catastrophic. Thus, using an external microphone, improving the dialogflow agent (e.g., more intents and entities) or changing the interface between the robot and the code could improve the user's experience and the robot's efficiency. Furthermore, part of the solution could be enhanced communication between the robot and the participants. The machine could communicate, for instance, that there is too much external noise, and that the human should either increase the volume of their voice or resume the interaction when the external noise has subsided. In addition, the robot could inform the participant if an error occurs, so the human can restart the conversation instead of trying to continue the interaction. With that feedback, participants should be able to adjust their behavior accordingly.

Remarks regarding Legibility and Predictability: The analysis of the predictability in a robot's motion described by Dragan et al. [8] were somewhat inapplicable in our case as the robot's motion is one dimensional (one can observe this by formulating the equations of motion in polar coordinates). A motion is described as predictable if the participant can predict the robot's motion when its goal is known. In our case our robot is mostly predictable (see APPENDIX A). Essentially, when the participant understands the physical constraints of our motor as well as the context of the game (i.e., the robot can play up to four moves) then the movement can be considered as predictable. However, as no guidance was given to the participants during the interview, it is plausible that some of the participants did not understand the physical design constraints (e.g., the servo rotor can move up to 180º) and therefore the motion was not fully predictable. Moreover, quantifying the predictability of the motion (and therefore concluding that the particular movement is fully predictable) can only be done when one tests specifically for this with human participants. A movement is completely predictable when it matches the prediction made by the participant. Thus, without knowing the latter, one cannot know whether the particular move is predictable or not.

Legibility in a social robot that plays rock, paper, scissors is bilateral. On the one hand, we want participants to instantly understand the intention conveyed by the robot through its facial expressions. The quicker the participants are able to discern the intent the more legible the behavior is. Thus, in this case legible motion is a desirable property. On the other hand, if our participants can "quickly and confidently infer the correct goal based on the robot's motion" [8], then they will always win the game since inferring the correct goal means predicting the rock, paper, scissors choice of the robot and thus countering it with the winning card. That would in turn nullify the whole point of the game whilst simultaneously rendering the robot's capabilities useless (e.g., no range of "emotions" would be displayed etc.). Our robot's servo rotor movements are illegible since each time the game starts, the robot defaults to the "go" (empty) state prior to making and showing its choice. Thus, there is only one possible move that it can do (i.e., move counterclockwise) and therefore the goal generally cannot be inferred based on the robot's motion (i.e., all three choices are possible). This coincides with Dragan's assumption that "Predictability and legibility are fundamentally different properties of motion, stemming from observer inferences in opposing directions" [8].

IV. CONCLUSION: DESIGN IMPROVEMENTS

Although our robot received generally positive comments, if one was to deploy it in real-life settings and environments the chances of it being successful would be minimal. Therefore, a number of communicative aspects of the robot that need to be changed as well as the rationale behind those changes are presented in this section.

Component improvements: The first hardware change would be the addition of a robotic arm and hand that perform similar movements to a human hand in order to show the robot's choice. This way our robot's movement would be performed in 3D space whilst simultaneously complete predictability is avoided. Consequently, the motion would be "*mindful of how it is interpreted by human collaborators*" as described by Dragan et al. [8]. Secondly, the existing hardware should be updated, i.e., high quality cameras and microphones should be used in order to minimize the number of times the robot defaults to its fallback intents, the interaction restarts abruptly due to a "lost" face or the sound is misinterpreted. This, in turn, would provide the user with a smoother interaction, resolving problems identified by the human participants, such as the robot not hearing the participant, restarting the conversation or listening to itself.

Software improvements: Further improvements with regard to the innate communication problems presented by hardware constraints, can be realized by using software solutions. This could be done, either by employing more sophisticated algorithms (i.e., noise cancelation, transforming the pre-existing microphone into a directional microphone and better face recognition algorithms) or by strengthening the communication between the robot and the participants. The robot could communicate, for example, that there is too much external noise, and that the human should act accordingly (e.g., increase the volume of their voice or resume the interaction when the external noise has subsided). In addition, the robot could inform the participant if an error occurs, so the human can restart the conversation instead of trying to continue the interaction. With that feedback, participants should be able to adjust their behavior accordingly. Moreover, based on the feedback we received, in some cases the robot was perceived as unintelligent: "*I didn't think the robot was smart, it just made its choice randomly*". Since we created a social toy robot, it would be interesting to see some form of reinforcement/deep learning built into our machine in order to make it act "intelligently". That way, our robot could learn the gaming patterns its enemies are exhibiting and use them to counter their moves. For instance, if a participant always chooses "scissors", then the robot could always choose "rock". Additionally, the robot would be able to assess various characteristics of the participants such as their emotions or whether the participant has interacted with the robot in the past and adjust its playing strategy accordingly.

Interaction Improvements: Recent research provides strong argumentation for implementing "emotions" into a social robot's behavior. According to Bartneck et al. [9] "*taking emotions into account in the design of a robot can help improve the intuitiveness of the human robot interaction*". Generally, emotions expressed by a human or a robot provide the outside world with information about their internal affective state, which is helpful to others in two ways. First, emotions convey information about them and their potential future actions. For example, displaying anger and frustration signals to others that you may be preparing for an aggressive response. In addition, emotions can convey information about the environment. Thus, successful communication of emotions promotes survival, enhances social bonds and minimizes the chances of social rejection and interpersonal physical aggression

[10] and would consequently be a major interaction improvement if implemented correctly.

Currently, our robot expresses four facial emotional expressions, namely, sad, angry, surprised and happy. This in turn creates a limited interaction experience, as the emotional range is significantly wider. A better approach to this would be to use the framework provided by Mehrabian where emotions are "captured" in a three-dimensional continuous space and therefore most of the emotions experienced can be accurately represented. By expanding the emotional range, the robot can both express and recognize in a human’s face we can now create a robot that responds to human emotions. The most straightforward way of programming emotional responsiveness for social robots is through mimicry. Mimicking in humans has been shown to create an idea of shared reality: you indicate that you fully understand the other person's situation, which creates closeness [9]. The exception here might be anger: responding to an angry person by yelling back usually does not facilitate mutual understanding or a resolution of the conflict. Thus, implementing mimicry in our robot (e.g., via the LED eyes and the neck servo) will improve the interaction while simultaneously managing the user’s expectations i.e., when users perceive the robot to be emotionally responsive, they may extend this observation to expectations about the robot's compliance with other social norms. Therefore, the robot’s emotional responsiveness should match its capabilities to fulfill other expectations. Lastly, another interaction improvement could be implementing some form of backchanneling. Backchanneling is a component of conversation and verbalization that is naturally embedded in our role as a listener in a conversation. Throughout a conversation, the listener may nod their head periodically or use short utterances, such as yeah, ok, uh huh, to show that they are engaged. Recent literature verified the importance of social cues like backchanneling [11] and therefore implementing such a feature would be a vital enhancement of passive communication, between the robot and the human participants, which is currently missing from our robot.

Inclusivity Improvements: When implementing our innovative design described in the previous paragraphs, we need to take into consideration the values of the users and improve the inclusiveness of our design. Utilizing user feedback and mindfully designing with inclusiveness in mind is important in order to account for the real possibility that during development, when the design team was effectively engaging in making the moral and ethical solutions, our robot and the software it contains may have automated and perpetuated unfair and discriminatory patterns [12]. In order to do that the “Reflexivity and Futures tool to reflect on the implicit assumptions” that was provided by Zaga et al., [13] was used to combat both ableism and algorithmic biases. Our novel design will include a computer vision algorithm that will detect both the movements of the human fingers to identify the three possible outcomes (rock, paper, scissors) and the face of the human participant to initiate and complete the interaction as well as perform sentiment analysis. Intrinsically in such AI algorithms one can observe algorithmic biases. For example, if the computer vision algorithm was trained primarily using images depicting either fingers or faces from Caucasian participants, then it will have a lower than average accuracy rate in identifying certain fingers (and thus game movements) or facial expressions from participants of different races. Moreover, some participants may have health problems (e.g., participant with amelia, or hearing loss). To ensure that differently abled people can interact with our robot and that our design is inclusive, multiple evaluations should be conducted with such participants in order to qualitatively and quantitatively assess the inclusivity and success of our machine. Lastly, following the “practices for fair robot learning” (see APPENDIX B) provided by Laura Londono et al. [14] and promoting a human centered approach when designing the robot can improve inclusivity significantly.

APPENDIX A

EXPLANATION OF PREDICTABILITY

Our robot’s chest hand only moves in one direction (i.e., it rotates around itself). Furthermore, our design is a semicircle where the four choices are displayed, namely, the rock, papers, scissors and go icons. Those icons are equally distanced from each other (in polar coordinates). Therefore, if the robot's hand was at the rock state and the robot made the choice to move to the paper state it would do so clockwise (only 60º rotation instead of 300º). Similarly, if it wanted to go to scissors it would do so clockwise (120º rotation instead of 240º). Similar argumentation can be provided when moving from one rock, paper, scissors, state to another rock, paper, scissors state. (i.e., distance between those states is always closer and thus the robot would never rotate towards the empty part of the circular chest). Thus, our robot is deterministic: if the robot's goal is known, then its motion can be deducted. In addition, our robot has another constraint: it can only move 180°. Thus, the above argumentation also holds true due to technical constrains. However, even if one analyzed the movement of the robot's hand completely mathematically (no technical constrains exist and the robot can turn 360°), as it will be seen by a random human participant, still the robot would never rotate towards the stateless part of the circular chest.

APPENDIX B

PRACTISES FOR FAIR ROBOT LEARNING [14]

- Human-centered design: robot learning solutions involving the human perspective in all stages.
- Interdisciplinary perspectives: for robot learning development and evaluation.
- Explicit guidelines and policies for inclusion and nondiscrimination: to increase diversity in work teams to provide integral solutions. Ideally, interdisciplinary and intercultural research groups will include different social perspectives to create empathetic solutions.
- Optimization techniques: different from optimization for efficiency and efficacy, explore approaches including constraints for optimization towards responsible robotics.

Methods for ethical decision making.

- Early detection and correcting unwanted robot behavior: explore robot learning frameworks that penalize unwanted robot behavior and testing environments with novel metrics.
- Online evaluation to supervise models for responsible robotics: aim for models with adaptability skills. Include constant evaluation that includes new possible scenarios.
- Informed consequences and impact of the developments: to clarify the social impact and possible reach of the development.
- Conscious development: since robots can physically interact with humans, this implies that they can also physically harm them. It is essential to keep this in mind throughout the development.

REFERENCES

[1] C. Zaga, "Non-Verbal Interaction in Human-Robot Communication," Enschede, 2022.

[2] A. Bonarini, "Communication in Human-Robot Interaction," *Current Robotics Reports,* pp. 279-285, 2020.

[3] K. Bergmann, F. Eyssel and S. Kopp, "A Second Chance to Make a First Impression? How Appearance and Nonverbal Behavior Affect Perceived Warmth and Competence of Virtual Agents over Time," in *International Conference on Intelligent Virtual Agents*, 2012.

[4] C. Zaga, The Design of Robothings: Non-Anthropomorphic and Non-Verbal Robots to Promote Children's Collaboration Through Play, Enschede: University of Twente, 2021.

[5] M. M. V. Pinxteren, M. Pluymaekers and J. G. Lemmink, "Human-like communication in conversational agents: a literature review and research agenda," *Journal of Service Management,* vol. 31, no. 2, pp. 203-225, 2020.

[6] M. Tomasello, Origins of human communication, MIT press, 2010.

[7] P. H. Kahn, N. G. Freier, T. Kanda, H. Ishiguro, J. H. Ruckert, R. L. Severson and S. K. Kane, "Design Patterns for Sociality in Human-Robot

Interaction," in *3rd ACM/IEEE international conference on Human robot interaction*, Amsterdam, 2008.

[8] A. D. Dragan, K. C. T. Lee and S. S. Srinivasa, "Legibility and predictability of robot motion," in *8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2013.

[9] C. Bartneck, T. Belpaeime, F. Eyssel, T. Kanda, M. Keijsers and S. Sabanovic, "Emotion," in *Human-Robot Interaction: An Introduction*, Cambridge University Press, 2019, pp. 114-126.

[10] P. A. Andersen and L. K. Guerrero, "Principles of communication and emotion in social interaction," *Handbook of communication and emotion: Research, theory, applications, and contexts,* pp. 49-96, 1998.

[11] H. W. Park, M. Gelsomini, J. J. Lee, T. Zhu and a. C. Breazeal, "Backchannel Opportunity Prediction for Social Robot Listeners," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2017.

[12] A. Birhane, "Algorithmic injustice: a relational ethics approach," *Patterns,* 2021.

[13] C. Zaga, M. L. Lupetti, N. Cila, M. Lee, G. Huisman, E. F. Villaronga and A. Arzberger, "Tools and Techniques to promote reflexivity with the embodied AI community".

[14] L. Londono, J. V. Hurtado, N. Hertz, P. Kellmeyer, S. Voeneky and A. Valada, "Fairness and Bias in Robot Learning," 7 July 2022.