Designing a Personality-Driven Robot for a Human-Robot Interaction Scenario

Hadi Beik Mohammadi*, Nikoletta Xirakia*, Fares Abawi, Irina Barykina, Krishnan Chandran, Gitanjali Nair, Cuong Nguyen, Daniel Speck, Tayfun Alpay, Sascha Griffiths, Stefan Heinrich, Erik Strahl, Cornelius Weber, Stefan Wermter

Abstract—In this paper, we present an autonomous AI system designed for a Human-Robot Interaction (HRI) study, set around a dice game scenario. We conduct a case study to answer our research question: Does a robot with a socially engaged personality lead to a higher acceptance than a competitive personality? The flexibility of our proposed system allows us to construct and attribute two different personalities to a humanoid robot: a socially engaged personality that maximizes its user interaction and a competitive personality that is focused on playing and winning the game. We evaluate both personalities in a user study, in which the participants play a turn-taking dice game with the robot. Each personality is assessed with four different evaluation tools: 1) the Godspeed Questionnaire, 2) the Mind Perception Questionnaire, 3) a custom questionnaire concerning the overall HRI experience, and 4) a Convolutional Neural Network analyzing the emotions on the participants’ facial feedback throughout the game. Our results show that the socially engaged personality evokes stronger emotions among the participants and is rated higher in likability and animacy than the competitive one. We conclude that designing the robot with a socially engaged personality contributes to a higher acceptance within an HRI scenario.

I. INTRODUCTION

Designing social robots involves two seemingly contrasting objectives. One is to cater to the functional expectations of users, and the other is to fulfill their social expectations. The former requires a rational intelligence focused on problem-solving and achieving a certain goal or outcome within a given scenario. The latter, however, requires a goal-independent framework for social intelligence [1]. We therefore develop both a social and a competitive robot personality and investigate whether these personalities lead to a different acceptance from the users.

Previous studies have shown that a robot’s appearance affects human perception with regard to social acceptance. Particularly, a robot’s physiognomy affects its likability, sociability, and safety [2], as humans often reject interaction with a “strange” or uncanny looking robot [3]. For example, the head can project non-verbal cues mediated through its face in social interactions [4, 5]. A faceless robot is anonymous [6], and therefore less prominent in meaningful social engagement. Moreover, robots with expressive and animated faces are perceived as more intelligent and are likely to capture the user’s attention [7, 8] in comparison to a less animated robot. Gao et al. [9] and Heider et al. [10] studied the perception of random moving shapes and association of the movements with social interactive behaviors. Gao argues that people can attribute animacy and lifelike characteristics while Heider claims that people attribute motives and personality to moving shapes. Significant findings were demonstrated by Visch et al. [11], who investigated the effects of animated films. The study argues that animacy is responsible for causing higher immersion and facilitates stronger emotions among humans. Interestingly enough, people tend to use social rules when interacting with computers and agents endowed with human-like characteristics [12, 13]. In fact, robots enriched with anthropomorphic features, verbal and non-verbal behaviors are perceived as social actors [12] manifesting a personality [13], and are found to be more suitable for meaningful social interactions with humans [14]. Nonetheless, a significant level of anthropomorphism in a robot’s appearance can create higher expectations towards its capabilities and consequently generate disappointment [2]. Supplementary HRI studies propose several interaction patterns [15–18] for designing social robots. Their aim is to enhance a robot’s sociality and manifest a personality, leading to a more pleasant interaction. In this study, we apply a variety of these proposed patterns with special attention to humor [19]. This is a complex challenge that requires the combination of multiple modalities [20] such as e.g. laughter in robots [21], as well as manipulating the voice’s pitch [22] based on the sentence’s sentiment. Related studies have shown that when humor is successfully portrayed, it can
TABLE I: Exemplary differences between the two personalities (competitive and socially engaged), based on their respective dialog acts. Generally, the socially engaged personality has more interactive dialog options while the competitive personality has more game strategy options.

<table>
<thead>
<tr>
<th>Utterance</th>
<th>Emotion</th>
<th>Motion</th>
<th>Comp.</th>
<th>Soc. eng.</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Hi! Finally, a person! My friends told me that you are here to play a game with me.”</td>
<td>Happiness</td>
<td>Short wave</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>“Do you know what is a robot’s favorite music? Heavy metal!”</td>
<td>Happiness</td>
<td>None</td>
<td>×</td>
<td>✓</td>
</tr>
<tr>
<td>“Your Current Score is X”</td>
<td>Neutral</td>
<td>None</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>“Oh my god! I cannot believe I won! I am really good at this!”</td>
<td>Happiness</td>
<td>None</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>“This is so exciting, winning makes me really happy, I wish I could have some ice-cream now.”</td>
<td>Happiness</td>
<td>None</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>“I’m very sad that I lost the game. I really need some ice-cream now.”</td>
<td>Sadness</td>
<td>Cover the face with the right hand.</td>
<td>×</td>
<td>✓</td>
</tr>
</tbody>
</table>

B. Designing NICO’s appearance

We modified NICO to look, behave, and sound like a 10-year-old boy. We concealed all mechanical body parts, using a sweatshirt and a child’s ski hat (see Fig. 2). We also assigned it the voice of a boy and manipulated its pitch and prosody making its utterances more lively [22].

To create the feeling of NICO being more animated, empathetic, and humanlike, we combined the execution of static gestures and facial expressions with specific utterances in the Dialog module. NICO uses LED matrices including two 8x8 matrices for the eyebrows and an 8x16 matrix for the mouth to display expressions like happiness. In our experiment, we focused on the expressions: neutral, happy, angry, surprised, and sad.

III. IMPLEMENTATION

The overall architecture consists of five main modules built on top of the ROS middleware [24]. A scheduler coordinates four distinct modules for controlling dialog, motion, vision, and emotion (see Fig. 3). The robot acts autonomously for the entire scenario. The following sections give a detailed description of each module.

A. Scheduler

The scheduler is a finite-state machine that splits the flow of the game into several episodes, each representing a logical unit of interaction and execution of different peripheral tasks. For example, the “introduction state” triggers a sequence of actions: motion task “head up”, dialog task “Hi!”, and emotion task “happy”. The scheduler assigns these tasks, synchronizes their execution, and chooses the next state.

B. Vision module

The Vision module captures visual input from a ceiling camera, providing frames of the entire table, and extracts visual information such as dice localization and number of dice. The dice’s location on the table in each frame is propagated to the Motion module and the number of pips is sent to the scheduler at 5 fps. The pip counting is achieved by identifying the dice from a set of contours detected by the Sobel operator for edge detection. As soon as the dice...
is localized, we extract the pips by another Sobel operator to separate the top surface. The pip number is counted with the built-in blob detector of OpenCV.

C. Motion module

The Motion module controls NICO’s physical interaction and also projects body language cues. It executes two types of tasks, grasping and gestures. Grasping enables the NICO to perform movements which are affected by external factors, e.g. “grasping the dice from the table”. In contrast, gestures are pre-defined movements designed to imitate the body movements observed in human-human interaction (e.g. co-verbal gestures), making our HRI experience more natural.

NICO is capable of grasping, with a three-fingered SR-RH4D hand, consisting of four DoF. Two Optoforce sensors are mounted on its left hand to detect failures.

Both tasks are executed using the same workflow. Input from the scheduler activates the corresponding controller, which generates discrete motion path configurations. Pypot is a low-level controller, completes, smooths and executes the generated paths. Finally, the inspection unit investigates any failures by using the Optoforce sensors and sends the resulting signal to the scheduler.

The grasping task starts by receiving the dice’s coordinates from the Vision module. The closest matching configuration is found from the lookup table of a scenario-specific dataset, consisting of dice positions and their corresponding grasp configuration. The inspection unit determines if the dice is reachable and decides if the task ended successfully. We define two types of constraints. A distance threshold between the dice’s position and the closest point in the dataset, and a dynamic pressure threshold for the Optoforce sensors. If grasping fails, the robot asks the participant to hand over the dice to continue the game.

D. Dialog module

The Spoken Dialog System (SDS) handles the direct verbal interaction with the user via multiple language-related submodules. The overall dialog is broken down into scenes, callables scripts of phrases NICO utters, which acquire the user’s response. The scenes act as an entry point to the SDS and are triggered by the scheduler’s currently active episode. The SDS structure follows the conventional flow of a general dialog system [25]. It communicates with the scheduler, triggers the scenes and forwards facial expressions as well as motion gestures back to the scheduler.

The input and output of the Dialog Manager (DM) are handled by the Automatic Speech Recognizer (ASR) and the Text-to-Speech (TTS) engines respectively. Speech is converted into textual form and then propagated to the SDS by the ASR engine. The ASR uses the Google Speech Recognizer with DOCKS as a post-processing technique [26]. We use Amazon Polly for the TTS task, due to the variety of voices it provides and the possibility to manipulate the voice’s pitch via Speech Synthesis Markup Language.

The Natural Language Understanding (NLU) unit processes the text received from the ASR and performs two tasks: extracting named entities and classifying the intent within the text. The Named Entity Recognition (NER) enables NICO to collect information about the user throughout the game. We use SpaCy to create a model for the NER subtask. We train our model on the Wikipedia corpus [27], and extend it with the Yelp dataset [28] which allows us to support food entities to imply that NICO feels hungry. This also allows the robot to engage the user in a personal conversation about their individual food preferences. For instance, if the user says “I am Erika and I like pizza”, NICO recognizes that Erika is a name and pizza is a food type. The intent in the user’s speech is extracted by the semantic parsing unit. We used the MIT Information Extraction (MITIE) [28] toolkit for this task, treating it as a supervised classification problem. Multiple phrases with a defined polarity are trained independently on two intent classes. For our application, we train a model to classify the polarity intents and repetition requests. For every phrase uttered by the participant, a polarity intention returns either a “yes” or a “no”, while a repetition intent returns a “true” or “false”. Finally, custom phrases were added due to the lack of labeled datasets for such tasks, to the best of our knowledge. The custom phrases were inspired by individual sequences selected from the NPS chat corpus [29] and were manually labeled.
The Natural Language Generation (NLG) unit pre-processes the sentences generated for speech synthesis. The phrases are generated in the scenes and passed with slot placeholders to the DM, which decides on the phrases that should be uttered. Before utterance, the NLG replaces the slot placeholders with entities acquired from the NLP output. An example of a phrase in the scenes would be: [Hey (PERSON)1]. After following the dialog manager pipeline, the NLG processor output would e.g. become: [Hey ERIKA!]. Several phrases can be provided for a single plan item while the one that is chosen for an utterance is based on a round robin policy which selects the phrases in a cycle.

The Dialog Manager (DM) is a frame-based dialog management system [30] due to our general system architecture. Since the scheduler controls the flow of the experiment, the functionality of the DM is restricted to controlling scenes, which require no interaction with other modules. The scenes define NICO’s responses according to the user’s utterance. Such responses are designed to handle the rejection and the acceptance of a sub-dialog, or even rephrasing the question if the user’s utterance did not satisfy the query.

The scenes were designed with the goal of maintaining succinctness. Griffiths et al. conducted a study on users interacting with virtual agents, showing that users responded with longer utterances in alignment with the agent’s utterance [31]. Longer utterances by the users would increase the possibility of erroneous speech recognition and were therefore avoided. However, the scenes explaining the game rules were intentionally designed to be more verbose as it is suggested that longer dialog conveys instructional content more clearly [31].

A mixed-initiative-based system would make the conversation more realistic and appeal to the users [32]. Frame-based dialog management systems generally expect a mixed initiative, a requirement which is fulfilled by our SDS. When NICO awaits a response from the user, they may decide to request their score or a repetition of the previous utterance. These requests are called universal commands, which are actively checked throughout the conversation, regardless of the current scene. They cause ongoing conversations to be briefly paused and resumed later.

E. Emotion module

The main purpose of the Emotion module is to gather additional raw data about the participants’ reactions to NICO. Questionnaires that are given after an experiment have multiple known disadvantages such as e.g. potential dishonesty, wrong interpretations, a lack of personalization, and the difficulty of capturing emotions [8]. By analyzing the participants emotions during the experiment, we hope to mitigate some of these issues (serving a similar function as the established practice of using EEG or fMRI in cognitive neuroscience studies).

The Emotion module consists of three different parts: face detection, face tracking, and facial expression/emotion recognition. Raw camera output is read and pre-processed using OpenCV. The face detection is realized with a Convolutional Neural Network (CNN) model from the Dlib library [33] running on the system’s GPU, which is held in memory to allow for fast re-detection of the participant’s face. Once a face is found, its region of interest gets propagated to the face tracker based on the correlation of a new frame compared to previous frames. This allows for real-time execution due to its computationally cheap nature. If the tracking quality falls beneath a threshold, the face detection gets activated again. When the scheduler needs information on the participant’s attention, i.e. either looking towards NICO (attentive) or elsewhere (distracted), it schedules a callback function to the Emotion module, which gives a rough estimate of the participant’s current attention. This is implemented by averaging over the last five frames with the face detection algorithm and the confidence scores. The callback method waits for the average attention to be equal to the scheduler’s goal. During tracking, the region of interest (participant’s face) is cropped out, converted to grayscale and fed to the emotion recognition CNN from Barros et al. [34]. This architecture extracts features from the face and body posture of the human. In contrast to the original paper, we only use the face channel. The network was trained on the FER+ dataset [35] and classifies each input into eight different emotion classes of which we use: Neutral, Happy, Surprised, Sad, and Angry.

IV. EXPERIMENTAL EVALUATION

We conducted a user study where the participants were asked to engage in a face-to-face interaction with NICO. The objective of the user study is to assess each personality and get a deeper understanding of how our robot is perceived.

A. Experiment setup

After introducing the general rules of communicating with the robot, participants were led to the experiment space, which was fenced by white covers decorated with posters to create the illusion of being in a child’s room. These covers also hid the experimenter from the participant to minimize influence from external factors. Each experiment consisted of two rounds of our dice game, one for each personality. The order of presentation for the personalities was randomly shuffled to rule out ordering effects. Overall, 22 participants between the age of 25-34 years took part in our user study. The participants rated their previous experience with robots as little (5%), some (50%) or frequent (45%). They filled out a questionnaire after each game and the goal of our study was revealed at the end of the experiment.

B. Evaluation tools

To evaluate our experiment, we used three questionnaires and analyzed the users’ emotional feedback by autonomously classifying their facial expressions when interacting with NICO. According to Bartneck et al. [8], a substantial drawback of questionnaires is the fact that they are given to a participant after an experiment. In contrast, a participant’s facial expressions capture the emotional reactions during the experiment.
Fig. 4: Results of Godspeed questionnaire. Shown are sample mean values (per scale) with 95% bootstrap confidence intervals. Used abbreviations: ANI - animacy, LIK - likability, INT - intelligence, ANT - anthropomorphism, SFT - safety.

Table II: Results of Godspeed questionnaire. Shown are means and standard deviations of the two samples (per scale) and p-values by one-tailed Welch's t-test.

<table>
<thead>
<tr>
<th>Scale</th>
<th>Soc. engaged</th>
<th>Competitive</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANI</td>
<td>3.51±0.54</td>
<td>3.12±0.57</td>
<td>0.01</td>
</tr>
<tr>
<td>LIK</td>
<td>4.21±0.58</td>
<td>3.87±0.56</td>
<td>0.03</td>
</tr>
<tr>
<td>INT</td>
<td>3.37±0.54</td>
<td>3.15±0.60</td>
<td>0.11</td>
</tr>
<tr>
<td>ANT</td>
<td>3.22±0.85</td>
<td>2.98±0.70</td>
<td>0.16</td>
</tr>
<tr>
<td>SFT</td>
<td>3.53±0.55</td>
<td>3.59±0.51</td>
<td>0.35</td>
</tr>
</tbody>
</table>

Table III: Results of the Mind Perception questionnaire. The means, standard deviations of the two samples (per scale) and p-values given by one-tailed Welch's t-test. Used abbreviations: AGN - agency, EXP - experience.

Table III shows the per-scale comparison of the two samples for the Mind Perception questionnaire and its corresponding statistics. The responses to this questionnaire have large variances in both samples, thus, neither the per-item [37, 38] test shows any statistically significant difference between samples. Nevertheless, on average the participants rated the socially engaged NICO higher on both Mind Perception dimensions.

Our custom questionnaire (see Fig. 5) highlights that the socially engaged NICO is indeed perceived as able to express humor, anger and joviality as well as able to experience empathy, happiness, and joviality. The analysis of the participants’ emotions reveals some interesting aspects concerning the effect of a robot’s personality on people’s emotional rapport. We performed the one-tailed two-sample Welch’s t-test on the recorded emotional expressions.

Table IV: Welch’s t-test result on the emotions which are captured from participant facial expressions during the game. Each emotion entry indicates the t-test result performed on the participant’s second interaction with the socially engaged competitive robot.

We observe that the participants display less neutrality and more happiness in their emotions while interacting with the socially engaged robot in comparison to the competitive robot. This can be seen in Table IV. From this, we may interpret that the participants realized the lightheartedness that the socially engaged robot introduced in its interaction through the dialog and the gestures. As a result, they experience more happiness when they interacted with the socially engaged robot. Furthermore, t-test results show that the participants displayed more happiness, surprise, and anger if they interact...
with the socially engaged robot personality (see Table IV). We speculate that this is due to participants being more immersed and showing stronger emotional reactions while interacting with the socially engaged personality, which could be an indication of animacy attribution [11].

V. Discussion

In this study, we investigate whether a socially engaged robot personality based on high social intelligence leads to better acceptance in a game scenario than a competitive personality based on mere rational intelligence. To answer our research question, we developed an AI system with two different robot personalities: one with augmented social capabilities and one task-oriented character.

Having a dedicated scheduler handling the communication between all other modules allowed us to design the two different personalities within the same scenario, by simply substituting the personality’s transcripts in the dialog. The design of our Dialog module combines a mixed-initiative architecture and a mechanism to unify utterances along with static gestures and facial expressions. This makes the conversation more realistic and NICO’s utterances more natural and life-like. Moreover, the ability to modify the pitch and prosody of the voice enabled NICO to produce vocal emotions and sound less robotic.

In the user study, the participants assessed the socially engaged personality as having slightly higher animacy and likability. Our automated analysis of the participants’ expressions suggests that the participants expressed stronger emotions when interacting with the socially engaged NICO, which could indicate the effects of animacy [11]. Although we could not measure any significant relationship between personality and the perceived intelligence, we observed a slight positive trend towards the socially engaged personality.

Some of the challenges that we faced during the study concerned technical limitations and the participant population. The former is a possible problem with almost all HRI studies. The personality design process was limited by the fact that some traits of personality could not be adequately expressed or have caused unfulfilled expectations for the participants [2]. Moreover, while the socially engaged personality was designed to have richer verbal communication, it was more prominent to errors which we believe resulted in a weaker assessment. Finally, we speculate that the differences between the two personalities were not apparent enough to allow the participants to a clearer preference.

Our overall findings suggest that a more humorous and animated personality is preferred for social interaction than a more serious and competitive one. However, anecdotal feedback provided by some participants showed that those of a more competitive nature felt more engaged by playing with the competitive personality, and perceived the socially engaged personality as distracting. We consider this being in connection to people’s tendency projecting their own personality onto the robot [3]. Therefore, we believe that developing a task-oriented character which can adapt to a user’s sociability level might be a suitable design. Future work will allow us to investigate a stronger correlation between NICO’s animacy and perceived intelligence by creating a better distinction towards the behavioral traits of the two personalities in an interaction scenario.

VI. Acknowledgements

The authors gratefully acknowledge partial support from the German Research Foundation DFG under project CML (TRR 169), the European Union under project SECURE (No 642667). We thank Sven Magg, Johannes Twiefel, Pablo Barros and Manfred Eppe for their involvement in this study.
REFERENCES


[38] B. L. Welch, “The generalization of student’s’ problem when several different population variances are involved,” *Biometrika*, vol. 34, no. 1/2, pp. 28–35, 1947.