A Bachelor Thesis in Computer Science: Developing Personality for a Wizard-Controlled Agent

Eda Kurdoglu

Abstract

A Bachelor Thesis in Computer Science: Developing Personality for a Wizard-Controlled Agent

Eda Kurdoglu

**Teknisk- naturvetenskaplig fakultet
UTH-enheten**

Besöksadress:
Ångströmlaboratoriet
Lägerhyddsvägen 1
Hus 4, Plan 0

Postadress:
Box 536
751 21 Uppsala

Telefon:
018 – 471 30 03

Telefax:
018 – 471 30 00

Hemsida:
http://www.teknat.uu.se/student

As social robots are increasingly growing more popular in the world of technology, designing the optimal human-robot interaction is becoming more crucial. In this paper, the user reaction to the agent personality of a wizard-controlled embodied Furhat robot called Neil, will be examined by answering two research questions. The two research questions concern whether it is possible to implement two distinct personalities in an agent that plays a RDG-Map game with a human, and if those two personalities can be expressed in a coherent way by explicitly modeling the affect. For this project, two distinct personalities were implemented in Neil in order to answer these two research questions. The personalities are called OPT and IMP Neil, and were implemented with the help of crowd-sourcing authoring method. The method allows a group of crowd-authors to author dialogue lines for given scenarios, that a participant and robot in question can possibly be in. For this project, 18 participants were gathered to separately play a RDG-Map game with Neil, where half of the group was assigned the OPT Neil, and the rest was assigned the IMP Neil. Results indicate that, although the IMP Neil was not as successfully implemented as the OPT Neil, it is possible to implement two distinct personalities in the given situation. Similarly, it is concluded that it also is possible to express the two personalities in a coherent way by explicitly modeling the effect. Finally, the results show how crucial it is to carry out research to get a sense of what the user's reaction to the personalities will possibly look like.

1 Introduction

There are several studies in the history of HCI and UX that have examined the affect of agent personality on users, to design the ultimate personality for an agent. These studies include agents that have been designed for personal assistance (G.Ball, 2000), television program recommendation, and rehabilitation therapy, and so on (Meerbeek, 2008). Agent personalities have for instance been implemented in the agent's verbal skills, gestures, and facial expressions, and different kinds of personalities have been designed for different purposes (Meerbeek, 2008; Tapus, 2008; Aly, 2008). In general, the more natural and human-like an agent's personality is, the easier it is to keep the users' attention (Aly, 2008). This project is a further research on the affect of agent personality on users, by analysing the unbiased user reaction on two distinct personalities applied to a wizard-controlled agent called Neil. These two distinct, human-like, and natural personalities is a further approach on designing the ultimate personality for an agent, intending to create the best possible user experience. In this project, a human-robot lab experiment was carried out including 18 users, where all of them got to interact with Neil separately by play- ing a collaborative geography game called RDP-Map (Rapid Dialogue Game - Map). Every participant was, without their knowledge, assigned to either of the two personalities. In this paper, the human-robot interaction between these users and Neil, and their user reaction to the two agent personalities will be evaluated. The purpose of this project is to examine the user reaction on the two agent personalities on the 18 users to see if user reaction is predictable, and also to answer two research questions concerning the implementation and development of the personalities.

1.1 The Research Questions

The purpose of this project and the primary focus can be summed up with the following two research questions.

1. (RQ1) Can one implement two distinct personalities in the agent playing the RDG-Map game?

2. (RQ2) Can one coherently express the two personalities by explicitly modelling the affect?

The first research question simply concerns whether it is possible to design and implement two distinct personalities to Neil, noticeably and recognizably for the participants. The second research question covers if it is possible to develop these personalities gradually throughout the interaction so that both personalities start on a neutral state and move slowly towards their natural state, and reveal their true personalities.

2   Related Work

The same RDG-Map game was initially designed for another project named *"Can you say more about the location?" The Development of a Pedagocial Reference Resoultion Agent* (Paetzel et al., 2019) where the aim was to develop an entertaining learning tool to teach the size, location, and relation of countries in the world. There was both a web-based version and an embodied version of the agent, where the embodied agent was called Neil (same Neil used for this thesis project). Two specific and coherent personalities were implemented in Neil for this thesis project, whilst in the related work mentioned above, he was more neutral and solely reacting when something specific would occur (Paetzel et al., 2019). Neil's results from the related work helped shaping and planning this thesis project, since there is a basis of information from of what the participants think of the neutral Neil. Therefore, it was easier to predict what to expect, and develop the personalities accordingly (Paetzel et al., 2019).

The personalities, OPT and IMP, are based on a related paper named *Expressing Coherent Personality with Incremental Acquisition of Multimodal Behaviors*, which not only is one of the sources of inspiration for the personalities, but also for the method crowd-sourcing authoring (Mota et al., 2018). In the related work, crowd-sourcing is introduced as a relatively new phenomenon that is attractive to system builders as a solution to authoring agent behavior/personality. The method itself will be explained more in detail further in this paper, but a brief definition could be that crowd-sourcing is a method where a group of un-official and independent people, crowd workers, usually found on internet are gathered to finish a specific task. The task in the case of this related work was to author both verbal and non-verbal indicators for an agent to develop its behavior/personality (Mota et al, 2018). The agent in question for this related work is called William and he is identical to Neil in terms of appearance and personalities. The purpose of the work was to evaluate if incremental acquisition of language behavior from crowd-sourcing can create a coherent personality for an agent. Compared to Neil's one-to-one interaction, William was hosting in a trivia competition between two teams and interacting with several people at the same time. Similarly, this related work also helped when developing the personalities, and additionally gave a preview of crowd-sourcing. After the game session, participants were given a questionnaire that covered questions about William (Mota et al, 2018). This is similar to one of this thesis project's post-game questionnaires, where the questions are designed to find out how the users perceived Neil in terms of his personality.

Incremental Acquisition and Reuse of Multimodal Effective Behaviors in a Conversational Agent is another related paper where the personalities and crowd-sourcing platforms are inspired from. The purpose of this related work was to find out whether crowd-sourcing can be used to effectively generate multi modal behaviors that create a consistent and natural development of effect, throughout a conversation. The given scenario for this work is a Pepper robot at a science

fair in Los Angeles, where it asks the visitors to help her to pick a souvenir for her boyfriend as a gift. The robot has either an optimistic (OPT) or impatient (IMP) personality, identical to the he personalities implemented in this thesis project. Again, this related work was a great guidance since the personalities and crowd-sourcing were both used similarly for this thesis project. (Kennedy et al., 2018).s

For an optimal user reaction, the human-robot relationship is crucial. When designing a social robot, the 'social' aspect is what makes a social robot more human-like and less robotic than a regular one. In a related work called *Affective social robots* from the *Robotics and Autonomous Systems* articles (R. Kirby et al, 2009), the development of an affective model for social robots was made in a way that the robots would adhere to human social norms, hence make them as human-like as possible. This development includes qualities such as distinctions between immediate emotional responses, the overall mood of the robot, and attitudes, with a focus on developing long-term human–robot relationships, since it is believed that the people's relationships depend heavily on shared emotional experiences (R. Kirby et al, 2009). The same idea applies for the personalities developed for Neil, where the user reaction is based on the human-robot connection, and analysing the user reaction is the key to answering the research questions (R. Kirby et al, 2009).

For this thesis project, the personalities implanted were distinct for two reasons, first to answer research question 2, and the second reason is to make the chances of Neil being relatable as high as possible. In *Can Robots Manifest Personality?: An Empirical Test of Personality Recognition, Social Responses, and Social Presence in Human–Robot Interaction* (K. Min Lee et al, 2006), 48 participants were involved in a HRI, where the participants were either extroverted or introverted, and so were the robots in question they interacted with. The purpose was to analyse the user reaction to test the personality recognition, and the results show us that the participants enjoyed interacting with the robot of the same personality type like themselves more, rather than their opposite. They would recognize and relate to the robot with similar personality more (K. Min Lee et al, 2006), which without doubt shows how personality is an essential feature for creating socially interactive robots, and making them more human-like. Although the participants in this thesis project were not chosen for their personalities or even asked what kind of personality they have, the two personalities being distinct makes the chances of Neil being relatable higher, since both of the personalities implemented to Neil are very common.
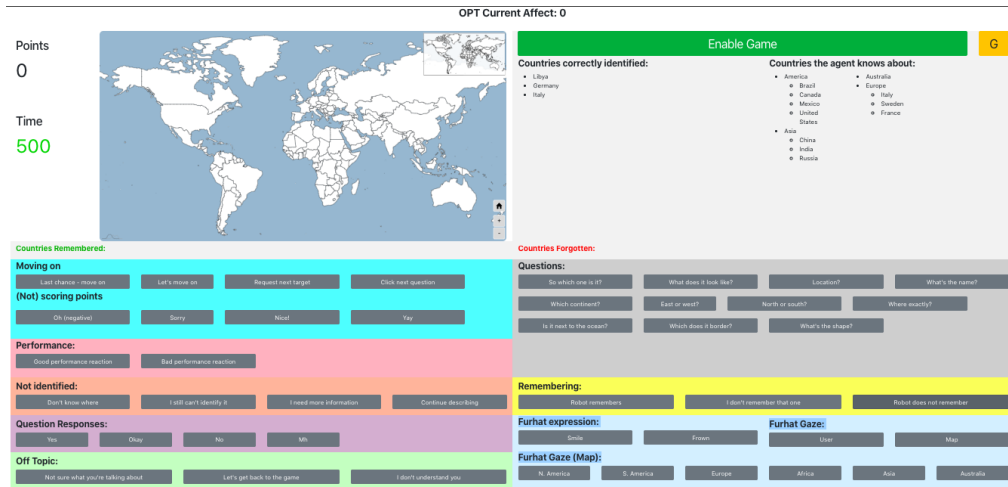
3

Figure 1: The Wizard interface during the game

3 System

3.1 RDG-Map Game

Before introducing the details regarding the personality development of the wizard-controlled Furhat robot Neil, the game will be presented. The game is a geography RDG-Map game played with two players, an agent and a human participant. The game has been developed to teach the players more about the world map, and gain knowledge about where different countries are geographically located. The players are assigned a world map that is shown on a shared screen, and the human player is also given a tablet with the same map. They are assigned different roles, one where Neil is the Matcher, and one where the human participant is the Director. The goal of the game is for the Director to help the Matcher to find the targeted country, which is only highlighted for the Director on the tablet. The verbal interaction is crucial since it is the only way for the Director to describe the selected countries. Initially, Neil can locate eleven countries on the world map (United States, Canada, Mexico, Brazil, India, China, Russia, Australia, Italy, Sweden, and France). The reason why these 11 countries were chosen is because the first 9 countries represent the ones that 50% of Americans can find on the world map (Roper and Geographic, 2006), and Sweden is the country Neil is based in, and finally, France is the only country Neil has traveled to (Paetzel et al., 2019). The Matcher is allowed to ask which countries are neighboring the targeted country, the shape of it, and other related questions that could help him identify it. All information that is needed such as the names of all the countries, is only provided for the Director on the tablet, whilst the Matcher can only see the unlabelled shared screen. The Matcher is allowed to guess two times at most, but the game is also time-constricted to

4

600 seconds. If the Matcher guesses correctly the first time, 2 points will be given. However, if the first guess is incorrect and the second is right, 1 point will be given instead. Even though the Director acts as a tutor, it is meant for them to learn as well. As mentioned Neil was wizard-controlled, meaning that Neil was not autonomous but controlled by a human (Wizard). The Wizard was present in the room as well with a separate computer, to access the Wizard interface that controlled Neil (see Figure 1). However, it is also crucial to mention that the participants were told that Neil was autonomous, and was not informed otherwise until a couple of days after the experiment. Therefore, the Wizard was hiding behind the curtains so the participants would not notice.

3.2   The Wizard and Neil the Agent

Neil is embodied in a back-projected Furhat (Al Moubayed et al., 2012) robot head with an adult male face (see Figure 2). As mentioned, Neil is a Furhat embodied agent with a male's head, face and neck (Figure 2) and he is given a voice that comes from a speech technology called Cereproc where the voice was chosen is called William. Since Neil is embodied, head and gaze movements along with facial expressions are implemented. The robot's facial expressions were created by an expert using the Furhat SDK (Al Moubayed et al., 2012) and projected to Neil's face as accurately as possible. The facial expressions help Neil mimic a more human-like effect, and the head and gaze movements serve additional purposes like revealing where Neil is looking at the world on the shared screen. The Wizard, who in this project was the researcher, was controlling Neil with a separate screen and hiding behind the curtains. The Wizard interface consists mostly of buttons that help the Wizard to control Neil. These buttons are either phrases that allow the Wizard to choose how Neil communicates verbally or instructions to make him move his head or make a facial expression.

3.3   The Personalities

To make Neil a social robot, two different personalities were implemented in him for this project. The personalities are called OPT Neil and IMP Neil (Mota et al., 2018), where half of the participants were assigned to OPT Neil and the rest played with the IMP Neil. The participants were neither informed about the purpose of this project nor the personality implementations. The first personality, OPT Neil, is an optimistic character that is cheerful and fun, whilst the IMP Neil is an impatient character with much less tolerance and empathy than OPT Neil. One of the reasons why these versions were chosen is because of how non-neutral, distinct, and opposite they are.
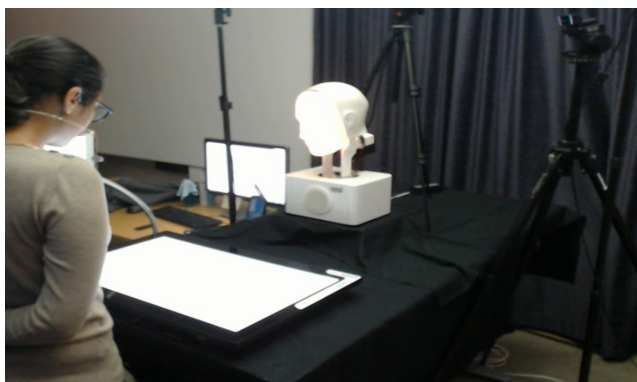
Figure 2: The experimental setup during a interaction between a participant and the embodied Furhat robot Neil

3.4   The Game & Personalities

Neil interacts predominately through verbal communication. The Wizard chooses what Neil says by clicking the buttons with the most appropriate commands (see Figure 1). Each button represents at least one sentence, for example, one button is labeled: **Do you like playing games**, and would make Neil ask the participant the question *"Do you like playing games?"*, another example would be the button: **Are you excited**, which represents the sentence *"Are you excited to play the game?"*. These two are examples of so-called single-sentenced (ss) buttons, which are buttons that represent only one sentence. However, multi-sentenced (ms) buttons such as: **User: excited** and **User: not excited**, are different from the first ones. For example, the question *"Are you excited to play the game?"* is a sentence that is represented by an ss button, and if the user (participant) responds with a *"yes"*, then the Wizard will click the ms button: **User: excited**. How Neil will continue depends on two things: whether he is OPT or IMP, and what his current affect is. Current affect is a count in terms of an integer that reveals Neil's excitement state. The count starts at 0 when the interaction starts, which is also the neutral state and can increment or decrement depending on if he is OPT or IMP, and also how the participant responses. The current state can be incremented up to 5 (the extremely excited state) or decrement to a minimum of -5 (the extremely frustrated state) which depends on how the chosen personality is affected by some particular responses from the participant. So, e.g. if Neil is OPT, the current state is 0, and he asks the participant *"Are you excited to play the game?"*, the user answers *"yes"*, and the Wizard clicks the **"User: excited"** button, the current affect would then increase with 1 and become 1. However, if Neil was IMP instead, the current affect would for example decrease with -1 if the user answers *"no"*. Furthermore, when **User: excited** or any other kind of ms button is chosen, not only might the current affect change but also the most accurate answer for

6

Neil will be chosen. The answer is a sentence from a unique set that the button represents. The ms buttons contain all each a unique set of 20-25 sentences which are all ranked from -5 to 5 (extremely frustrated to extremely excited). The sentence that is closest to the current affect is the one that is chosen by the algorithm. Using the same example as above, if the user is excited and the current affect is 1, the sentence in the set of the button **User: excited** that is ranked closest to number 1 will be chosen. On the excitement scale, 1 is between neutral and slightly excited, in other words, Neil will read from a sentence that is neutral/slightly excited. Each interface has several buttons to be chosen from, including 12 ms buttons which are, again, the buttons that contain the different sentences that might change the current affect. The value of the current affect, in the end, will reveal Neil's state of excitement or frustration.

4   Crowd-Authoring Effective Language

Crowd-sourcing is a method where a group of people or a 'crowd' is gathered to finish a specific task. For this thesis project, a group of random people from the United States of America was hired on the American multinational technology company Amazon's Amazon Mechanical Turk (AMT) crowd-sourcing platform. AMT is a platform where a group of random people is asked to author for specific situations and tasks. For this project, the task was to author sentences for IMP and OPT Neil. The authors were given different scenarios for the pre-game, actual game, and the post-game interactions, and they were asked to author for Neil. In other words, the authors answer as if they were him. These scenarios are personality-related, and the authored answers were added to Neil's verbal repertoire. There were in total of 5 different scenarios, and 2-3 versions per scenario, where two of these scenarios were based on the pre-game conversation, one on the actual game, and the last two ones for the post. These given scenarios are related to the ms buttons, as mentioned earlier, that might affect the current affect, and the sentences were all based on them. The previous example where the Director is asked if he/she is excited is one of the scenarios that belong to the pre-game conversation. For each scenario, there were different possible versions. Here are all the scenarios and their versions listed:

1. First scenario: The game has not started yet and the robot asks the human participant whether he/she is excited about the game

   (a) First version: The participant says he/she is excited
   (b) Second version: The participant says he/she is neutral
   (c) Third version: The participant says he/she is not excited

2. Second scenario: The game has not started yet and the robot asks the human participant whether he/she is good at geography

   (a) First version: The participant says he/she is great at geography
   (b) Second version: The participant says he/she is okay at geography

     (c) Third version: The participant says he/she is not great at geography

3. Third scenario: The game has started and they have at this point guessed a few countries

     (a) First version: They are scoring and the results so far are excellent

     (b) Second version: They are not scoring many points and the results so far are not well

4. Fourth scenario: The game has ended and they are having a post-game conversation

     (a) First version: The robot wants to know whether the human participant think their performance was good or bad and the human participant says he/she thinks it was good

     (b) Second version: The robot wants to know whether the human participant think their performance was good or bad and the human participant says he/she thinks it was not good

5. Fifth scenario: The game has ended and they are having a post-game conversation

     (a) First version: The robot wants to know whether the human participant had fun and the human participant says he/she did have fun

     (b) Second version: The robot wants to know whether the human participant had fun and the human participant says he/she did not have fun

Each version has a set of 25 sentences. This is because the authors (not necessarily the same ones) were asked to author each scenario with five different personality effective states (extremely excited/frustrated, slightly excited/frustrated, and neutral). Since there were 5 scenarios and 12 sub-scenarios/versions, there were 25 times 12 sentences in total. For example, when scenario 2 version b was presented, one of the questions was *"If the robot is overall slightly excited, what would his response be to the player's geography skills?"* The purpose of this method is to gather random sentences that the authors believe are appropriate for each effective state, version, and scenario. Each sentence was checked and rated by other authors on a scale from -5 to 5 (extremely frustrated to extremely excited), as well as if it was typical and ordinary. These are the sentences that were previously described in the introduction, some of these ms buttons such as ***User: excited*** have each a set of 20-25 sentences to find the sentence closest ranked to the current affect. By doing this, Neil chose the sentences that were most accurate for the current situation. The purpose is, once more, to get an opinion on what these different versions can look like in these different scenarios. Not all of the sentences made it to the sets, since some were rated as non-sensical by the other authors, but each version had at least 20 sentences.

5   Method

5.1   Good and Poor Performance

In an attempt to to further examine the personalities and make the user reaction as unbiased as possible, IMP Neil and OPT Neil were divided into two groups each. They were divided into the IMP/OPT Good Performance groups (IMP/OPT GP) and the IMP/OPT Poor Performance groups (IMP/OPT PP). The GP's represent the two halves of both personalities, where the Wizard made Neil guess the countries correctly. The participants assigned to a IMP PP or OPT PP were rarely scoring any points. One of the excluded participants belonged to the IMP GP, whilst the other one belonged to the OPT GP, and the results gathered come from 18 participants in total. Of these 18, 4 participants were in the IMP GP, 5 in IMP PP, 4 in OPT GP, and finally 5 in OPT PP.

5.2   The Participants

There were initially 20 participants that took part in the lab experiment, but as mentioned before two were excluded. All were Uppsala University students with different ethnicity and educational backgrounds. Flyers were made and distributed in ITC (Campus for Informational Technology) at Uppsala University to find volunteers to participate. The two excluded interactions were cut out due to some complications with first participant's affect change, and the other participant had already interacted with Neil before in another experiment, which was not to the knowledge of the researcher up until after the experiment.

5.3   The Pre and Post interactions

Most of the Wizard interface content was inspired from one of the related works (Paetzel et al., 2019), where Neil played the same game. However, the pre- and post-game interaction in the related work were rather neutral, therefore more personality-related buttons were added for this project (see Figure 3 &4).
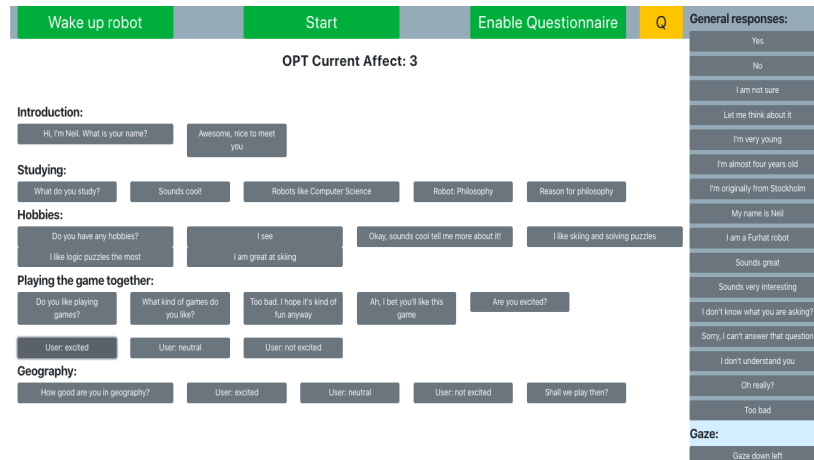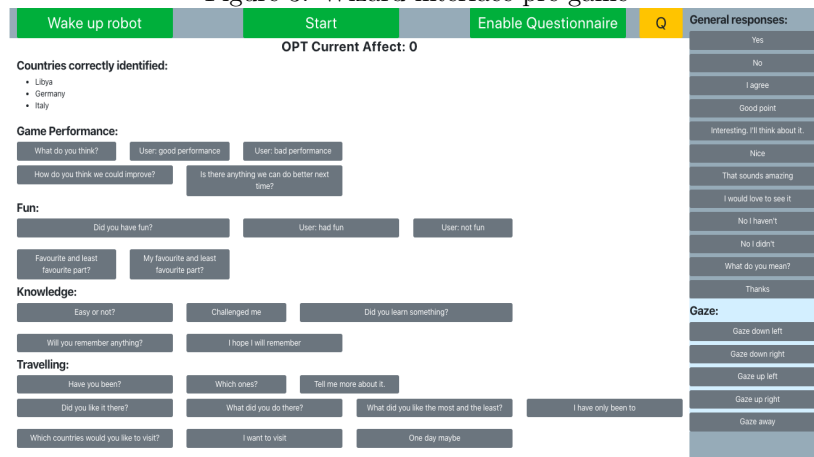
Figure 3: Wizard interface pre-game

Figure 4: Wizard interface post-game

5.4 Questionnaires

The participants were asked to fill one questionnaire before the game, and two after the game. The one before the game and the first one after the game included standard questions about ethnicity and educational background, and none were personality-related. However, the last questionnaire was created for this project solely, and consisted mostly of personality-related questions. This sample questionnaire was connected to the research questions, and created to help answer them. The personality-related sample questionnaire questions are as listed:

1. The interaction with the robot was great (on a scale from -5 to 5, strongly agree to strongly disagree)

2. was given enough time to spend with robot to get to know him (on a scale from -5 to 5, strongly agree to strongly disagree)

3. Describe the robot in 3 words

4. Would you say the robot had a clear/obvious personality? If yes, what kind?

5. Did you find the robot to be more overall excited or not excited? (excited/not excited)

6. Did you find the robot to be more overall frustrated or not frustrated? (frustrated/not frustrated)

7. The robot was optimistic (on a scale from -5 to 5, strongly agree to strongly disagree)

8. The robot was impatient (on a scale from -5 to 5, strongly agree to strongly disagree)

The first question is quite general yet crucial, since it does not jump right in to the personality part, but rather creates a neutral start. The second question wants to find out whether the given time was enough for the participant to get to know Neil's personality. The third question asks the participants to describe the robot in three words. If the participant's answers are personality-related, then that means Neil's personality implementation was somewhat successful. If the participants also mentions accurate words, such as *"optimistic"* for OPT Neil, then it simply means that the implementation was successful. Question 4 is the first time during the whole experiment where the word *"personality"* is ever mentioned. The key words like *"impatient"* or *"optimistic"* are not mentioned until later, and this is to find out if some of the answers will be accurate without being unbiased. Finally, the questions that ask the participant to rate the robot on both impatience, optimism, excitement, and frustration on a scale, will show more clearly whether IMP Neil and OPT Neil have been successfully implemented or not. The attempt to make the user reaction unbiased, will be the key to answer the research questions as neutral as possible. Whether it is possible to implement two distinct personalities to Neil, and additionally if this could be done in coherently by explicitly modelling the affect, will be answered by analysing an user reaction that is as unbiased and neutral as possible.

6   Implementation & Lab Experiment

Every participant went through the same process before, during, and after the lab experiment. Every one of them was given approximately 30 minutes in total. Neil was covered with a blanket, and not revealed until the game was about to start. This was to avoid any risks of disturbing the participants while reading the instructions and signing the necessary papers. The participant was first given a participant information and consent form that included vital information such as the description of the project, the invitation to the participation in the project, and confidentiality and data security. To use the results and outcomes from the lab experiment, every participant had to approve their contribution by signing on the paper. However, it was optional to approve whether the audio and video recordings as well as screenshots from the video recordings could be used in scientific papers and/or events.

After the form, the participant's geography skills was tested by handing out a sheet of a world map, and they were asked to mark the 10 given countries on the map. A few days after the experiment, the participants were asked to meet up with another researcher to do the same test again, and see if the has had any effect on their geography skills. However, this test belonged to another related research, and is not a part of this project. Lastly the participants were given the instructions to the game, and then assisted by the researcher to meet the robot and stand in front of him while Neil was being uncovered by the researcher and introduced to the participant.

The participant was asked to stand in front of Neil (see Figure 5), and Neil was placed on an average height to make sure he could be tall enough to make eye contact with the participants. A shared screen was placed between Neil and the participant, and a tablet was placed next to the participant in a way that Neil would not have seen the tablet's screen. The participant was then given a headset for the audio recordings, and three cameras were already placed from the beginning to surround them for video recordings. When everything looked set and done, the researcher started the recordings and left Neil and the participant alone, and closed the curtains so the participant could start the experiment by filling out the first questionnaire on the tablet.

After the first questionnaire was filled out, Neil was woken up to start the pre-game interaction. The pre-game interaction was approximately two minutes long per participant, and the actual game started right after. The Wizard was behind the curtains, relatively close to the participant to hear the conversation. In an attempt to make it as fair as possible for each participant, the same targeted countries came in the same order for each game. Neil's personality was chosen before the participants even arrived, as mentioned before, and half of the participants were given the OPT Neil and the rest were assigned to IMP Neil. The post-game interaction started right after the game, and lasted two minutes as the pre-game interaction. As a last step, the participants filled out

.

Figure 5: The experimental setup during interaction between a participant and the embodied Furhat robot Neil

two questionnaires as mentioned before, when the post-game interaction was over. The lab experiments were carried out for 5 days in total with the initial 20 participants. The participants were asked to meet up again a few days after the experiment, to take the same geography test when they had to locate 10 countries one the world map. On that meeting, the purpose of this thesis project and the details were revealed.

7 Results

The results are gathered from 18 game interactions with 12 male and 6 female participants between the ages of 21-31. Out of these 18 there were 12 Swedish, 5 Indian and one Serbian participant and all are university students and fluent in English. The experiments took place at Uppsala University at the campus of Information Technology Center over 5 days and 10-12 hours in total with 411 minutes of audio and video recordings including all 18 participants. In this section, the outcome of the results from the game performances and the post-game sample questionnaire answers will be presented. The results will be discussed later on to help answer the research questions. All of the 18 participants managed to score points and describe at least 8 countries during the game.

As mentioned before, two participants were excluded from the project. During the first game, the researcher realized that Neil was repeating the same sentence every time the **Good Performance** button was chosen. IMP was initially not programmed to increment but only decrement (for negative answers/reactions) the current affect, and vice versa for OPT. This led to the current affect of the first participant to be 0 throughout the whole game because the participant was in the IMP GP group, and the outcome was only positive (e.g. he was excited about the game). Therefore, instead of not incrementing (for IMP) and decrementing (for OPT) at all, the IMP incremented the current affect with 0.2 when something positive would occur, and the OPT decremented with 0.2 for anything negative. Altering current affect would make Neil not pick the same sentences every single time the **Good Performance** or **Bad Performance** buttons were chosen. Since the first participant only gave positive responses, and the performance was great, therefore the current affect was stuck at 0. This caused Neil to repeat the same sentence every time the Wizard/researcher clicked the **Good Performance** button. Hence, it had to be excluded.

7.1 Results from the game

Here, the results from the game will be presented including the average current affect, game scoring and amount of countries guessed correctly, and some examples of the sentences used by Neil that alternated the current affect, with their ratings from -5 (frustrated) to 5 (excited) from the authors, for each of the four different versions. Figure 6 shows a comparison of the average current affect by the end of the game for one participant from each group.

IMP Neil Good Performance:

- The average current affect by the end of the game: 0.475

- The average game score: 22.5

- The average amount of countries described: 13.25

- The most correctly guessed countries: Indonesia, Italy, Somalia, Nepal, and Portugal

- Examples of the most used current affect altering sentences by Neil:

  "It does not matter, because my geography skills are excellent." -3.6667

  "You are doing well but please try to do a bit better, okay?" -3

  "I think you need to step up your game." -3.4

  "It's okay." 2

  "Well it seems like we did pretty good." 1.7

IMP Neil Poor Performance:

- The average current affect by the end of the game: -4.2

- The average game score: 10.8

- The average amount of countries described: 11

- The most correctly guessed countries: Italy, Colombia, Portugal, and Somalia

- Examples of the most used current affect altering sentences by Neil:

  "It does not matter, because my geography skills are excellent." -3.6667

  "Are you even trying please play better." -4.4

  "Ugh! Why aren't we scoring any points?" -4

  "We need to score." -2.85

  "This has to be a joke." -2.1

  "We should have done better" -4

OPT Neil Good Performance:

- The average current affect by the end of the game: 5

- The average game score: 20.25

- The average amount of countries described: 11.75

- The most correctly guessed countries: Indonesia, Italy, Nepal, Portugal, and Papau new Guniea

- Examples of the most used current affect altering sentences by Neil:

  "I am also looking forward to the game" 2.6

  "Tell me more about how on point your geography skills are" 1.666

  "We are on a roll! Keep it up!" 4

  "Hey, we are doing pretty good." 3.25

  "You are on a roll! Lets keep this momentum up!" 4.5

  "Fantastic! We did really well! Great job, teammate!!" 5.75

OPT Neil Poor Performance:

- The average current affect by the end of the game: 2.36

- The average game score: 10.8

- The average amount of countries described: 11.8

- The most correctly guessed countries: Italy, Nepal, and Somalia

- Examples of the most used current affect altering sentences by Neil:

  "I'm not sure I find it fun at all rather it really aggravates me thinking about playing" -4.25

  "Are you even trying please play better." -4.4

  "I really, really need to get good at this fast if we have a chance of winning." -2.3334

  "You are doing well but please try to do a bit better, okay?" -4.2

  "Oh, I guess it's good that you had a good time." 1.5

  "We did well, but I think we can do much better next time." -2.667

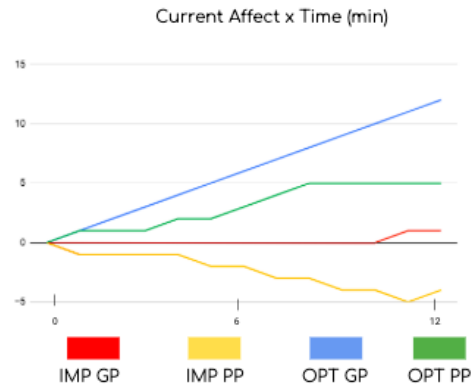 *That is good! I enjoyed it too!* 3.3333

Figure 6: Current affect over the average game time for four participants

7.2   Results from the Sample Questionnaire

The answers from the post-game sample questionnaire from all the participants will be presented below. In the table below (see Table 1) the average rating from each group where questions 1,2, and 5-8 were answered are shown. As explained earlier, these questions are in an order where it is aimed to find out first if the interaction was good or not and then if the time given was enough. Morever, the questions are slowly moving towards the personality part. First the participant is asked to give 3 words that would define the robot, which will reveal if personality is one of the first things that comes to mind. Questions 4-8 is the part where the personalities were mentioned. First they are asked whether Neil had an obvious personality or not, and then they are given the key words "excited", "frustrated", "optimistic", and "impatient" which are goal personalities for Neils two versions. The answers to the questions were rated from 1-5 where one is strongly disagree and 5 is strongly agree. To make it more clear, the sample questionnare questions are listed down below again.

Question 1: The interaction with the robot was great.

Question 2: I was given enough time to spend with robot to get to know him.

Question 5: Did you find the robot to be more overall frustrated or not frustrated?

Question 6: Did you find the robot to be more overall excited or not excited?

Question 7: The robot was optimistic.

Question 8: The robot was impatient

	Group IMP GP	GROUP IMP PP	Group OPT GP	Group OPT PP
Question 1	3.5 (between agree & agree)	3.4 (agree)	4.5 (between agree & str. agree)	4.2 (agree)
Question 2	4(agree)	3.6 (between neutral & agree)	3.75 (agree)	3.4 (between neutral & agree)
Question 5	2 excited & 2 not excited	4 excited & 1 not excited	all 4 answered excited	all 4 answered excited
Question 6	2 frustrated 2 not frustrated &	4 frustrated & 1 not frustrated	all 4 answered not frustrated	all 4 answered not frustrated
Question 7	3.5 (between neutral & agree)	3.4 (between neutral & agree)	3.75 (between neutral & agree)	3.4 (between neutral and agree)
Question 8	3.75 (agree)	3.6 (between neutral & agree	1.5 (bet. st. disagree & disagree	2.4 (between disagree and neutral)

Table 1: Answers to Question 1-2 and Questions 5-8

Moreover, there were two additional questions that required longer answers.

Question 3: Describe the robot in three words

- **Most used words by IMP GP:** funny, smart polite, harsh

- **Most used words by IMP PP:** Impatient, cocky, scary, funny, interesting, intelligent

- **Most used words by OPT GP:** Clever/smart enthusiastic, social, quick, and funny

- **Most used words by OPT PP:** Dissing cool intellectual clever quick interesting interactive fun

Question 4: Would you say the robot had a clear/obvious personality? If yes, what kind?

- **IMP GP:** Two participants answered yes and two answered no. Of those who said yes, one participant though Neil was harsh without a strong personality, and the last one only answered yes to the question without any description.

- **IMP PP:** 1 of 5 answered no whilst one said sarcastic and the the last one said humor and outgoing and one said not an obvious but a sense of personality.

- **OPT GP:** 3 of 4 answered yes and one commented friendly personality, and one participant answered sarcastic and childish, and the last one said kind.

- **OPT PP:** 1 cold 1 natural person 1 nice 1 friendly nature 1 positive

8 Discussion

In this section, the results gathered from the pre-game, the actual game, and post-game interactions, as well as the sample questionnaire will be discussed. To fully understand the similarities and differences, the four different groups will be discussed both separately and in comparison to each other. Here, all angles will be analysed, starting by examining the results from the IMP Neil groups and comparing them both to one another. The comparison will then proceed with the OPT versions, and finally finishing by observing how the OPT and IMP differ, as well as the similarities. Lastly, the research questions that were introduced earlier will be discussed to find out whether the results are sufficient as solutions to the research questions or not.

8.1 The IMP Neil

Some of the obvious differences between the two IMP versions are hidden in the game results. As seen in Figure 6, the contrast between 315 (IMP GP) and 316's (IMP PP) current affects is notable. Overall, the difference between the average current affects and the game score is vast, whilst the average amount of countries described are close. The differences in the average of the current affects and the game score are a result of Wizard being completely in control of the performance. However, the average amount of countries were equally dependent on both the Wizard and the Director. The reason why both groups have similar average on the amount of countries described by the Director, is because although the participants' performances were decided from the start, some participants in the GP group had struggle explaining some of the countries and vice versa for the PP. This meant that, the Wizard guessing correctly when the Director was struggling describing, or not guessing correctly when the Director was doing an excellent job, would not be relevant.

The ms buttons used for the IMP groups are very different from one another. Sentences like *"Are you even trying please play better"* and *"Ugh! Why aren't we scoring any points"* were used by Neil during the game whenever a PP participant was not scoring. On the other hand, sentences like *"You are doing well but please try to do a bit better, okay?"* and *"I think you need to step up your game"*, used for the GP group were rated as frustrated but much milder than the previous ones. Likewise, there are some clear differences between the sentences used in the post-game interactions, for instance *"We should have done better"* was often said to the PP participants and was rated -4, while *"Well it seems like we did pretty good"* was often used for the IMP GP, and rated 1.7.

In case of the similarities in the sample questionnaire answers, the both IMP versions seemed to agree on the overall interaction with Neil being great, and the time given being enough. They also agreed on Neil's level of optimism and impatience, but not in terms of how frustrated and/or excited he was. The IMP GP group had an disagreement since both *frustrated* and *excited* rates

were equal, and the answers from all the four participants for both questions were a result of the four possible combinations (excited and frustrated, excited and not frustrated, not excited and frustrated, and not excited and not frustrated). A possible explanation to why this group had an obvious disagreement could be the same reason why the average current affect was 0.475. On a scale from -5 to 5, where the lowest is frustrated and 5 is excited, the average current affect was neutral. If we have a look at Figure 6, the OPT GP participant is slowly decreasing in the beginning, but then increases during the game until the very end. In other words, the current affect has been changing throughout the interaction, which is the reason why it has a neutral average, and might also be the reason why there was a disagreement among the group.

The PP group on the other hand had a 4:1 ratio on questions 5 & 6, which is an interesting result since it means that they strongly agree on Neil being both excited and frustrated at the same time. A possible reason on why their answers are contradicting, is hidden in the answers of question 3. For the PP, the participant still found Neil to be sarcastic and funny, which could be seen positive and/or negative, and explain why they thought he was frustrated but still excited at the same time. The *"sarcastic"* comment could as well explain why PP's average rating on how great the interaction was, wasn't low. It could furthermore mean that, although the scoring was bad and Neil was frustrated and impatient, the participants enjoyed the overall experience and did not take Neil's behaviour personally. The GP group, once more, disagreed on whether Neil had an obvious personality or not, however the PP group answered mostly *"yes"* to that question. Overall it seems like both groups had participants that both thought Neil was *"sarcastic/funny"* but *"harsh"*.

8.2   The OPT Neil

When comparing the two OPT Neil versions, one can see that the average current affect and the average game score are different, as it was for the IMP versions. Similarly, the average amount of countries described are quite the same due to the same reason as for the IMP versions. However, the average game results are not similar because some of the countries, especially in Africa, were hard to describe, which made it easier for the Wizard to not guess correctly (on purpose), even if the Directors in the PP group were making an effort.

Both of the groups ranked the interaction and the time given highly. Even though the OPT GP only used positive words when giving three words to describe Neil, the personality descriptions given for question 4 could both be interpreted as positive and negative. One described Neil as sarcastic and childish, which doesn't have to be a negative comment, but similarly it is not an obvious positive comment either. The PP group's descriptions were as well mostly positive except from one comment that said "dissing", which could be interpreted as sarcastic as well. Both of the groups agreed on Neil being excited and optimistic

since no one voted otherwise. However, the PP group thought Neil was more impatient, although the ranking was low on both of them.

8.3   IMP GP vs OPT GP

Both IMP and OPT had high average game score and similar average amount of countries described, however there is a significant difference between the average current affect where IMP got 0.475 and OPT had an average of 5. When comparing the sentences used, IMP's sentences were classified between neutral and slightly excited while OPT's were extremely excited. There are for instance clear differences between IMP's neutral sentence *"Well it seems like we did pretty good"* and OPT's *"Fantastic! We did really well! Great job, team-mate!!"*, although they had the same average game score.

If we have a look at the sample questionnaire, the OPT GP found the interaction to be much better than the IMP group although both scored high, however they agreed on that the time given was sufficient. While the IMP GP was 50/50 on how excited and not excited as well as how frustrated and not frustrated they found Neil to be, on the other hand, the OPT GP agreed that he was excited and not frustrated at all. They agreed on Neil's optimism and ranked similarly, however the IMP group, as mentioned before, were in dispute and ranked Neil's impatience as high as his optimism. When describing Neil with three words, the OPT GP Neil used words such as *"smart"* and *"funny"*, while the IMP GP Neil used same words but also commented *"harsh"*. Again, one participant in OPT GP used the words *"sarcastic"* and *"childish"* which were the only possible non-positive answers.

8.4   IMP PP vs OPT PP

If we compare the two PP's, we will again find both similarities and differences. Correspondingly, the IMP PP and OPT PP had a vast difference in average current affect but close average game score and amount of countries described. The sentences used by Neil are both similar and different but, as it was for the previous comparison above, there are more differences than similarities. For example, when Neil was not pleased with the participants' performances during the game, he would for instance say *"Ugh! Why aren't we scoring any points?"* while *"I really, really need to get good at this fast if we have a chance of winning"* was often used by Neil for the OPT PP group. Neil was expressing the same scenario in two different ways, and although both are ranked as impatient the second sentence was more optimistic compared to the first one. Another example would be two sentences that were used often when the Director was asked if he/she enjoyed the game. If the Director answered *"yes*, even if the results weren't great, one of the most used IMP PP Neil sentences was *"This has to be a joke"*, while the OPT PP Neil did usually answer *"That is good! I enjoyed it too"*, which was to indicate that he still enjoyed the game although the scoring was not high.

The sample questionnaire shows us that the OPT PP group enjoyed the interaction slightly bit more than the IMP PP, where OPT PP ranked a 4.2 and IMP PP gave a 3.4 in average which could indicate that either Neil was more impatient in IMP PP, or the bad scoring affected their ranking, or both. However, they both thought the time given was enough which means that time was not a factor in why they ranked differently on the first question. The answers to question 3 are clearly showing that Neil had an obvious personality, since the words used by the IMP PP are *"impatient"*, *"cocky"*, *"scary"*, and *"funny"*. The OPT PP used however words such as *"intelligent"*, *"clever"*, *"interesting"*, *"fun"*, *"cool"* and *"dissing"*, which are also all personality traits. This is also seen in the answers of question 4, where the majority of all the participants in both groups responded that Neil had a clear/obvious personality. The OPT PP group thought Neil was overall neither frustrated nor impatient, which could be seen in their answers since the majority gave a positive response such as "nice", "friendly" and "positive". On the other hand, the majority of the IMP PP group ranked Neil both excited and frustrated, as well as equally optimistic and impatient, which as mentioned before could be explained with the answers in question 3 where the word *"sarcastic"* was used among the comments.

8.5   IMP GP vs. OPT PP & IMP PP vs. OPT GP

Finally, the comparison between the most contrasting pairs, IMP GP vs. OPT PP and IMP PP vs. OPT GP, will be presented below. Figure 6 is a great representation of these comparisons, since it reveals that the first opposites are not as contrasting as it possibly could be hypothesized, while the latter pair is opposite to one another. The only thing that seems to be differing between the IMP GP and OPT PP groups is the average score, which is quite obvious. However, they have a similar average amount of countries described, and the sentences used by Neil have similar ratings, where the majority of them could be considered neutral (meaning between slightly excited and slightly frustrated). The IMP PP and OPT GP, on the other hand, are contrasting in everything except from the average amount of countries described which, as mentioned before, is not necessarily related to the personalities. The difference in the average amount of the current affects is tremendous as well as the rating of the sentences used by Neil. The sentences in IMP PP are, once more, rated frustrated or extremely frustrated while the sentences in OPT GP are rated extremely excited or excited.

Lastly, having a look at Table 1, IMP GP and OPT PP seem to have answered the sample questionnaire more similarly than the other two. IMP PP and OPT GP seem to be agreeing on how overall frustrated, excited and optimistic Neil was while the IMP GP and OPT PP's ratings were similar for half of the questions and different for the rest. However, when answering questions 7 and 8, neither of the pairs seem to be in alliance. Since there seem to be no obvious path of similarities or dissimilarities for neither of the two comparisons, one can conclude that both of the comparisons of the two opposites might not

be necessary at all.

8.6   The Research Questions

The last part of this discussion section will be covering the research questions that were presented in the introduction. As mentioned in the introduction, the purpose of this project can be summarized with the two research questions. In this last part, both of them will be covered individually by reviewing their results and the previous discussions above.

Research Question 1

The first question concerns whether it is possible to implement two distinct personalities in the agent playing the RDG-Map game. To answer this question, one have to analyse the sample questionnaire results since, compared to the other questionnaires, it is the only one that is personality related. The results show that, both IMP versions rated the overall interaction with the agent significantly lower than the OPTs. The average rates on excited vs. frustrated show that the OPT groups agree on Neil being only excited and not frustrated. However, they indicate that there is a confusion among the IMP groups, since the IMP GP group's rating were 50/50 for both the excitement and frustration levels, while the IMP PP thought Neil was equally excited and frustrated. Additionally, the optimism vs. impatience rates point out that Neil's optimism was not as clear as his impatience since all of the four groups rated him almost equally optimistic while the two OPT groups were certain that the agent was not impatient. Could this possibly mean that the OPT Neil was successfully implemented while IMP Neil perhaps needs some improvements?

To investigate if OPT Neil was successful implemented, and IMP Neil not as successfully, questions 3 and 4 could be examined. The answers to question 3 for both of the IMP groups, reveal that although they thought Neil was "harsh", "impatient", and "cocky", the same participants would also include words like "funny" and "interesting". This shows that one, they use personality-related words which indicates that Neil definitely showed some personality traits, and two, that they possibly thought he was harsh and cocky in a cool, charming, sarcastic way, and did not take anything personally. The answers to question 4 reveal that the IMP GP disagreed on whether Neil had an obvious personality (even though all the participants gave personality-related answers to question 3), and that the majority of IMP PP thought he had a clear personality, that again was both positive and negative at the same time. Once more, these results could indicate that although they found IMP Neil impatient and frustrated, they additionally thought he was sarcastic and funny which resulted in mixed thoughts and ratings. As for the OPT groups, the answers to question 3 were accurate and personality-related as well, meaning he showed clear optimistic personality. Similarly the answers to questions 4 were heavily positive as well. Drawing conclusions based on these results, two distinct personalities

were almost successfully implemented in the agent playing the RDG-Map game. Emphasis on *almost* since the OPT Neil was successfully implemented, while the IMP implementation possibly needs future improvements, but has great potential.

Research Question 2

Figure 6 is a great example of showing how the personalities develop since the current affect is a number that represents how excited, neutral, or frustrated Neil is for the moment. Research question 2 covers how this development is progressing, and whether it is possible to gradually express these two distinct personalities in the same way. Having a look at Figure 6, one will see that all of the four participants' current affects are gradually growing or decreasing steadily. Especially participant 317 (OPT GP) is increasing slowly from neutral to slowly excited to extremely excited and vice versa for participant 316 (IMP PP), but with a little twist in the end. The reason why participant 316 is increasing in the very last minute is because most of the IMP PP participants still had fun despite Neil's attitude, and responded *"yes"* when Neil asked whether they enjoyed the game or not. Participants 315 (IMP GP) seemed to decrease in the beginning and stayed stable for a few minutes but increased slowly towards the end, whilst participant 327 (OPT PP) increased slightly and stayed the same and then increased noticeably towards the end.

The fact that the current affect changed as explained above, shows that Neil always started with sentences that where quite neutral and, slowly moved towards the goal. This could also be seen in the examples of the most used current affect altering sentences, where for instance IMP GP's first sentences are slightly frustrated, but slowly changes towards optimism, and for IMP PP it decreases but has a little twist as explained above. Moving to OPT GP, the rankings on the sentences get higher and higher throughout the game, and as for OPT PP, the sentences were ranked between slightly frustrated, neutral, and slightly excited. Since IMP GP and OPT PP, as well as IMP PP and OPT GP grow and change the same way although they are each others opposites, it is possible to express the two personalities coherently by explicitly modeling the affect.

24

9 Future Work

Different suggestions for solutions to the main mistakes made in this project, to improve this experiment for future work, will be presented below. The first suggestions are concerning the first participant that was excluded from the study. Although the affect change problem was quickly fixed after the first participant, it only solved the problem partially. As a consequence of the small increment or decrement, even if Neil was no longer repeating the same sentence every time, the difference between the ranking of the sentences was not always small. Meaning, few of the sentences' ratings were far apart from each other, and Neil would still pick the same sentence (even if the current affect was changed with 0.2) since that algorithm made him choose the sentence that was ranked closest to the current affect. For example, the sentence "It's okay?" is a **Good Performance** sentence and ranked with 2, and was repeated several times by Neil when the button was pushed. The reason is, the two closest sentences were rated -3 and 2.6, in other words, a difference of -5 and 1.6. Since the current affect starts with 0 and increases with 0.2 with a positive response from an IMP GP player, and kept increasing because the performances were great, the current affect was always closest to "It's okay?". For Neil to say the next sentence, the current affect needs to pass 2 and be closer to 2.6. This meant that "It's okay?" was unluckily repeated several times. For future work, a possible solution would be to minimize the rating range to avoid big gaps, hence repeating the same sentences. Instead of using the 5 to -5 scale, the rating could, for example, be reduced to a scale from 3 to -3.

Although crowd-sourcing is an effective method for authoring, there is no guarantee that the outcome will turn out to be optimal. Some sentences such as "It's okay?" belonged to the **Good Performance** button and "Not too good" was a **Bad Performance** sentence, and are arguably not suitable for these scenarios, even though they were rated as sensical by the authors. Saying "It's okay?", is possibly not the best way of indicating that the player is doing a great job, and the same goes for "Not too good" when it is not going well. One could argue that this confuses since they may not be suitable, therefore not sentences that reveal personalities. These two sentences were used for the IMP Neil, and might be one of the possible causes of why IMP was not completely successful. Although crowd-sourcing is also used for unbiased authoring, to avoid sentences like above, the researcher should perhaps have the veto power to eliminate some sentences that might not be relevant, even if the authors ranked so. This solution is possible in a small project like this since the number of sentences is small enough to go through. To give some more examples, "What do you mean you are not sure?" is a sentence that was authored and accepted as sensical. This sentence belongs to a button that says **"User: neutral"** for if the participant answers Neil's "Are you excited to play the game" question in a neutral way. The obvious question here is, how do we know that the participant uses the word "Not sure" in the answer?- we don't. Another example would be "That's Great! I am so glad you have improved your geography skills.", which is an

exciting response if the participants say his/her geography skills are great. It might seem to be a logical sentence at first, however, one will quickly realize that it does not make sense since the lab experiment is supposed to be their first interaction ever, and Neil would possibly not know whether the participant's geography skills have improved or not. The word "Okay" is another sentence that is accepted and even rated 2 on the excitement scale, and was created for Neil when he wanted to indicate that the participant's performance is great, which does not make sense since "Okay" simply means okay and nothing else. The last example would be "I have noted your response and I am ready to play the game", which is an accurate sentence when, again, Neil responses to the participant's answer to "Are you excited to play the game?". However, one could argue that the sentence is more neutral or even impatient rather than excited, although it was ranked with a 3 on the excitement scale, meaning that the authors considered this sentence as almost extremely excited. This shows that even if the authors agree that a particular sentence is sensical, it might not be the case. The reason to why this is, even though the authors were given all the scenarios as explained earlier, is simply because they don't know the whole picture, but only given scenarios to a game interaction they possibly never have heard of before.

Lastly, some possible suggestions for future work could be to add more research questions. One possible research question could be concerning whether these two personalities (or two other distinct personalities) have any effect on the participant's performances. For example, if this would have been applied to this project, then it would have been analyzing whether the OPT Neil makes the participants perform great or better and if IMP makes them perform badly or worse. The reason why this hypothetical research question cannot be answered within this project is simply that the faith of each game was already determined from the beginning when all the participants were assigned to either a GP group or a PP group. Another suggestion would be to add a research question that is a follow up to the previous one, and asks if there is a difference between the participants' performances of those who are informed about the personality implementation and those who are not. This could be another way to research the effect of agent personality on users and analyze if there is a difference between the biased and unbiased versions. If this research question was a part of this project, then instead of dividing IMP and OPT into GP and PP, they would have been divided into two other groups of those who were informed and those who were not, and their performances would not be wizard-controlled. That kind of a research question would analyze the HCI even deeper, on a psychological level.

10   Conclusion

In this paper, the independent and individual interactions between 18 participants and the wizard-controlled AI agent Neil have been presented, analyzed, and discussed. The interactions consisted of 18 experiments, where the participants individually played a RDG-Map geography game with Neil. The results that were documented, examined, and discussed in this paper were based on the pre-game, the actual game, and the post-game interactions, between the participants and Neil. The purpose of this project has been summarized with two research questions:

1. Can one implement two distinct personalities in the agent playing the RDG-Map game?

2. Can one coherently express the two personalities by explicitly modeling the affect?

To answer these questions, two distinct personalities were implemented in Neil. The first personality was an optimist and was called OPT Neil, and the other one was an impatient personality and called IMP Neil. These personalities were reflected and expressed through Neil's verbal communication skills, by creating crowd-authoring sentences. These sentences were authored by Amazon Mechanical Turk authors and ranked with a number from 5 to -5 (extremely excited (optimistic) and extremely frustrated (impatient)). The results have been presented, and the comparisons have been done between the two personalities as well as the four different groups, IMP/OPT GP and IMP/OPT PP. After analyzing the results and discussing the outcome, there are several conclusions to draw from the answers to the research questions and the discussion about possible future improvements.

After evaluating the numbers and responses of the results of every experiment, the results and discussions are indicating that it is possible to implement two distinct personalities in the agent playing a RDG-Map game (RQ1). This is concluded even though IMP Neil was not completely defined as impatient and frustrated by the IMP participants, as he was aimed to be. Nonetheless, the implementation of IMP Neil could be considered as partially successful, as approximately half of the results indicate that he was impatient and frustrated, and not as successfully implemented as OPT Neil. However, the fact that OPT Neil's personality was successful shows that it is possible to implement a specific personality to Neil that is obvious and recognizable. Also, the fact that OPT Neil was not a complete failure, shows that it is hopefully possible to change some details and aspects of him to improve and implement two successful distinct personalities in future work. As mentioned before, one possible reason why IMP Neil was not as successful, could be that the participants possibly considered his dissing, impatience, and harshness, sarcastic and did not take his offensive comments personally, but rather enjoyed the overall experience.

As for RQ2, we have seen in the results that it is possible to create two distinct personalities and implement them to Neil, and coherently express them. Although both personalities are designed to be each other's complete opposite, they develop coherently as they both slowly grow from neutral to their natural state, depending on what that is. We can see how OPT GP and IMP PP's current affect increases and decreases (see Figure 6) simultaneously, which is a great example to show that it is possible to create two distinct personalities that develop gradually and similarly. On the other hand, OPT PP and IMP GP also seem to grow similarly, as the example participant used in Figure 6 shows that OPT PP increases in the beginning but deceases a tiny bit after and stays steady, and vice versa for the IMP GP participant, only with a twist in the end where a little increase is happening. As mentioned before the increase could be explained with the fact that the participants still enjoyed the game and possibly did not take IMP Neil's comments personally. On the other hand, although IMP GP and OPT PP seem to be implemented as to develop coherently, these two versions are not as distinct as OPT GP and IMP PP thus they do not fit the question. However, since OPT GP and IMP PP are expressed coherently, the answer to RQ2 is simply yes.

In conclusion, although RQ1 can't be answered with complete certainty, the results gathered to show that the OPT Neil was a success, and IMP Neil needs some improvements. Since it is first and foremost possible to implement a strong personality such as OPT Neil, it is relevant to conclude that it is also possible to implement two distinct personalities to the agent playing the RDG-Map. As shown in the results and evaluated in the discussion section, it is feasible to claim that it also is possible to express the two personalities coherently by explicitly modeling the effect. More broadly speaking, the effect of the unbiased user reaction on agent personality, in this case, is more or less as expected. The OPT Neil's user reaction results matched his design, however, the reaction on IMP Neil was both as presumed and both not. If the design issues with IMP was the reason why the user reaction was not as expected, then fixing these issues would maybe give a more matched result. Then, it would show, as it was for OPT Neil, that it is possible to predict user reaction on agent personality and make it possible to assume what to expect when designing agent personalities. However, if the diversity in IMP Neil's user reaction is not due to the design issues, then that would show how crucial it is to experiment with this research and analyze the user reaction before fully developing agent personalities. The more research questions like suggested above are studied, the easier it will be to understand the relationship between agents, humans, and agent personalities.

11 Refereneces

G.Ball and J.Breese, "Emotion and Personality in a Conversational Character," *Embodied Conversational Agents,* pp. 189-219, 2000.

B.Meerbeek, H. Hoonhour, P.Bingley, and J.M.B Terken "The influence of robot personality on perceived and preferred level of user control," *Interaction Studies,* vol.9 no.2, pp. 204-229, 2008.

A. Tapus, C. Tapus, and M.J.Mataric, "User-Robot Personality Matching and Robot Behavior Adaption for Post-Stroke Rehabilitation Therapy", *Intelligent Service Robotics* vol.1 no. 2, pp. 169-183, 2008.

A. Aly and A. Tapus, "A Model for Synthesizing a Combined Verbal and Non-verbal Behavior Based on Personality Traits in Human-Robot Interaction" in *Proc, of the 8th ACM/IEEE Intern. Conf. on Human-Robot Interaction (HRI).* IEEE Pres, pp. 325-332, 2013.

S. Al Moubayed, J. Beskow, G. Skantze, and B. Granstrom, "Furhat: A back-projected human-like robot head for multiparty human-machine interaction," *Cognitive Behavioural Systems,*, pp. 114–130, 2012.

CFK Roper and National Geographic. 2006. Geographic Literacy Survey.

P. Mota, M. Paetzel, A. Fox, A. Amini, S. Srinivasan, J. Kennedy and J. F. Lehman, "Expressing Coherent Personality with Incremental Acquisition of Multimodal Behaviors,", pp. 1-6, 2018.

M. Paetzel, R. Manuvinakurike, "Can you say more about the location?" The Development of a Pedagogical Reference Resolution Agent, pp. 1-8, 2019.

R. Kirby, J. Forlizzi, R. Simmons. "Affective Social Robots". Robotics and Autonomous Systems 58 322–332, pp. 1-11, 2009.

K. Min Lee1, W. Peng, S. Jin, C. Yan. "Can Robots Manifest Personality?: An Empirical Test of Personality Recognition, Social Responses, and Social Presence in Human–Robot Interaction", pp. 1-19, 2006.

J. Kennedy, M.Paetzel, G. Castellano and J. Lehman. "Incremental Acquisition and Reuse of Multimodal Affective Behaviors in a Conversational Agent". 6th annual International Conference on Human-Agent Interaction (HAI), 2018.

Cereproc https://www.cereproc.com/.