Personalized and Adaptive Cognitive Human-Robot Interaction with a Novel Fuzzy Logic Control and Reinforcement Learning-Based Paradigm

Master Thesis Report

Marcel Munster

Personalized and Adaptive Cognitive Human-Robot Interaction with a Novel Fuzzy Logic Control and Reinforcement Learning-Based Paradigm

Master Thesis Report

by

Marcel Munster

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Wednesday May 10, 2023 at 10:00 AM.

Student number:	4467949	
Thesis committee:	Dr. ir. R. van Paassen,	TU Delft, chair
	Dr. A. Jamshidnejad,	TU Delft, supervisor
	Dr. ir. Y. Eisma,	TU Delft, external examiner

TU Delft

Preface

With this thesis I conclude my studies at the TU Delft. I started my thesis research during the Covid-19 pandemic, which has been tough. As I moved back to my hometown, which is a two hour commute to Delft, I had to conduct to the entire research from home. Having to plan a zoom meeting to discuss small details of the research has slowed it down significantly. Furthermore, I started working a full-time job a little over a year ago, which might not have been the best idea. As combining a job with a thesis research is very intense. However, I am proud of what I have achieved and I look back at a really enjoyable time at the TU Delft.

First of all, I would like to thank Dr. Anahita Jamshidnejad. Although the first time we met in person was over two years into the research, she always provided me with valuable insights and solutions when I was stuck on a problem. Furthermore, I would like to thank Dr. Davide Dell'Anna for the discussion and insights on the implementation of simulated patients. Finally, I would like to thank my girlfriend, Maartje, and my parents for keeping me motivated and providing me with a welcome distraction whenever I needed a break.

Marcel Munster
Obdam, April 2023

Contents

I

Research Paper

Personalized and Adaptive Cognitive Human-Robot Interaction with a Novel Fuzzy Logic Control and Reinforcement Learning-Based Paradigm

M.S.W. MUNSTER[1]
[1]Delft University of Technology, 2629 HS Delft, The Netherlands (e-mail: M.S.W.Munster@student.tudelft.nl)

ABSTRACT An aging population puts a pressure on health-care workers working with dementia patients globally. A potential solution is to provide care with Socially Assistive Robots (SARs), i.e. robots who help people through social interaction. However, for effective care these SARs must be able to personalize their behavior to individual patients and adapt this behavior to changes in the patient's preferences. This paper presents the decision making process of a SAR that enables the SAR to personalize its behavior to the personality of the patients and adapt this behavior to their current state-of-mind. The system consists of a Fuzzy Logic Control personalization module, which personalizes the SAR's behavior and a Reinforcement Learning based decision making module together with a Fuzzy Logic Control reward module for the adaption of the personalized behavior. The personalization of the SAR's behavior is assessed by comparing the output of the system with answers from a survey. The average scatter index over all different behavioral parameters of the SAR is 20.4%. The adaptation of the behavior is assessed with computer-based simulations, where an overall accuracy of 81.8% is achieved. A third experiment is carried out to assess the effect of adding the Fuzzy Logic Control personalization module to the system. This experiment shows that adding the personalization module to the decision making system of the SAR decreases the time for the learning process to converge with 13.3%. Although the first assessments of the system look promising, more extensive experiments should be held in later stages of the research. A crucial experiment that must be held in future research is performing real-life interactions between dementia patients and the SAR, in such an experiment the functionality of the system really can be assessed.

INDEX TERMS Socially Assistive Robots, Personalization, Adaptation, Reinforcement Learning, Fuzzy Logic Control, Dementia Care

I. INTRODUCTION

The world population is aging, meaning that the percentage of people over 65 is growing [1]. Consequently, there is a rise in the number of people suffering from dementia. Ferri et al. in [2] estimate the global number of dementia patients to double every 20 years, with a total rising up to 81 million by 2040. Providing personalized care to dementia patients is a demanding task for health care workers. An aging society and increasing number of dementia patients confine care-givers from giving personalized one-on-one care. Mierlo et al. in [3] show that personalized care positively affects the well-being of dementia patients, whereas lack of personalized care reduces the quality of life of these patients.

Epp in [4] describes the method of person-centered dementia care to achieve personalized care. In contrast to the traditional culture of dementia care, where the focus lies on the disease and in which individuality is depreciated, person-centered dementia care focuses on the appreciation of individual patients. Care-givers take into account the desires and capabilities of patients to provide care that fits their specific preferred ways of living. Person-centered dementia care improves the quality of life of dementia patients. Furthermore, it reduces agitation among the patients as they are given more freedom in scheduling the activities they do, example given meal time, bed time or activities in the caring home. However, person-centered dementia care is time consuming and expensive [3]. With the current rise in the number of dementia patients, innovative technologies for personalized care are required to reduce the pressure on health-care workers.

An emerging field in robotics called Socially Assistive Robots (SARs), introduces a potential solution for person-centered dementia care despite the growing number of dementia patients. SARs are robots that assist people through social interactions [5]. To date, their main applications are health care, education, and entertainment [6].

SARs have various applications in dementia care. A widely used application of SARs in dementia care is assistance during therapy sessions. Animal-like robots are given to patients in therapy sessions to engage the patient in the session. Marti et al. in [7] use Paro, a seal robot that reacts to touch by moving its tail and eyes. Paro is given to patients during therapy sessions, after which the therapists records the interactions between the patient and robot. The results show that patients seem less stressful when interacting with Paro. Furthermore, Paro makes it easier for the therapists to initiate a conversation with the patients and the patients are more engaged in the therapy sessions.

Moyle et al. in [8] used Paro in a similar way. However, here Paro was not given in a therapeutic setting, instead the patients were free to interact with it as they pleased. Their results stress the importance of person-centered care. Different patients had different responses to Paro. For example, one patient was so engaged with the robot, she started talking to it and hugged it a lot, whereas another patient did not even want to accept Paro when it was handed to her. Moreover, patients responded differently on different occasions where they interacted with the robot. This stresses two crucial considerations when applying SARs in dementia care. First of all, different patients respond differently to the SAR. Therefore, it is necessary to personalize the SAR's behavior to an individual patient, i.e. the SAR creates a behavior that is distinctive for patients with different characteristics. Secondly, individual patients respond differently in different sessions. Therefore, the SAR's behavior must be adaptive to handle changes in the responses of patients, i.e. the SAR must display different behaviors in different scenarios.

Moro et al. in [9] perform research on a SAR that helps people with cognitive impairments in daily life activities and personalize its behavior to the preferences of the patient. The robot is taught behaviors necessary to assist a patient by means of demonstration. Currently, the SAR is only taught to assist patients in the process of tea making, in the future additional activities will be investigated. Example behaviors that the SAR has learned from demonstration are different steps in the process of tea making, correcting patients when an incorrect step is performed, and re-engaging patients when they are distracted. The SAR decides which behavior is required in a certain situation using a decision tree. A second characteristic of the SAR is that it can personalize the behaviors in terms of speech content (assertive versus suggestive) and movement activity. The demonstrations of the behaviors are performed by multiple graduates with distinctive ways of assisting the elderly. Therefore, the demonstrations differ in speech content and movement activity, each demonstration is labeled by its level of assertiveness versus suggestiveness and its level of movement activity. With Reinforcement Learning (RL), the robot learns to select which type of behavior is preferred by an individual patient. The personalization of the SAR behavior is assessed with computer simulations of cognitive models of potential patients. The results of the study show that the SAR was able to display the correct behavior in 93% of the times. In future research, real-life user studies are performed to see whether the robot functions correctly on real patients.

Tapus et al. in [10] performed research on a SAR that can adapt the difficulty of a game in cognitive training for dementia patients. They have designed a SAR to encourage dementia patients during a game called 'Name that tune', where the patients hear songs after which they must click the corresponding button and sing along. The SAR assists the patients by explaining the rules of the game and providing them with hints helping them to click the right button. With these hints, the SAR was able to adapt the difficulty of the game to the performance of the patient, measured by the patient's number of correct answers and their reaction time. Possible hints that the SAR could give are reminding the patients they must press a button and point in the direction of the right button. The results show that the performance of the patients increases over a series of training sessions, indicating improvement in their cognitive abilities. Furthermore, the patients became more engaged over the sessions.

These last two examples show that personalization and adaptation both have been used in SARs for dementia care. However, to the best of our knowledge no research on SARs in dementia care has been looking into a SAR that both personalizes its behavior to individual patients and adapts this behavior to changes in the preferences of this patient in different situations. Likewise, research on personalized and adaptive SAR behavior in fields other than dementia care also is limited to either personalization or adaptation, not both.

Clabaugh et al. in [11] developed a SAR system to assist children with autism spectrum disorder (ASD) in a set of mathematical games. The system personalized difficulty of the games and hints provided to complete the games to the mathematical competence of the children. The aim of the system was to develop the mathematical skills of the children by finding the right difficulty for each individual child, where they were sufficiently challenged but still engaged in the game. The results show that the children were more engaged when interacting with the system when it was personalized to their level of challenge and feedback. Hence, the children improved their mathematical skills significantly.

Tsiakas et al. in [12] discuss the developmental process of a system with a similar aim as [11]. As part of cognitive training, for children with learning issues for example, the difficulty and type of feedback in a sequencing task is personalized to the skill of the participant. Based on real-life data from students, user models are created that describe the success ratio and engagement levels at different difficulty levels of the sequencing task. Policies are learned to select the right task difficulty and type of feedback for these user models. Similar as in [11], the difficulty and the type of feedback should be chosen such that the participant is challenged but does not become disengaged with the task. When the policies are learned for each user model, the difficulty and feedback can than be personalized by selecting the right policy after conducting a skill assessment on the participant. The result of their research is that they came up with a set of user models which contained policies to change the difficulty of the game and the SAR's feedback for different performance and engagement levels of the participant.

Tapus and Mataric in [13] assist patients in post-stroke rehabilitation therapy by motivating the patient through social interaction. The robot personalizes its behavior to the personality of the patient. Participants in the experiment were asked to fill in a questionnaire to determine their level of extroversion. Based on how extrovert or how introvert the participants were, the vocal content, the activity level and the proxemics, i.e. the amount of distance kept between the robot and the participant, could be changed. For example, for a participant showing high levels of extroversion, the robot gave challenging comments with high pitch and volume, showed high levels of activity, and kept less distance towards the participant. Whereas for a more introvert participant the robot gave more nurturing comments on low pitch and volume, showed lower activity levels, and kept more distance to the participant. Every participant was assisted by the robot in two different situations, one where the behavior was matched to the personality of the participant and one where the behavior was selected randomly. The results of the experiment indicated that the participants preferred the behavior that matched their personality over the random behavior.

As the results of [8] show that every dementia patient responds differently to a SAR and that an individual patient responds differently on different occasions, this research focuses on designing the decision making process for a SAR, which allows it to personalize its behavior to an individual patient and adapt this behavior to changes in this patient's preferences. Similar to [13] the SAR personalizes its behavior to the personality of the patient. After personalizing the behavior to the patient's personality, the SAR adapts this behavior to changes in the preferences of the patients based on their current state-of-mind.

II. MAIN CONTRIBUTIONS AND STRUCTURE

Research has been performed on either personalizing SAR behavior or adapting it. However, to the best of our knowledge no research has been performed on both personalizing the behavior to the user as well as adapting it to changing circumstances. This paper presents the decision making process for a SAR that can personalize its behavior to the personality of a dementia patient and adapt it to the current state-of-mind of this patient. In order to do this, two decision making modules are developed: (1) a decision making module relying on Fuzzy Logic Control (FLC) for personalizing the SAR behavior and (2) a RL based decision making module to adapt the personalized behavior to the current state-of-mind of the patient. Moreover, the SARs in the established researches are designed for particular tasks, such as cognitive games. The decision making system introduced in this paper is designed to assist patients in their daily routines.

The paper starts with a description of the designed system where the working principles of the modules of the decision making process are discussed, and the reasoning for choosing the selected control methods is explained. The system description is followed by an assessment of the decision making modules, where experiments on the individual decision making modules and the system as a whole are explained and their results are discussed. Finally, the conclusions of the research are drawn and recommendations for future research are made in the conclusion.

III. SYSTEM DESCRIPTION

As previously mentioned, the problem for the SAR behavior is twofold. First of all, the SAR must personalize its behavior to the personality of the patient. Secondly, the SAR must adapt its behavior to changes in the patient's state-of-mind. The behavior of the SAR focuses on interacting with the patient in their daily routines. Therefore, the behavior of the SAR consists of the following parameters:

- Amount of speech
- Volume of speech
- Gestures
- Interactive comments (energetic versus cautious)
- Motivating comments (cooperative versus challenging
- Feedback (realistic versus nurturing)
- Proxemics

Examples for the comments that the SAR might give are:

- Energetic comment: *'Hey, I've got an amazing idea on what we could do right now. Let's play this real fun board game.'*
- Cautious comment: *'Maybe we can play this board game, it could be fun.'*
- Cooperative comment: *'Let's see if we can extend our walk with one more round around the block. Together we can make it.'*
- Challenging comment: *'I bet you cannot do another round around the block, but feel free to prove me wrong.'*
- Realistic feedback: *'This week the exercises did not go well, try to improve next week as it helps you to stay fit.'*
- Nurturing feedback: *'This week's exercises did not go as well as usual, maybe you were a little tired this time. Next week they'll probably go better again.'*

The proxemics is the distance that the SAR maintains to the patient.

Fig. 1 illustrates the decision making process of the SAR. The decision making system consists out of two modules: (1) an FLC module for the personalization of the SAR's behavior and (2) an RL module for the adaptation of it.

A. FUZZY LOGIC CONTROL PERSONALIZATION MODULE

Personalizing the SAR behavior is required based on the first observation made in [8] which indicates that each dementia patient responds differently to interacting with SARs. Therefore, the personalization side of the decision making

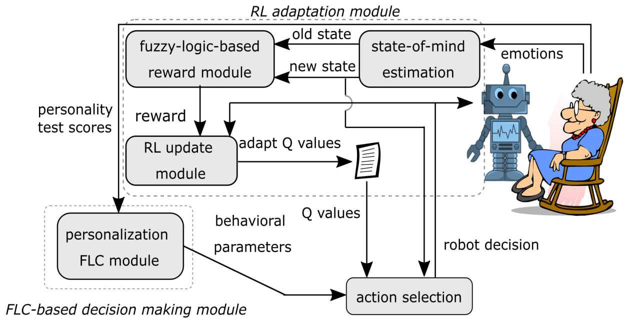

FIGURE 1. Schematic view of the different elements used to develop the proposed decision making module for SARs.

process generates an initial behavior for the SAR that suits the personality of the patient.

The reason to personalize the SAR's behavior to the personality of the patient is because [13] shows that people prefer to interact with a robot that has a personality which resembles their own personality. In their research the target was to motivate patients in post-stroke therapy. Therefore, they focused on personalizing the SAR's vocal content, activity level, and proxemics, which are parameters that are related to the extroversion-introversion personality trait. In our research the focus is on assisting dementia patients via social interaction in daily routines, making other personality traits besides extroversion relevant too [14]. Therefore, this research considers the Big-Five personality traits [15] for the personalization of the SAR's behavior to its patients.

The Big-Five personality is a widely used approach to describe someone's personality and it includes five main traits: Openness to experience, conscientiousness, extroversion, agreeableness, and neuroticism.

Openness to experience describes a person's imagination and acceptance for new experiences. People scoring high on openness are open-minded, imaginative and tolerant. They are curious to try out new things and engage in new ideas. People low on openness are more down-to-earth and conventional.

Conscientiousness is concerned with someone's feeling of responsibility. Conscientious people are more careful, organized and scrupulous. They pay attention to details and aim to finish tasks successfully. People with low levels of conscientiousness are irresponsible and disorganized.

Extroversion measures the level of sociability and excitability. People with high levels of extroversion are more sociable, assertive and active. They pursue excitement and challenge. People scoring low on the extroversion trait are more reserved and cautious.

Agreeableness relates to someone's intentions. Individuals high in agreeableness are modest, cooperative and trustworthy. They are concerned about the feeling of others and show interest in them. People low on agreeableness are suspicious, irritable and competitive. Sometimes people low on agreeableness can even be manipulative.

Neuroticism marks someone's emotional stability. People with high neuroticism are anxious, insecure and depressed. People scoring low on neuroticism are more laid back, emotionally stable and can handle stress very well.

Currently the research is in a preliminary phase and focuses on personalizing and adapting the way in which the SAR interacts with the patients. The traits extroversion, agreeableness and neuroticism are traits that come out during interactions with other people, for example whether someone is talkative or anxious. Whereas the traits openness to experience and conscientiousness are more in line with motivational goals such as stimulation and achievement values, making these more task related [15]. Therefore, only the traits extroversion, agreeableness and neuroticism are included in the system for now. In later stages of the research the SAR should also be able to suggest activities to the patients. Then the latter two traits, openness to experience and conscientiousness, are included in the system too.

Moreover, Tapus and Mataric in [13] mirrored the personality traits of the robot to the personality of the patient. However, research on the influence of the personality of therapists, especially on the traits of agreeableness and neuroticism, does not confirm a positive impact when having congruence between the personalities of clients and patients [16]. Therefore, instead of mirroring the personality of the SAR, knowledge that humans use in their daily interactions is implemented in the personalization module of the SAR.

This is achieved with FLC. FLC is a rule-based method that is comparable to the way humans think. In FLC the rules rely on linguistic terms instead of mathematical rules [17]. The linguistic terms are described with fuzzy sets, which unlike classic sets allow partial membership to the set. The FLC process consists out of three steps: (1) fuzzification, i.e. transforming real-life crisp input values into fuzzy input values; (2) inference, i.e. generating fuzzy output values using the system's fuzzy rule-base and the fuzzy input values; (3) defuzzification, i.e. transforming the fuzzy output values back to real-life crisp output values. The fuzzy rule-base of the inference system consists of a set of *"If ..., then ..."* rules, the first part of these rules (*"If ..."*) is called the antecedent and the second part of the rules (*"then ..."*) is called the consequent.

FLC is selected for personalizing the SAR's behavior for the following main reasons:

1) No mathematical models exist that describe the relation between someone's scores for the Big-Five personality traits and what type of behavior they prefer in social interactions. Human knowledge on these relations does exist, making it possible to capture the relations with linguistic terms in the fuzzy if-then rule base. Example given, Roccas et al. in [15] describes the preferences people generally have when they possess the different personality traits.
2) The personality traits are measured along a spectrum, someone's personality is not defined by having certain traits but by the extent someone displays the traits.

Therefore, interpreting the personality scores as precise crisp values is prone to mistakes. In FLC the personality traits can be treated as fuzzy variables, where the variables can be described with fuzzy sets that allow partial membership.
3) Rule-based approaches based on fuzzy logic, such as FLC, are the closest methodologies to the reasoning and decision making of humans [18], [19]. Therefore, it is assumed that FLC provides the SAR with personalization capabilities as close as possible to the personalization procedure by humans.

Other control techniques, such as Artificial Neural Networks [20] and RL [21] have been considered for the personalization of the SAR behavior as well. However, these methods require a lot of training and for Artificial Neural Networks also a large data set with example relations between a person's personality scores and their behavioral preferences. Collecting such data sets can be very time consuming. FLC only requires the expert knowledge, which readily is available. Therefore, FLC is selected for personalizing the SAR's behavior over Artificial Neural Networks and RL.

Rules are generated for the FLC module to describe the relation between the personality traits of the patient and the behavioral parameters of the SAR. All rules have the following general formulation: "If *personality trait 1* is $A_{1,i}$ and/or *personality trait 2* is $A_{2,j}$ and/or *personality trait 3* is $A_{3,k}$, then *behavioral parameter* is B_m" In this formulation *personality trait 1, 2* and *3* refer to extroversion, agreeableness and neuroticism respectively. $A_{1,i}$, $A_{2,j}$, $A_{3,k}$ for $i = 1, ..., N_{A_1}$, $j = 1, ..., N_{A_2}$, and $k = 1, ..., N_{A_3}$ are fuzzy sets corresponding to the linguistic terms used to categorize the personality traits. Example given, when the personality traits are all categorized as *low, medium,* or *high* then $N_{A_1} = N_{A_2} = N_{A_3} = 3$. Moreover, B_m for $m = 1, ..., N_B$ is a fuzzy set corresponding to the linguistic terms that are used to describe the nature of the behavioral parameters, i.e, amount of speech, volume of speech, gestures, interactive comments, motivating comments, feedback, and proxemics. Here N_B is the number of categories to describe the behavioral parameter.

The rules for the FLC module are inspired by expert knowledge taken from literature. Roccas et. al in [15] give a broad overview of the characteristics that belong to the five personality traits of the Big-Five model. Here the relation between the personality traits and parameters such as amount of speech, volume of speech and the different types of comments are addressed. Hostetter and Potthoff in [22] describe the relation between the personality traits extroversion and neuroticism and the number of gestures a person makes. Tapus and Mataric in [13] investigated the relation between the trait extroversion and several elements of the SAR's behavior. The elements they considered are a SAR's vocal content, activity levels and proxemics. The vocal content included the amount of nurturing versus challenging feedback, pitch, and volume. The activity level was described by the amount of movement. The knowledge from these sources are used to set-up the rules to define the relationship between a patient's personality and the behavior they wish to see from the SAR.

Currently, the focus of the decision making system lays on the social interaction between the patient and the SAR. However, later on in the research the SAR should also be able to suggest activities to keep patients entertained. Which type of activities a person likes to do also depends on their personality. For example, people scoring high on openness to experience like to do more creative activities [23]. Therefore, in later stages of this research, the list of behavioral elements for the SAR behavior is extended with elements that capture types of activities and other activity related behavioral elements, such as spontaneity. The rules to describe the relation between the personality traits and which activities a patient might like will also be inspired on expert knowledge from literature.

The FLC module uses Mamdani inference and the center of gravity method for defuzzification [24]. The literature could not be used to construct the membership functions of the fuzzy sets for the linguistic terms. Therefore, a survey is held to collect data on the relation of people's personality and their preferences on the different behavioral elements. The survey was completely anonymous and has been approved by the Ethics committee of the TU Delft. About two thirds of the responses of this survey are used to optimize the membership functions of the fuzzy rules. The remainder of the responses is used for verification of the FLC personalization module, which is discussed in the results. More on this survey and the construction of the membership function is explained in Section IV-A.

B. REINFORCEMENT LEARNING BASED DECISION MAKING MODULE

The second observation made by [8] was that the same patient responded differently in different sessions with the SAR. Therefore, personalizing the SAR's behavior to an individual patient is not sufficient but the behavior must also be made adaptable to changes in the preference of this patient in different circumstances. As this research focuses on providing care to the dementia patients by means of social interaction and, in the future, suggesting activities, the patient's state-of-mind is considered as the circumstance to adapt the initial behavior to because a person's state-of-mind affects its social interactions [25]. Thus, the SAR changes its behavior to changes in the state-of-mind of its patient. Adapting the behavior of the SAR is achieved with RL. RL is selected for the behavior adaptation as it is able to tackle the adaptivity problem for each individual patient without requiring large amounts of training data.

FLC is suitable to personalize the SAR behavior to the patient's personality as this can be captured with rules to link the behavior to the different personality traits. However, how the SAR should adapt its behavior in a certain state-of-mind is dependent on the preferences of the individual patient. For example, one patient might want the SAR to lower the amount of speech when they feel sad, while another patient

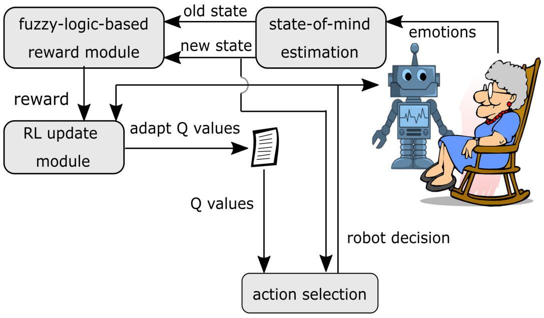

FIGURE 2. Process of the RL based decision making module.

might want the SAR to increase the amount of speech in this situation. Therefore, it is impossible to set up rules that lead to a solution that fits all patients. Methods such as Artificial Neural Networks require a large data set for training the SAR. Collecting such a data set is a time consuming process and is out of the scope of this research. Moreover, the SAR would learn the preferences of the patients included in the data set. In case a patient has totally different preferences than the patients in the data set, the SAR will not be able to adapt its behavior to the wishes of the patient. RL is able to learn to adapt the behavior according to the preferences of each individual patient. Therefore, RL is selected as the method to adapt the behavior of the SAR to the current state-of-mind of the patient.

The RL based decision making module works in a loop which is illustrated in Fig.2. At the start the RL based decision making module receives the personalized values for the behavioral parameters from the FLC module. The RL module explores how to adapt this behavior for different states-of-mind of the patient by interacting with the patient and analyzing how the updated behavior affects the state-of-mind of the patient. The effect of the changed behavior is used to compute a reward (see Section III-B2 for details) with which the SAR can determine which actions are effective and which are not.

For the RL framework, the state s_k is the state-of-mind of the patient and the action a_k is the change in one of the behavioral parameters. The subscript k is used to specify the interaction step. It relies on Q-learning with an ϵ-greedy policy [21] to adapt the initial behavior of the FLC module to the current state-of-mind of the patient. Q-learning assigns a Q-value to each state-action pair, the pair with the highest Q-value is selected as most favorable action in the respective state. In the iterative loop described above the Q-value is updated according to:

$$Q^{updated}(s_k, a_k) = Q(s_k, a_k) + \alpha \left(r(s_k, a_k) + \gamma \max_{a \in A} Q(s_{k+1}, a) - Q(s_k, a_k) \right) \quad (1)$$

In this equation $Q(s_k, a_k)$ is the current Q value for the state-action pair, α is the learning rate of the system, $r(s_k, a_k)$ is the reward that the system receives when selecting action a_k in state s_k, γ is a discount factor and $\max_{a \in A} Q(s_{k+1}, a)$, with A the set of all actions, is the maximum Q-value in state s_{k+1} which is obtained by selecting the optimal action for that state. The SAR determines the next action, i.e. a behavioral element change, according to an ϵ-greedy policy. Here in ϵ% of the cases the SAR chooses a random action, the remaining cases the SAR chooses the action with the highest Q-value.

It is the task of the RL module to find the best action for all combinations of parameters and the current emotional state of the patient. After sufficient interactions, the module should have converged and the best actions of all parameters can be selected to get the optimal robot behavior for the patient in each emotional state.

1) Feedback for the Reinforcement Learning adaptation module

The RL module adapts the SAR's behavior based on the state-of-mind of the patient. However, the state-of-mind is an abstract concept, which refers to the overall cognitive state of a person. Therefore, it is more intuitive and straightforward to provide quantified scores for several basic emotions and to compute the state-of-mind from these emotions.

Retrieving feedback on the actions taken by the SAR is done by asking the patient to provide scores for 8 basics emotions, which are:

- Excitement
- Joy
- Satisfaction
- Relaxation
- Boredom
- Sadness
- Stress
- Anger

These emotions are selected as they allow to create degrees into the state-of-mind of the patient. For example, if the patients gives back a high score for excitement this leads to a very high state-of-mind, but if it has a more neutral state-of-mind it will provide a high score for relaxation or boredom. Humans can experience emotions in a fuzzy way, for example when feeling joyful someone can still feel stressed to some extent. Consequently, to incorporate the feedback of the patient into the adaptation module of the SAR, the emotions are represented as fuzzy variables. The scores for the emotions are fed into a fuzzy logic inference system that establishes the current state-of-mind of the patient. According to the estimated state-of-mind of the patient, a reward is computed and assigned to the performed action of the SAR. The details of this FLC reward module are explained in Section III-B2.

Another reason for selecting emotions to retrieve feedback from the patients is that, unlike the state-of-mind of a person, emotions can be automatically detected by computer-vision systems. Also for the Nao robot, an automatic emotion detection system has been developed [26]. Perhaps, when using automatic emotion detection, the system detects other emotions than the ones currently used in the decision making system. In that case, the rule-base from the fuzzy inference system in Section III-B2 must be adapted to incorporate the detected emotions. When the SAR can detect the emotions automatically, it does not have to rely on the feedback given by the patients. However, before automatic emotion detection can be implemented in the decision making system of the SAR, more research has to be performed to see if the emotion detection systems also work for people with dementia. Automatic emotion detection might not be applicable for patients in advanced stages of dementia or dementia combined with other cognitive impairments such as Parkinson's Disease [27].

2) Fuzzy logic based reward module

For the RL module to learn which action to use in which state-of-mind, it is dependent on the rewards it receives for its actions. However, no mathematical relations exist that explicitly describe how emotions, state-of-mind and a successful social interaction with the patient are related to each other. Instead the relationships can be formulated intuitively with linguistic fuzzy rules. Hence, the RL adaptation module uses a FLC reward module to compute the rewards for its actions. The FLC reward module has two tasks: (1) to compute the state-of-mind of the patient from the emotional feedback the SAR received and (2) to compute a reward from the new state-of-mind and the change in state-of-mind. The FLC reward module has a fuzzy inference system for both tasks.

The first system determines the patient's state-of-mind based on the scores provided on the various emotional states. The rule-base for the fuzzy inference system to compute the patient's state-of-mind consists of rules that relate one or two of the retrieved emotions with the realized state-of-mind of the patient. The rules can be written in the following general form: "If *emotion 1* is $E_{1,i}$ and *emotion 2* is $E_{2,j}$, then *state-of-mind* is S_k." In this formulation *emotion 1* and *emotion 2* refer to the feedback from the patient on the respective emotions, $E_{1,i}$ and $E_{2,j}$, for $i = 1, ..., N_{E_1}$ and $j = 1, ..., N_{E_2}$, are the fuzzy sets corresponding to the linguistic terms used to categorize the emotions, *state-of-mind* refers to the realized state-of-mind of the patient after the interaction and S_k, for $k = 1, ..., N_S$, is the fuzzy set corresponding to the linguistic terms used to categorize the realized state-of-mind of the patient. The inference system uses Mamdani inference with center of gravity method for defuzzification.

The second task of the FLC reward module is to compute the rewards based on the change in the patient's state-of-mind after interacting with the SAR. The reward must favor actions that improve the patient's state-of-mind or stabilize their positive state-of-mind. For instance, if the state-of-mind of the patient improves after interacting with the SAR, the corresponding action/behavior of the SAR should be rewarded positively. On the contrary, whenever the state-of-mind of the patient worsens, the corresponding action/behavior of the SAR should be rewarded negatively. Moreover, the intensity of the change of the state-of-mind is factored into the reward, so actions that result in a big change in the state-of-mind are rewarded/penalized heavier than actions that result in a smaller change in the state-of-mind. Therefore, the reward for the RL based decision making module has two components: (1) the new state-of-mind of the patient, a better state-of-mind results in a higher reward, and (2) the change in the state-of-mind of the patient, the reward is higher when the patient goes from a bad state-of-mind to a good state-of-mind than when the patient stays in a good state-of-mind. The first component that is purely related to the new state-of-mind of the patient is treated as the quality related part of the reward, the component that is related to the relative change in the state-of-mind of the patient is treated as the intensity related part of the reward. As a result, when the initial state-of-mind of the patient is s_k and the SAR takes action a_k, which results in the new state-of-mind s_{k+1}, the reward $r(s_k, a_k)$ for the state-action pair is given by:

TABLE I. Reward components based on the new state-of-mind

New state-of-mind	$r^{\mathrm{q}}(s_{k+1})$
Very good	2.5
Good	1.5
Neutral	0
Bad	-1.5
Very bad	-2.5

$$r(s_k, a_k) = r^{\mathrm{q}}(s_{k+1}) + r^{\mathrm{i}}(s_k, s_{k+1}) \tag{2}$$

In this equations $r^{\mathrm{q}}(s_{k+1})$ is the quality related reward component, which is based on the new state-of-mind of the patient. Table I shows the values for $r^{\mathrm{q}}(s_{k+1})$ based on the new state-of-mind of the patient. $r^{\mathrm{i}}(s_k, s_{k+1})$ is the intensity related component of the reward. This part of the reward is computed with the second fuzzy inference system of the FLC based reward module and depends on the change in the state-of-mind of the patient. The rules for this inference system are written in the following general form: "If *old state-of-mind* is O_i and *new realized state-of-mind* is R_j, then $r^{\mathrm{i}}(s_k, s_{k+1})$ is I_l." In this formulation O_i and R_j, for $i = 1, ..., N_O$ and $j = 1, ..., N_R$, are the fuzzy sets corresponding to the linguistic terms used to categorize the old and realized state-of-mind of the patient respectively, I_l, for $l = 1, ..., N_I$, is the fuzzy set corresponding to the linguistic terms used to categorize the intensity related reward component for the adaptation module. The inference system uses Mamdani inference with center of gravity method for defuzzification.

IV. EXPERIMENTS

The different modules of the SAR's decision making process are assessed by a set of experiments. The FLC based decision making module for personalizing the behavior is assessed with a survey to compare the output of the SAR with the answers provided by the participants of the survey. The participants are not real dementia patients, as finding such a

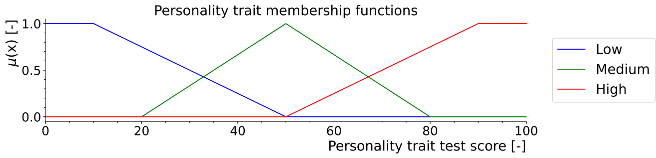

FIGURE 3. Membership functions of the antecedent of the second Mamdani inference system.

group is considered out of the scope of this research. For the experiment the relation between someone's personality and the desired behavior of the SAR is investigated. Therefore no special requirements are set for the participants of the survey. A group of 20 people, including both male and female participants aged between 20 and 70, have completed the survey for the assessment of the decision making system of the SAR. The RL based decision making module and the FLC reward module to adapt the SAR's behavior are assessed with an experiment in which simulated patients interact with the SAR. This section discusses the setup of the experiments and their results.

A. FLC DECISION MAKING MODULE EXPERIMENT

The first experiment is to assess the functioning of the FLC decision making module, which sets up an initial behavior based on the patient's Big-Five personality scores. This section starts with the implementation of the experiment after which the corresponding results are discussed.

The rule-base used for the personalization is presented in Table II. The membership functions of the FLC personalization module are shown in Fig. 3 and Fig. 4.

1) Experiment Description

For the assessment of the FLC decision making module a survey [28] is distributed in which the participants are asked to provide the results of a Big-Five personality test [29] along with their preferences in situations that might occur between a patient and a SAR. The survey is completely anonymous and has been approved by the Ethics Committee of the TU Delft. In total 20 participants have completed the survey, these participants were not dementia patients. Finding dementia patients willing to participate in this experiment was considered out of the scope of this research. also it is assumed that if the personalization to a person's personality works for non-dementia patients, it works for dementia patients as well. However, in future stages of the research the system must also be tested on dementia patients.

The survey contained questions that sketched social situations in which the different behavioral elements of the SAR play a crucial role. The participants are presented with two example behaviors in these social situations and are asked to provide their preference between the two behaviors on a scale of 1 to 5. An example question of the survey is:

Imagine you need a nurse to explain you how to use your new medications. One nurse has a strong loud voice, while the other nurse is more soft-toned. Which of the two nurses would you prefer to explain you the instructions?

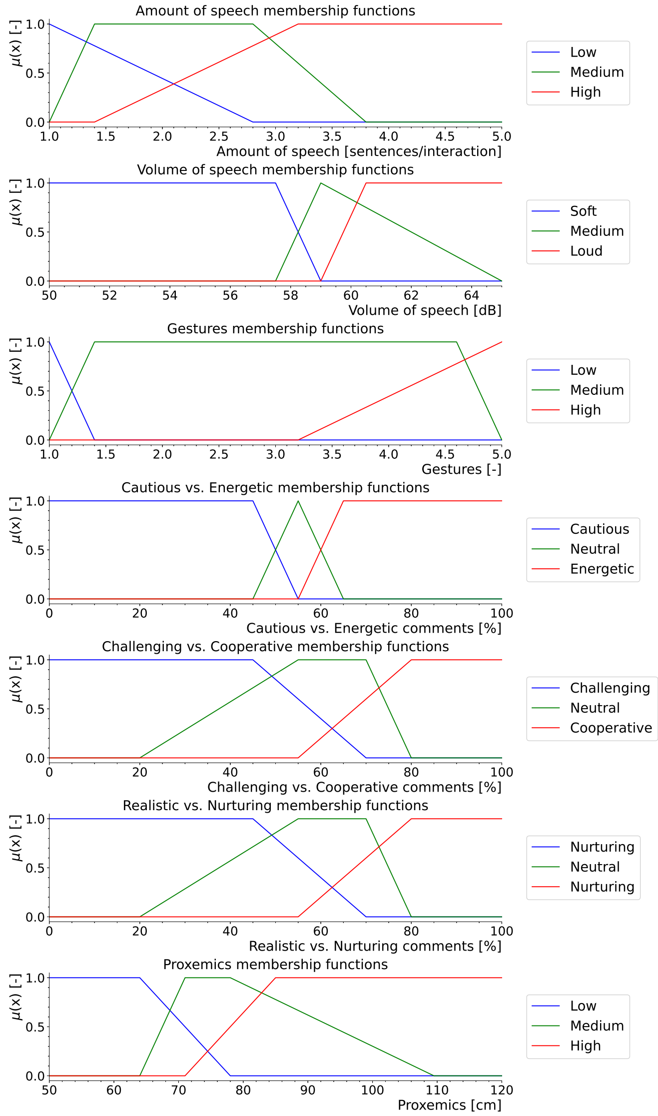

FIGURE 4. Membership functions of the consequent of the second Mamdani inference system.

In this question the participants are asked to provide their preference on the behavioral element volume of speech, the survey contained questions like these for all behavioral elements of the SAR.

In order to assess the functioning of FLC decision making module the answers of the survey are compared to the output of the system. The personalization of the SAR behavior is assessed based on the Root Mean Squared Error (RMSE) over the 6 participants used for verification. The RMSE is calculated according to:

$$\text{RMSE(i)} = \sqrt{\frac{\sum_{p=1}^{n}(b_p^{\text{SAR}}(i) - b^{\text{survey}_p}(i))^2}{n}} \tag{3}$$

In this equation i is an index, with $i = 1, \ldots, 7$, corresponding to the different behavioral parameters of the SAR, n is the

TABLE II. Rule base for the FLC personalization module.

If	Antecedent 1	and/or	Antecedent 2	Then	Consequent
If	*extroversion is Low*	-	-	Then	*Amount of speech is Low*
If	*extroversion is Medium*	-	-	Then	*Amount of speech is Medium*
If	*extroversion is High*	-	-	Then	*Amount of speech is High*
If	*extroversion is Low*	-	-	Then	*Volume of speech is Soft*
If	*extroversion is Medium*	and	*Neuroticism is Low*	Then	*Volume of speech is Medium*
If	*extroversion is Medium*	and	*Neuroticism is Medium*	Then	*Volume of speech is Medium*
If	*extroversion is Medium*	and	*Neuroticism is High*	Then	*Volume of speech is Soft*
If	*extroversion is High*	and	*Neuroticism is Low*	Then	*Volume of speech is Loud*
If	*extroversion is High*	and	*Neuroticism is Medium*	Then	*Volume of speech is Loud*
If	*extroversion is High*	and	*Neuroticism is High*	Then	*Volume of speech is Medium*
If	*extroversion is Low*	and	*Neuroticism is Low*	Then	*Gestures is Low*
If	*extroversion is Low*	and	*Neuroticism is Medium*	Then	*Gestures is Medium*
If	*extroversion is Medium*	and	*Neuroticism is Low*	Then	*Gestures is Medium*
If	*extroversion is Medium*	and	*Neuroticism is Medium*	Then	*Gestures is Medium*
If	*extroversion is High*	or	*Neuroticism is High*	Then	*Gestures is High*
If	*extroversion is Low*	-	-	Then	*Interactive comments is Cautious*
If	*extroversion is Medium*	-	-	Then	*Interactive comments is Neutral*
If	*extroversion is High*	-	-	Then	*Interactive comments is Energetic*
If	*Agreeableness is Low*	-	-	Then	*Motivating comments is Challenging*
If	*Agreeableness is Medium*	-	-	Then	*Motivating comments is Neutral*
If	*Agreeableness is High*	-	-	Then	*Motivating comments is Cooperative*
If	*Neuroticism is Low*	-	-	Then	*Feedback is Realistic*
If	*Neuroticism is Medium*	-	-	Then	*Feedback is Neutral*
If	*Neuroticism is High*	-	-	Then	*Feedback is Nurturing*
If	*extroversion is Low*	and	*Agreeableness is Low*	Then	*Proxemics is High*
If	*extroversion is Low*	and	*Agreeableness is Medium*	Then	*Proxemics is High*
If	*extroversion is Low*	and	*Agreeableness is High*	Then	*Proxemics is Medium*
If	*extroversion is Medium*	and	*Agreeableness is Low*	Then	*Proxemics is High*
If	*extroversion is Medium*	and	*Agreeableness is Medium*	Then	*Proxemics is Medium*
If	*extroversion is Medium*	and	*Agreeableness is High*	Then	*Proxemics is Low*
If	*extroversion is High*	and	*Agreeableness is Low*	Then	*Proxemics is Medium*
If	*extroversion is High*	and	*Agreeableness is Medium*	Then	*Proxemics is Low*
If	*extroversion is High*	and	*Agreeableness is High*	Then	*Proxemics is Low*

number of participants, b_p^{SAR} is the output behavior of the SAR for participant p and b^{survey_p} is the preferred behavior of participant p received from the survey. From the RMSE it still can be difficult to determine how well the personalization works. Therefore, also the scatter index (SI) is computed. The SI divides the RMSE with the average value of the observed value to account for the range over which the values are observed. The SI is computed according to:

$$\text{SI(i)} = \frac{\text{RMSE(i)}}{\frac{1}{n}\sum_{p=1}^{n} b_p^{\text{SAR}}(i)} \tag{4}$$

From the total of 20 answers received on the survey 2 were incomplete, making these unusable. From the remaining 18 answers, two-thirds are used for optimization of the membership functions of the fuzzy logic inference system. One-third of the answers is used for the assessment of the FLC based decision making module.

The optimization of the membership functions of the FLC personalization module focuses on optimizing the membership functions of the consequent, i.e. the seven behavioral elements. The membership functions for the antecedent, i.e. the personality traits, are illustrated in Fig. 3. These membership functions are not optimized as a single personality trait influences multiple behavioral elements. Therefore, it is easier to optimize the membership functions of the consequent. During the optimization, different values for the membership functions have been investigated. The values that resulted in the lowest error between the output of the FLC personalization module and the answers provided to the survey have been selected. The final membership functions for the sub-module are displayed in Fig. 4.

2) Results

In order to compare the outputs of the survey with the outputs of the FLC decision making module. The output for each behavioral element of the FLC decision making module has been scaled from their respective range to the range of the survey answers (1 to 5). Having all behavioral elements in the same range allows for better comparison between the different elements, that is why the range of 1 to 5 from the survey answers has been selected. The results of the comparison between the output of the SAR's FLC decision making module and the answers of the survey are illustrated in Fig. 5 to Fig. 12.

Fig. 5 illustrates the scores for the three personality traits, extroversion, agreeableness and neuroticism of the participants of the survey that were used for validation. Fig. 6 to Fig. 12 each illustrate the outcomes of the personalization from the FLC decision making module and the answers of the survey of one the seven behavioral elements.

The RMSE's and the SI's of the different behavioral elements are displayed in Table III, the values in the second

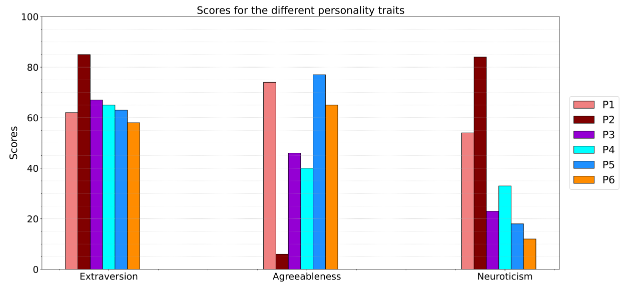

FIGURE 5. Personality scores of the survey participants used for validation.

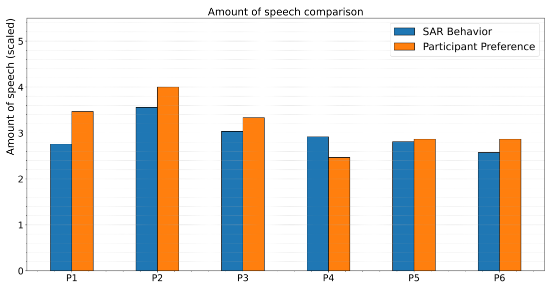

FIGURE 6. Validation results for the Amount of Speech.

column of the table is the RMSE in the range of 1 to 5. With these values the personalization of the different behavioral elements can be compared to each other. The values in the third column of the table display the RMSE in the range of the behavioral element, with the respective ranges in brackets behind them. The fourth column shows the RMSE as a percentage and the fifth column shows the SI of the behavioral parameters.

From the bar plots and the values for the RMSE and SI it can be concluded that personalization works well for all behavioral elements apart from the Feedback comments. Most elements have one or two small outliers among the participants, such as P2 and P4 for the Gestures and P5 for Motivating comments for example. This is expected as the literature describes the general preferences of the personality

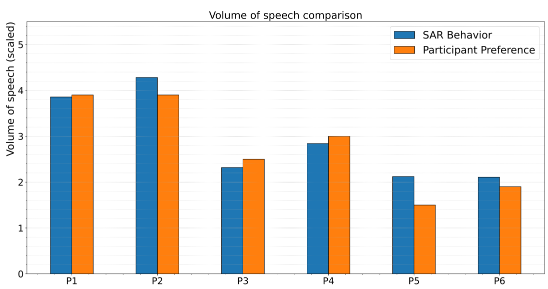

FIGURE 7. Validation results for the Volume of Speech.

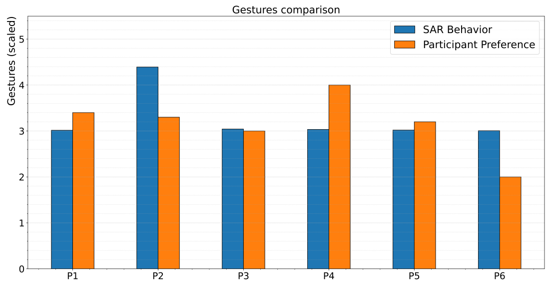

FIGURE 8. Validation results for the Gestures.

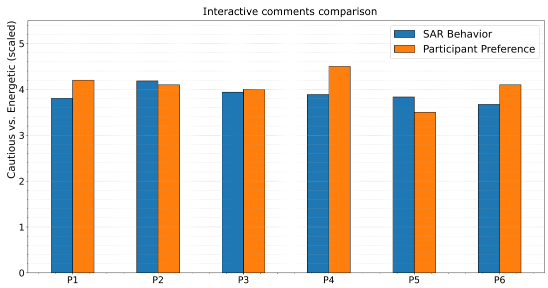

FIGURE 9. Validation results for the Interactive comments.

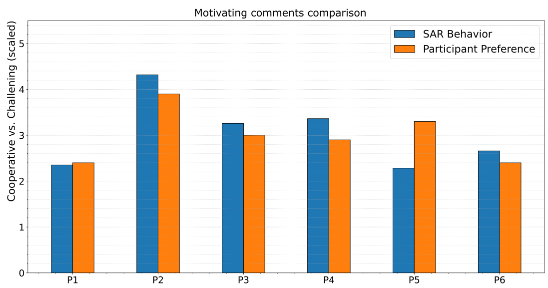

FIGURE 10. Validation results for the Motivating comments.

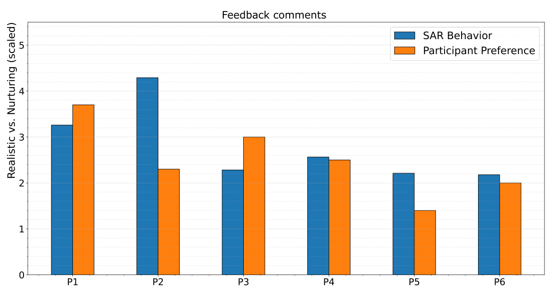

FIGURE 11. Validation results for the Feedback comments.

TABLE III. RMSE's of the different behavioral elements.

Behavioral element	RMSE [range 1-5]	RMSE [element range]	RMSE as % of range	SI
Amount of speech	0.42	0.42 [1-5]	10.6%	14.4%
Volume of speech	0.33	1.22 [50-65]	8.14%	2.13%
Gestures	0.74	0.74 [1-5]	18.6%	22.9%
Interactive comments	0.37	9.33 [0-100]	9.33%	12.9%
Motivating comments	0.51	12.8 [0-100]	12.8%	25.0%
Feedback	0.95	23.6 [0-100]	23.6%	52.6%
Proxemics	0.58	10.1 [50-120]	14.5%	13.2%

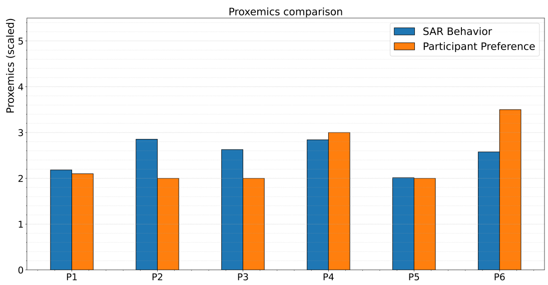

FIGURE 12. Validation results for the Proxemics.

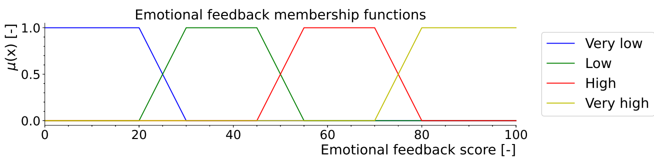

FIGURE 13. Membership functions for the patient's emotional scores.

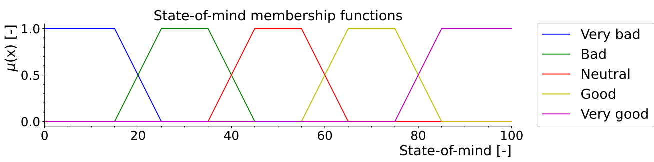

FIGURE 14. Membership functions for the patient's state-of-mind.

traits, but in reality it can be the case that a person deviates from this general preference.

However the Feedback comments have a significant outlier with P2. The RMSE, as calculated in (3), considers the squares of the errors, hence larger outliers are penalized heavier. This explains the higher RMSE and SI for the Feedback comments.

A reason for the large outliers for the Feedback comments can be that in the survey participants are asked to imagine themselves in a social situation instead of really experiencing this situation. The survey describes the situation in words, which might make it difficult for a participant to imagine the situation. It could be the case that for the question of the feedback participants had difficulties to imagine the sketched situation or interpreted the situation differently than intended, leading to inaccurate results.

Therefore, in further stages of the research it is crucial that a test group interacts with the SAR over a longer period of time, such that they really experience the situations with the robot instead of having to imagine them. After the interactions a survey can be held, which asks about how they experienced their interactions with the SAR. However, in the current stage of the research there are not enough resources nor time for such an experiment.

A second reason for the large outliers in the personalization of the Feedback comments is that the expected relation between neuroticism in the preference in Feedback comments is not present in the group of participants. People who score high on neuroticism are more anxious and insecure [15]. Therefore, a more nurturing behavior would be preferred by people who score high in neuroticism. However, the results of the personalization do not show this relation.

For example, the preference for the Feedback comments for P2 and P6 is approximately the same, while P2 has the highest level of neuroticism and P6 has the lowest level of neuroticism.

Hence, it can be the case that there is no relation between the preference in the Feedback comments and the level of neuroticism of the participants. Further testing, such as the experiment where participants interact with the SAR in real life, can clarify whether the relation exists or not.

B. REINFORCEMENT LEARNING DECISION MAKING MODULE AND FUZZY LOGIC CONTROL REWARD MODULE EXPERIMENT

The RL module and the FLC reward module are assessed together as these modules depend on each other and together they are responsible for the adaptation of the SAR behavior to the patient's current state-of-mind. The rule-base for the two fuzzy inference systems of the FLC reward module are presented in Table IV and Table V. The membership functions of the inference system are illustrated in Fig. 13, Fig. 14 and Fig. 15.

1) Experiment Description

The modules are assessed based on computer-based simulations of interactions between the SAR and its patient. The reasons for using computer-based simulated patients, which were also used in [9], are: (1) With computer-based simulated patients a large number of interactions can be held in short periods of time. (2) A wide variety of potential responses

TABLE IV. Rule base for the first inference system of the FLC reward module.

If	Antecedent 1	and/or	Antecedent 2	then	Consequent
If	*excitement is very high*	-	-	then	*state-of-mind is very good*
If	*excitement is high*	and	*joy is very high*	then	*state-of-mind is very good*
If	*excitement is high*	-	-	then	*state-of-mind is good*
If	*joy is very high*	-	-	then	*state-of-mind is good*
If	*joy is high*	and	*satisfaction is very high*	then	*state-of-mind is good*
If	*joy is high*	and	*satisfaction is high*	then	*state-of-mind is good*
If	*joy is high*	and	*relaxation is very high*	then	*state-of-mind is good*
If	*sadness is high*	and	*boredom is very high*	then	*state-of-mind is bad*
If	*sadness is very high*	-	-	then	*state-of-mind is bad*
If	*stress is very high*	or	*stress is high*	then	*state-of-mind is bad*
If	*anger is high*	-	-	then	*state-of-mind is bad*
If	*anger is very high*	-	-	then	*state-of-mind is very bad*
If	*anger is high*	and	*sadness is very high*	then	*state-of-mind is very bad*
If	*stress is very high*	and	*sadness is very high*	then	*state-of-mind is very bad*

TABLE V. Rule base for the second inference system of the FLC reward module.

If	Antecedent 1	and/or	Antecedent 2	Then	Consequent
If	*old state-of-mind is Very bad*	and	*new realized state-of-mind is Very bad*	Then	$r^i(s_k, s_{k+1})$ *is Negative*
If	*old state-of-mind is Very bad*	and	*new realized state-of-mind is Bad*	Then	$r^i(s_k, s_{k+1})$ *is Neutral*
If	*old state-of-mind is Very bad*	and	*new realized state-of-mind is Neutral*	Then	$r^i(s_k, s_{k+1})$ *is Positive*
If	*old state-of-mind is Very bad*	and	*new realized state-of-mind is Good*	Then	$r^i(s_k, s_{k+1})$ *is Very positive*
If	*old state-of-mind is Very bad*	and	*new realized state-of-mind is Very good*	Then	$r^i(s_k, s_{k+1})$ *is Very positive*
If	*old state-of-mind is Bad*	and	*new realized state-of-mind is Very bad*	Then	$r^i(s_k, s_{k+1})$ *is Negative*
If	*old state-of-mind is Bad*	and	*new realized state-of-mind is Bad*	Then	$r^i(s_k, s_{k+1})$ *is Negative*
If	*old state-of-mind is Bad*	and	*new realized state-of-mind is Neutral*	Then	$r^i(s_k, s_{k+1})$ *is Neutral*
If	*old state-of-mind is Bad*	and	*new realized state-of-mind is Good*	Then	$r^i(s_k, s_{k+1})$ *is Positive*
If	*old state-of-mind is Bad*	and	*new realized state-of-mind is Very good*	Then	$r^i(s_k, s_{k+1})$ *is Very positive*
If	*old state-of-mind is Neutral*	and	*new realized state-of-mind is Very bad*	Then	$r^i(s_k, s_{k+1})$ *is Very negative*
If	*old state-of-mind is Neutral*	and	*new realized state-of-mind is Bad*	Then	$r^i(s_k, s_{k+1})$ *is Negative*
If	*old state-of-mind is Neutral*	and	*new realized state-of-mind is Neutral*	Then	$r^i(s_k, s_{k+1})$ *is Neutral*
If	*old state-of-mind is Neutral*	and	*new realized state-of-mind is Good*	Then	$r^i(s_k, s_{k+1})$ *is Positive*
If	*old state-of-mind is Neutral*	and	*new realized state-of-mind is Very good*	Then	$r^i(s_k, s_{k+1})$ *is Very positive*
If	*old state-of-mind is Good*	and	*new realized state-of-mind is Very bad*	Then	$r^i(s_k, s_{k+1})$ *is Very negative*
If	*old state-of-mind is Good*	and	*new realized state-of-mind is Bad*	Then	$r^i(s_k, s_{k+1})$ *is Negative*
If	*old state-of-mind is Good*	and	*new realized state-of-mind is Neutral*	Then	$r^i(s_k, s_{k+1})$ *is Neutral*
If	*old state-of-mind is Good*	and	*new realized state-of-mind is Good*	Then	$r^i(s_k, s_{k+1})$ *is Neutral*
If	*old state-of-mind is Good*	and	*new realized state-of-mind is Very good*	Then	$r^i(s_k, s_{k+1})$ *is Positive*
If	*old state-of-mind is Very good*	and	*new realized state-of-mind is Very bad*	Then	$r^i(s_k, s_{k+1})$ *is Very negative*
If	*old state-of-mind is Very good*	and	*new realized state-of-mind is Bad*	Then	$r^i(s_k, s_{k+1})$ *is Very negative*
If	*old state-of-mind is Very good*	and	*new realized state-of-mind is Neutral*	Then	$r^i(s_k, s_{k+1})$ *is Negative*
If	*old state-of-mind is Very good*	and	*new realized state-of-mind is Good*	Then	$r^i(s_k, s_{k+1})$ *is Neutral*
If	*old state-of-mind is Very good*	and	*new realized state-of-mind is Very good*	Then	$r^i(s_k, s_{k+1})$ *is Positive*

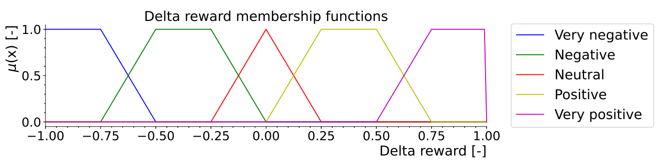

FIGURE 15. Membership functions for the rewards for the RL based decision making module.

to the actions selected by the SAR can be simulated and assessed, without the need of finding a large group of people that may still not possess preferences with the wide variations possible for simulated patients. (3) Performing experiments based on trial and error on real human beings can potentially be harmful.

The RL decision making module and the FLC reward module require feedback on the patient's emotions, i.e. excitement, joy, satisfaction, relaxation, boredom, sadness, stress, and anger. Therefore, the simulated patient must be able to generate responses to the changes in the SAR's behavior that consists of scores for these eight emotions.

For each behavioral element of the SAR a table is generated that includes probabilities for the emotional scores for the different actions of the SAR, i.e. increase the behavioral element a lot, increase it moderately, decrease it moderately, and decrease it a lot. Table VI shows an example response table for the responses of a simulated patient to changes in the amount of speech, although it is kept concise by leaving out several emotions and actions.

For each action the scores for the emotions are divided into four equal ranges. Probabilities are assigned to these score ranges, which represent the likelihood that the patient provides a score within this range as feedback. Using the probabilities one range is selected, the score for the action is a random integer in the selected range. The response of the patient depends on his or her state-of-mind. Therefore, the different rows of the table represent the patient's states of mind.

Thus, the way to read this table is: When the SAR increases the volume of speech a lot and the patient is in a good state-of-mind, then the probability for the score for the emotion excited to be in the range 0-25 is 0%, the probability for it to be in the range 26-50 is 20%, the probability for it to be in the range 51-75 is 30%, and the probability for it to be in the range 76-100 is 50%.

When one of the sore ranges is selected based on the probabilities in the table, the exact score is computed by taking a random number in that range. Again it should be noted that the other emotions but those are left out to keep Table VI concise.

The data for the response tables of the virtually scripted patients are produced in two different ways: (1) Based on survey answers to gather data from real human beings. (2) Randomly generated responses to gather a wide variety of responses including extreme cases, which might not be included in the survey group.

The survey [30] is held among the same group as the survey for the assessment of the FLC decision making module. This survey contains questions on how the current state-of-mind of the patients changes when the person they are interacting with behaves differently than usual. An example of the questions is:

You are initially in the very good state-of-mind explained before because of your new car. Imagine a friend asks you to play a board game with her/him. The way in which your friend asks to play the board game is described below. Please indicate how the following ways of asking affect your very good state-of-mind.

Your friend is significantly more energetic versus cautious than usual. For instance, the way she/he asks you to play the game is: 'Let's play this super fun board game, I am sure you'll love it.'

The response tables are generated based on the answers provided by the participants. Besides the data from real human beings, also random response tables are generated. In this way a lot of computer based simulated patients can be created and a lot of simulations can be run in short periods of time.

Also with randomly generated response tables more extreme cases can be captured to see whether the learning process of the SAR converges in these cases too.

For the randomly generated response tables, the tables are slightly adjusted halfway through the experiment to see whether the RL decision making module can deal with changes in the responses of the patients, something that might happen in real life.

In the experiment it is assessed whether the SAR is able to adapt its behavior to the preferences of the simulated patient in different states of mind. In the interactions, the simulated patient is put into different states of mind. Feedback is generated based on the response tables of the virtual patients. From the feedback rewards are computed with the FLC reward module based on the change in the patient's state-of-mind.

Over time the learning process of the RL decision making module should converge, where the SAR selects the best actions to improve the patient's state-of-mind or stabilize it in a positive state-of-mind.

2) Results

The RL decision making module and the FLC reward module are assessed in the same experiment. In interactions between the SAR and computer-based simulated patients, the modules are assessed based on convergence of the learning process and the performance of the final behavior of the SAR in the different emotional states of the patients. The two modules are tested both on the computer-based simulated patients generated from the survey answers and on the randomly generated simulated patients.

The convergence of the learning process of the SAR is assessed by plotting the absolute change of the Q-values of the RL decision making module. Over time, the absolute change in Q-value should decrease and the plot of the learning process should flatten out. When the slope of the plot of the change in Q-value is approximately 0 (below 0.00005) it is assumed that the learning process has converged.

The performance of the RL decision making module and the FLC reward module is determined by comparing the selected actions after learning with the most optimal action. This most optimal action is determined by using the fuzzy inference system to compute state-of-mind of the patient. For each action a set of emotional scores is determined using the state transition table of that patient. These emotional scores are implemented in the fuzzy inference system to compute the state-of-mind of the patient and the action that results in the highest realized state-of-mind is considered the most optimal action. By comparing the action selected from the Q-values after the learning phase to this most optimal action, the performance of the RL decision making module can be determined.

The SAR should select an action for every combination of behavioral elements (7) and current states of mind (5), making a total of 35 actions. Therefore, the performance p of the output of the RL decision making module is calculated according to:

$$p = \frac{n_{\text{correct}}}{35} \tag{5}$$

In this equation n_{correct} is the number of correct actions selected by the SAR.

First the results for the computer-based patient simulations generated from the survey data are discussed. Fig. 16 illustrates the absolute Q-changes of the first participant of the survey.

The simulations consist of 100,000 interactions with the computer-based simulated patients. In order to keep the figures readable, the average absolute change in Q-value is plotted of every 100 interactions. The blue line in Fig. 16

TABLE VI. Response of a specific simulated patient to the changes in the volume of speech.

	Emotions																	
	Excitement																...	**Anger**
	Actions																...	...
	Increase a lot				**Increase**				**Decrease**				**Decrease a lot**				...	...
	Score ranges				*Score ranges*				*Score ranges*				*Score ranges*				...	...
State-of-mind	**0-25**	**26-50**	**51-75**	**76-100**	**0-25**	**26-50**	**51-75**	**76-100**	**0-25**	**26-50**	**51-75**	**76-100**	**0-25**	**26-50**	**51-75**	**76-100**	...	...
Very good	0%	0%	30%	70%	0%	0%	40%	60%	5%	15%	30%	50%	20%	25%	25%	30%	...	...
Good	0%	20%	30%	50%	0%	10%	40%	50%	10%	10%	40%	40%	25%	35%	25%	15%	...	...
Neutral	10%	40%	30%	20%	15%	25%	30%	30%	15%	35%	35%	15%	60%	30%	5%	5%	...	...
Bad	60%	35%	5%	0%	65%	25%	10%	0%	40%	35%	15%	10%	85%	10%	5%	0%	...	...
Very bad	85%	15%	0%	0%	75%	10%	15%	0%	50%	25%	25%	0%	90%	10%	0%	0%	...	...

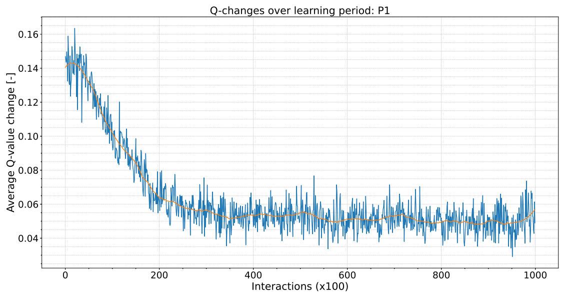

FIGURE 16. Results for the learning process of participant 1.

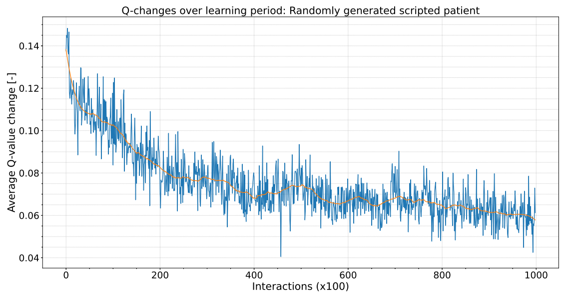

FIGURE 17. Results for the learning process of a participant with a randomly generated response table.

TABLE VII. Results of the learning process for all participants.

Participant	**Convergence [interactions]**	**Performance [%]**
P1	41,600	91.4
P2	37,300	85.7
P3	31,000	77.1
P4	39,300	82.9
P5	31,600	88.6
P6	28,700	80.0
Average	34,917	84.3

TABLE VIII. Results of the learning process for the randomly generated simulated patients.

	Convergence before change [interactions]	**Convergence after change [interactions]**	**Performance [%]**
Average	28,472	54,060	81.7

illustrates the average of the absolute change in Q-value over the last 100 interactions, while the orange line illustrates a trend line which eases the analysis. Table VII illustrates the results for the convergence of the learning process and the performance of the outcome for all participants of the survey.

In Fig. 16 it can be seen that the change in the Q-values decrease rapidly, which indicates that the learning process is converging. Around 36400 interactions the slope of the trend line is below 0.00005, so that is when the learning process is assumed to be fully converged. In Table VII it can be seen that the learning process converges for all 6 participants used for validation, with an average of 34,917 interactions. The average performance of the selected actions by the RL decision making module is 84.3%.

Besides the computer-based patients generated from the survey data, also simulations have been performed with computer-based patient with randomly generated state transition tables. For these simulations, the response tables are slightly adjusted halfway through the experiment (at interaction 50,000) to see if the RL decision making module can also deal with changing responses of the patient.

Fig. 17 illustrates an example of the learning process of one of the 100 simulated patients with a randomly generated response table. The absolute change of the Q-value decreases nicely, but around interaction 50,000 there is a slight increase in the absolute change of the Q-value after which the it starts to decrease again. The performance of the output for this simulated patient is 82.9%.

For the simulated patients with a randomly generated response table the average convergence time and the average performance of the outputs are shown in Table VIII. The learning process for all 100 simulated patients converged before the adjustments in the response tables, so an extra column has been added with the average convergence time after the adjustments in the response tables.

The learning process of the SAR has converged for all of the 100 passengers. In Table VIII it can be seen that the time to converge after the adjustments is significantly smaller than the initial convergence time. This makes sense as the adjustments were small. Hence, for many of the combinations of the behavioral elements and the current state-of-mind of the patient the optimal action has not changed. Therefore, the SAR requires less time to learn and converge.

A second thing to notice from the results for the randomly generated computer-based simulated patients is that the performance for these patients is lower than for the simulations where the simulated patients are based on the survey answers.

Two possible reasons exist for this: (1) the adjustments in the response tables lead to lower performance values, (2) the cases for the randomly generated simulated patients are more complex than the simulated patients based on the survey data.

First of all, the adjustments made to the response tables halfway through the learning process could result in a lower performance of the output of the RL decision making module and the FLC reward module.

With the decaying ϵ it can be the case that the RL decision making module does not explore sufficiently anymore to recover all adjustments in the response tables. In further research it should be explored if different exploration strategies or RL methods yield an even higher performance.

A second reason for the lower performance with the randomly generated simulated patients is that the RL decision making module is faced with more complex cases than for the simulated patients based on the survey data.

As mentioned before the randomly generated simulated patients contain more diverse responses. Consequently, more extreme cases are included in the randomly generated simulated patients. Presumably, the RL decision making module struggles more with these extreme cases resulting in a lower performance.

Moreover, for some of the combinations, the participants answered that their state-of-mind would be influenced the same for multiple actions. As a result, the optimal action is not a single action but can be multiple actions. When a participant does not have a preference for an individual action for many of the combinations, the performance for this participant automatically will be higher.

Similar to the experiment for the personalization of the SAR's behavior, an experiment in which real human beings interact with the SAR must also be carried out for further assessment of the adaptation module. In such an experiment direct feedback can be asked to the participants on how they experience their interactions with the SAR.

C. COMBINED FUZZY LOGIC CONTROL AND REINFORCEMENT LEARNING DECISION MAKING MODULES VERSUS THE INDIVIDUAL REINFORCEMENT LEARNING DECISION MAKING MODULE

In the previous sections the FLC decision making module for personalization and the RL decision making module with the FLC reward module for adaptation have been assessed separately. This section compares the functioning of the personalization module and the adaptation modules combined versus a SAR equipped with just the RL decision making module and the FLC reward module to assess the benefit of adding the FLC decision making module for personalization.

1) Experiment

It was observed that the system as a whole is able to converge to an optimal solution with a performance of around 82%. However, from that experiment it cannot be assessed what the influence of the individual decision making modules is. In the comparison between the individual RL decision making module and the two combined decision making modules, the effect of adding the FLC decision making module is assessed. When combining the RL decision making module with the FLC decision making module, the initial behavior of the robot should be closer to the preference of the patient. Therefore, the learning process should converge quicker for the combination of the two decision making modules than for the RL module individually.

Similar to the experiment of the RL decision making module, the combination of the two decision making modules is assessed with computer-based simulated patients. The patients are based on the survey answers from the assessment of the RL module. However, for better comparison of the results of the individual RL decision making module and the two combined decision making modules, randomness has been reduced as much as possible. Therefore, unlike with the assessment of the RL decision making module, the preference of the patient does not change halfway through the experiment. The randomness in the selection of the actions for the RL decision making module has not been removed as this is crucial for the exploration of the system. Furthermore, as the individual RL decision making module does not have the FLC decision making module to initialize the behavior of the SAR, the initialization of the SAR's behavioral elements is done randomly for the individual RL decision making module.

2) Results

The convergence of the combination of the two decision making modules is compared to the RL decision making module individually for each combination of patient state-of-mind and robot behavioral element.

Three patterns stand out in the convergence of the two setups. First of all, there is the favored convergence pattern, where the initial behavior of the two combined decision making modules is closer to the behavior preferred by the patient making the learning process converge quicker than the individual RL module. This is illustrated in Fig. 18.

However, in some cases the personalization of the FLC decision making module deviates from the real preferred behavior of the patient. Consequently, the random initialization of the individual RL decision making module might be closer to the preference of the patient making its learning process converge faster than for the two combined decision making modules. This is illustrated in Fig. 19.

Finally, in some cases the personalization actually leads to a better initialization but due to the randomness in the exploration phase of the learning process, the individual RL decision making module converges faster. In the exploration phase random actions are taken to learn which actions give high rewards. Generally, the learning process converges faster when the initial behavior is closer to the preference of the patient. However, when the random actions in the exploration phase of the learning process takes the behavior

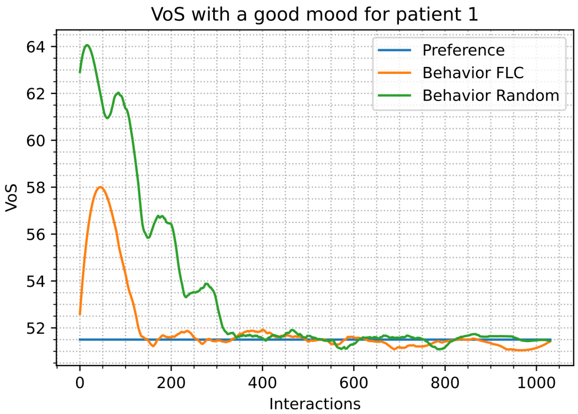

FIGURE 18. The personalization is closer making the process converge faster (VoS stands for the parameter Volume of Speech).

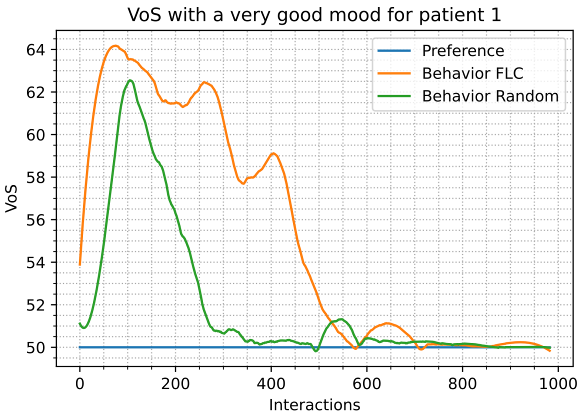

FIGURE 19. The random initialization is closer than the personalization (VoS stands for the parameter Volume of Speech),

of the SAR in the wrong direction, this might lead to a longer convergence period. This is illustrated in Fig. 20. It should be noted, that this also works the other way around when the randomly initialized behavior is closer to the preference than the initial behavior from the personalization.

Table IX shows the overall results for the comparison of the learning process of the two decision making modules combined and the individual RL module for one-third of the participants of the survey, as was done in the assessment of the individual decision making modules. The comparison of the combined decision making modules to the individual RL decision making module is based on three aspects:

- The overall convergence rate r, i.e. in how many percent of the cases the learning process converges
- The average convergence time t, i.e. how fast the learning process converges
- The performance p of the system, similar to the performance in section IV-B

The difference in convergence time between the two combined decision making modules and the individual RL module is expressed in a percentage which is calculated with:

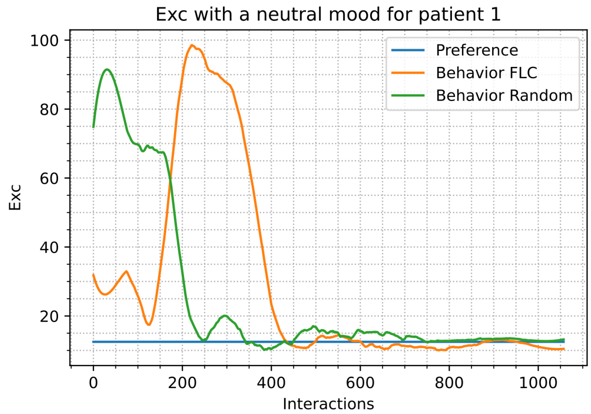

FIGURE 20. The personalization is closer but due to exploration does not converge faster (Exc stands for the parameter Interactive comments).

$$\Delta\%_t = \frac{t_{rl} - t_{combined}}{t_{combined}} \tag{6}$$

First of all, though it is minimal, the overall convergence rate of the combined decision making modules is higher than the individual RL decision making module. For the combined decision making modules there was only one combination (out of 210) of patient state-of-mind and behavioral element where the learning process did not converge.

Secondly, and most importantly, the convergence time is on average 13.3% lower for the combined decision making modules than for the individual RL decision making module. As expected, the FLC decision making module created a better initial behavior leading to a shorter convergence time. As mentioned before, for individual combinations of patient state-of-mind and behavioral parameter the randomness in the exploration phase or the random initialization can lead to better convergence. However, overall it can be stated that adding the FLC decision making module to the RL decision making module leads to a decreased convergence time. This means behavior of the SAR matches the preference of the patient quicker when applying the system in real time.

V. CONCLUSION

An aging world population puts an increasing pressure on health care workers working with dementia patients. The increasing number of dementia patients confine care-givers to provide personalized care to their patients. As a result the quality of the care provided to dementia patients reduces and the well-being of dementia patients is negatively affected.

As a solution to this increasing pressure on health-care workers, this paper investigates the possibilities to use socially assistive robots (SARs) to provide care to dementia patients. SARs are robot that assist people through social interaction. Currently they are already used to assist therapists in dementia therapy. The results of these researches show that different patients respond differently to the SAR and that an individual patient responds differently to the SAR

TABLE IX. Results of the comparison between the combined decision making modules and the individual RL module.

Participant	$r_{combined}$[%]	r_{rl}[%]	$t_{combined}$[-]	t_{rl}[-]	Δt [%]	$p_{combined}$[%]	p_{rl}[%]
P1	100	100	422.6	472.9	11.9	92.2	84.3
P2	100	97.1	414.7	485	16.9	83.3	82.0
P3	100	97.1	425.3	480.1	11.3	87.8	85.7
P4	97.1	97.1	398.0	424.9	6.7	89.5	86.5
P5	100	100	408.3	479.9	17.5	76.1	70.1
P6	100	100	399.4	461.0	15.4	85.7	80.3
Average	99.5	98.6	411.4	467.3	13.3	85.8	81.5

on different occasions. Therefore, it is important that a SAR can personalize its care to an individual patient and adapt this behavior over changing circumstances. In this paper the decision process of a SAR is designed that personalizes its behavior to three of the five Big-Five personality traits (extroversion, agreeableness and neuroticism) and adapts this behavior to the current state-of-mind of this patient.

The decision making process of the SAR consists of three main modules: a Fuzzy Logic Control (FLC) personalization module for personalizing the SAR's behavior to the personality of the patient, and a Reinforcement Learning (RL) based decision making module together with a FLC reward module for adaptation of the SAR's behavior to the current state-of-mind of the patient.

The FLC personalization module personalizes the SAR's behavior to three personality traits of the patient using predefined rules that relate the different elements of the SAR's behavior to the patient's personality traits.

The RL based decision making module adapts the personalized behavior to the current state-of-mind of the patient. The module uses Q-learning to select the right way of adapting the personalized behavior, such that the adapted behavior matches the preferences of the patients in their current state-of-mind.

The FLC reward module computes the reward for the RL adaptation module, this reward is dependent on the change in the state-of-mind of the patient. When the state-of-mind of the patient improves it receives a higher reward than when the state-of-mind of the patient decreases.

The personalization of the SAR is assessed by means of a survey. In this survey participants were asked to provide their scores for the different personality traits, together with their preferences in certain social situations where the SAR's behavioral elements played a crucial role.

For most behavioral elements this first assessment looks promising. However, for the behavioral element of Feedback comments the results were not sufficient. Further experiments are required to investigate whether the personalization does not work for the Feedback comments because no relation exists between this behavioral element and the trait neuroticism or that the question in the survey was unclear for the participants such that the answers did not match the output of the FLC personalization module.

The adaptation of the SAR is assessed by letting the SAR interact with computer-based simulated patients. Two different kinds of simulated patients have been used for the experiment: simulated patients where the responses are based on data from a second survey and simulated patients where the response was random. For the simulated patients with a random response, the responses were slightly adjusted halfway through the experiment to see if the SAR could deal with such changes.

The results of the experiment show that the RL adaptation module manages to have a converging learning process for all patients. The performance of the final output of the RL based decision making module is measured by the amount of combinations of behavioral element and current patient state-of-mind for which the selected action actually was the optimal action. For the simulated patients based on the survey data an performance of 84.3% is achieved, for the simulated patients with random responses an performance of 81.7% is achieved.

The reason for the lower performance of the simulated patients is on the one hand more extreme cases are included in the randomly generated simulated patients, on the other hand the fact that the responses of these simulated patients were changed halfway through the learning process can cause a lower performance. Different exploration techniques or perhaps a different RL method should be investigated to see whether these can improve the performance for the randomly generated simulated patients.

When comparing the learning process of the individual Reinforcement Learning based decision making module to the combination of the FLC personalization module and the RL adaptation module it becomes evident that adding the FLC personalization module makes the learning process of the SAR converge with on average 13.3%.

The research still is in an early stage, for later stages of the research several points have to be carried out. First of all, experiments with real human beings, preferably real dementia patients, must be conducted. Although the initial results of the assessment look promising, they are based on surveys and computer-based simulations. The survey relied on human interpretation, where participants had to imagine being in a situation where they interacted with the SAR. Imagining such situations can be difficult and is sensitive to the participant's interpretation of the question. Therefore, an experiment where people interact with the SAR in real-life is crucial for further assessment of the decision making process.

Furthermore, besides the behavioral elements for the social interaction, behavioral elements to select activities for the dementia patients should be added. When the SAR is able to suggest activities to the dementia patients, the extent to which it can provide care to the patient is increased and their well-being is further improved. However, for the behavioral elements to suggest activities, the remaining two traits of the Big-Five personality traits (openness and conscientiousness) should also be included in the FLC personalization module as openness is related to a person's creativity levels for example. When the SAR must be able to suggest activities these traits also become relevant. Suggesting activities to the dementia patients keeps them entertained, which makes it an important feature to add in later stages of the research.

Finally, instead of asking the patients to provide feedback on their emotions to compute the rewards for the RL adaptation module, which is achieved with the response tables in the computer based simulations, the SAR should be able to detect these emotions automatically. Techniques already exist that can detect emotion from facial expressions. For the patients automatic detection of their emotion is more convenient as they do not have to provide verbal feedback to the SAR. Moreover, the system likely works faster as there is no hassle with asking for feedback. Therefore, having automatic emotion detection would be a great advantage for the SAR.

REFERENCES

[1] United Nations, Department of Economic and Social Affairs, Population Division, *World Population Ageing 2019*. New York: United Nations, 2020. [Online]. Available: http://digitallibrary.un.org/record/3846855

[2] C. Ferri, M. Prince, C. Brayne, H. Brodaty, L. Fratiglioni, M. Ganguli, K. Hall, K. Hasegawa, H. Hendrie, Y. Huang, A. Jorm, P. Menezes, E. Rimmer, and M. Scazufca, "Global prevalence of dementia: A delphi consensus study," *Lancet*, vol. 366, no. 9503, pp. 2112–2117, 2006.

[3] L. Mierlo, H. van der Roest, F. Meiland, and R.-M. Dröes, "Personalized dementia care: Proven effectiveness of psychosocial interventions in subgroups," *Ageing research reviews*, vol. 9, no. 2, pp. 163–183, 2009.

[4] T. Epp, "Person-centred dementia care: A vision to be refined," *The Canadian Alzheimer Review*, vol. 5, no. 3, pp. 14–18, 2003.

[5] D. Feil-Seifer and M. J. Mataric, "Defining socially assistive robotics," in *9th International Conference on Rehabilitation Robotics*. Chigago, United States of America: IEEE, 2005, pp. 465–468.

[6] T. Fong, I. Nourbakhsh, and K. Dautenhahn, "A survey of socially interactive robots," *Robotics and Autonomous Systems*, vol. 42, no. 3-4, pp. 143–166, 2003.

[7] P. Marti, M. Bacigalupo, L. Giusti, C. Mennecozzi, and T. Shibata, "Socially assistive robotics in the treatment of behavioural and psychological symptoms of dementia," in *The First IEEE/RAS-EMBS International Conference on Biomedical Robotics and Biomechatronics*. Pisa, Italy: IEEE, 2006, pp. 483–488.

[8] W. Moyle, C. Jones, J. Murfield, L. Thalib, E. Beattie, D. Shum, and B. Draper, "Using a therapeutic companion robot for dementia symptoms in long-term care: reflections from a cluster-RCT," *Aging & Mental Health*, vol. 23, no. 3, pp. 329–336, 2017.

[9] C. Moro, G. Nejat, and A. Mihailidis, "Learning and personalizing socially assistive robot behaviors to aid with activities of daily living," *ACM Transactions on Human-Robot Interaction*, vol. 7, no. 2, pp. 1–25, 2018.

[10] A. Tapus, C. Tapus, and M. J. Matarić, "The use of socially assistive robots in the design of intelligent cognitive therapies for people with dementia," in *IEEE International Conference on Rehabilitation Robotics*. Kyoto, Japan: IEEE, 2009, pp. 924–929.

[11] C. Clabaugh, K. Mahajan, S. Jain, R. Pakkar, D. Becerra, Z. Shi, E. Deng, R. Lee, G. Ragusa, and M. Matarić, "Long-term personalization of an in-home socially assistive robot for children with autism spectrum disorders," *Frontiers in Robotics and AI*, vol. 6, no. 110, pp. 1–18, 2019.

[12] K. Tsiakas, M. Abujelala, and F. Makedon, "Task engagement as personalization feedback for socially-assistive robots and cognitive training," *Technologies*, vol. 6, no. 2, pp. 1–17, 2018.

[13] A. Tapus and M. Matarić, "Socially assistive robots: The link between personality, empathy, physiological signals, and task performance." in *AAAI spring symposium: emotion, personality, and social behavior*. Stanford, United States of America: AAAI, 2008, pp. 133–140.

[14] C. Esterwood and L. P. Robert, "Personality in healthcare human robot interaction (h-hri): A literature review and brief critique," in *Proceedings of the 8th International Conference on Human-Agent Interaction*. Virtual Event, NSW, Australia: Association for Computing Machinery, 2020, p. 87–95.

[15] S. Roccas, L. Sagiv, S. Schwartz, and A. Knafo-Noam, "The big five personality factors and personal values," *Personality and Social Psychology Bulletin*, vol. 28, no. 6, pp. 789–801, 2002.

[16] J. Delgadillo, A. Branson, S. Kellett, P. Myles-Hooton, G. E. Hardy, and R. Shafran, "Therapist personality traits as predictors of psychological treatment outcomes," *Psychotherapy Research*, vol. 30, no. 7, pp. 857–870, 2020.

[17] Y. Bai and D. Wang, "Fundamentals of fuzzy logic control — fuzzy sets, fuzzy rules and defuzzifications," in *Advanced Fuzzy Logic Technologies in Industrial Applications*. London, United Kingdom: Springer, 2006.

[18] L. Zadeh, *Fuzzy sets*. Information and Control, 1965, vol. 8, no. 3.

[19] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 15, no. 1, pp. 116–132, 1985.

[20] I. N. da Silva, D. H. Spatti, R. A. Flauzino, L. H. B. Liboni, and S. F. dos Reis Alves, *Artificial Neural Networks: A Practical Course*, 1st ed. Cham, Switzerland: Springer Nature, 2017.

[21] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, United States of America: A Bradford Book, 2018.

[22] A. B. Hostetter and A. L. Potthoff, "Effects of personality and social situation on representational gesture production," *Gesture*, vol. 12, no. 1, pp. 63–84, 2012.

[23] U. Wolfradt and J. Pretz, "Individual differences in creativity: Personality, story writing, and hobbies," *European Journal of Personality*, vol. 15, no. 4, pp. 297–310, 2001.

[24] C. Wang, "A study of membership functions on mamdanitype fuzzy inference system for industrial decision-making," Master's thesis, Lehigh University, Bethlehem PA, 2015.

[25] J. Vittengl and C. Holt, "A time-series diary study of mood and social interaction," *Motivation and Emotion*, vol. 22, no. 3, pp. 255–275, 1998.

[26] A. Lopez-Rincon, "Emotion recognition using facial expressions in children using the nao robot," in *2019 International Conference on Electronics, Communications and Computers (CONIELECOMP)*, 2019, pp. 146–153.

[27] M. W.-R. Ho, S. H.-L. Chien, M.-K. Lu, J.-C. Chen, Y. Aoh, C.-M. Chen, H.-Y. Lane, and C.-H. Tsai, "Impairments in face discrimination and emotion recognition are related to aging and cognitive dysfunctions in parkinson's disease with dementia," *Scientific Reports*, vol. 10, p. 4367, 2020, available: https://www.sciencedirect.com/science/article/pii/S0952197622002251.

[28] M. Munster, "Fuzzy logic based decision making module survey," https://tudelft.fra1.qualtrics.com/jfe/form/SV_ehUK09qhgzXHkyi, accessed: 07-03-2022.

[29] E. van Thiel, "Personality test," https://www.123test.com/personality-test/, accessed: 07-03-2022.

[30] M. Munster, "Reinforcement learning based decision making module survey," https://tudelft.fra1.qualtrics.com/jfe/form/SV_09tvBCdiQBRiD1s, accessed: 07-03-2022.

.

II

Literature Study

(This section has already been grade for the course AE4020 Literature Study)

1
Introduction

According to the [1] the world population is ageing, meaning that the percentage of people over 65 is growing. Consequently, there is a rise in the number of people suffering from dementia. Ferri et al. in [2] estimate the number of dementia patients to double every 20 years, with a total rising up to 81 million by 2040. Providing personalized care to dementia patients is a demanding task for health care workers. An ageing society and increasing number of dementia patients confine care-givers from giving personal one-on-one care. Mierlo et al. in [3] show that personalized care positively affects the well-being of dementia patients. Therefore, the lack of personalized care reduces their quality of life.

Epp in [4] describes person-centred dementia care, an approach to achieve personalized care. In contrast to the traditional culture of dementia care, where the focus lies on the disease and in which individuality is depreciated, person-centred dementia care focuses on the appreciation of individual patients. Care-givers take into account the desires and capabilities of patients to provide care that fits their preferred way of living. Person-centred dementia care improves the quality of life of dementia patients. Furthermore, it reduces agitation among the patients as they are given more freedom in activity scheduling. However, person-centred dementia care is time consuming and expensive [3]. With the current rise in the number of dementia patients, innovative technologies for personalized care are required to reduce the pressure on health-care workers.

An emerging field in robotics, Socially Assistive Robots (SARs), can form a solution to the increase of the number of dementia patients. SARs are robots that assist people through social interactions [5]. To date, their main applications are health care, education and entertainment [6]. But, to the best of our knowledge, SARs have not been used to provide personalized care to improve the quality of life of the patients. Several researches have been performed on personalizing SAR behavior for short-term interactions. However, little research has been performed on personalizing care of SAR's on long-term basis, as is required for person-centred dementia care.

This literature review is written as part of a research looking into a SAR to improve the quality of life of people who suffer from dementia. The robot helps patients by providing personalized and adaptive care in their daily routines. For example, it suggests activities that fit the user's preferences to decrease boredom. This paper conducts a literature review on SARs and methods to achieve personalized care.

1.1. Research Question

Current research on personalized and adaptive SARs in dementia care do not consider long-term interactions. Instead they are limited to short term interactions, such as individual therapy sessions. This creates a gap in literature with the aim of the intended research, which concerns personalization and adaptation for long-term SAR interactions to improve the quality of life of dementia patients.

The focus of this research is to develop the decision making module of a SAR that is able to personalize its behavior and interactions with a dementia patient on the patient's preferences. Taking the gap of long-term SAR personalization into consideration, the main research question is as follows:

How can a decision making module be developed for a Socially Assistive Robot by means of techniques from control theory and Artificial Intelligence, such that the robot personalizes its decision making towards long-term personality traits of dementia patients and exhibits adaptive behavior to handle changes in the needs and emotional status of patients?

More information about control methods suitable for personalization and adaptation must be gathered to be able to create the proposed SAR decision making module. Therefore, the following sub-questions should be addressed to be able to answer the main research question:

1. Which novel control methods exist for personalization of SAR behavior?
2. Which novel control methods exist for adaptation of SAR behavior?
3. Which user attributes and personality traits should be considered in the adaptation and personalization of the SAR behavior?

1.2. Research Objective

With the research questions established, the following research objective is introduced:

Develop the decision making module of a Socially Assisitive Robot for people with dementia, which allows the robot to display personalized and adaptive behavior by means of intelligent robot control.

This objective can be divided into several smaller sub-objectives:

1. Develop a controller that is able to personalize the behavior of a SAR with respect to the long-term personality traits of a person with dementia.
2. Develop a controller that is able to adapt the behavior of a SAR to the short-term variations in the needs and emotional status of a person with dementia.
3. Implement the developed decision making module in the humanoid robot Nao.

Nao is a human-like robot equipped with several features that allow for human robot interaction. Moreover, the Nao robot is fully programmable, thus it allows for easy implementation of the developed decision making module.

1.3. Report Structure

This literature review covers literature on SARs and techniques from control theory and Artificial Intelligence (AI) suitable to achieve personalization and adaptation. Chapter 2 contains literature on SARs. It starts by elaborating on the concept of SARs, in which the definition of SARs, challenges in the design of SARs and the current applications of SARs are discussed. The concept is followed by literature in which personalization and adaptation of SARs are used. Chapter 3, 4 and 5 cover the theory behind novel control methods for personalization and adpatation of robot behavior. Moreover, they include literature in which these control methods are applied in SARs. The control methods that are considered are; Fuzzy Logic Control, Artificial Neural Networks and Reinforcement Learning for chapter 3, 4, and 5 respectively. Chapter 6 discusses the proposed solution for the controller for personalize and adaptive robot behavior for people with dementia.

2

Socially Assistive Robots

This chapter explains the concept of Socially Assistive Robots (SARs) by means of existing literature. It gives the formal definition of SARs, challenges that exist for designing SARs and their applications. Furthermore, it covers literature on personalization and adaptation in SARs.

2.1. Definition of Socially Assistive Robots

SARs can be seen as the intersection between Socially Interactive Robots (SIRs) and Assisitive Robots (ARs) [5]. ARs are robots that help human users in various ways, including physical tasks and memory tasks. SIRs are robots whose tasks main task is social interaction, via gestures or speech for example. SARs are a combination of ARs and SIRs; their purpose is to assist people through social interaction.

2.2. Challenges in Socially Assistive Robots

Ordinary robotics face challenges such as navigation and decision making, SARs also face challenges arising from social interaction between the human user and the robot [6].

Fist off all, the robot must understand feedback provided by the user. Humans have different forms of communication [7]. The most important communication methods are speech and non-verbal communication through gestures and facial expressions. However, also small social cues, such as the way someone dresses, can be very informative. For smooth interaction between a user and a robot, the robot should be able to recognize and understand these forms of communication. This enables the robot to determine the state or intention of the user and to display the desired behavior.

Secondly, the communication from the robot towards the user should be understandable and in line with the user's expectations. The most common forms of communication for SARs are speech and gestures. The embodiment of a SAR determines which communication types can be used. For example, [8] uses a SAR that consists solely of a head, this restricts it from using communication methods such as hand gestures. Moreover, the embodiment of a SAR inflicts an expectation on the user. For example, a person treats a pet robot differently than a humanoid robot [6].

Finally, the robot should have real time performance. If processing of received data and decision making take a lot of time, the interaction (if realized at all) is not considered pleasant by the user. Therefore, the robot must be able to perform in real time to allow for smooth interactions with humans.

2.3. Applications of Socially Assistive Robots

In literature SARs assist people in various ways. These applications differ in form and intention of the interaction between the robot and humans. The robots are used in short-term interactions to provide assistance for therapists in individual therapy sessions or to provide help for children in short educational games. However, the robots may also be used on a long-term basis, for example to assist people in daily life or for entertainment purposes [6]. This section elaborates on different applications of SARs found in literature.

2.3.1. Socially Assistive Robots in Therapy Sessions

SARs can be used as therapy robots, for example in therapy for dementia patients, therapy for children with Autism Spectrum Disorder and in post-stroke rehabilitation therapy.

Socially Assistive Robots in Dementia Therapy
A review on the use of SARs in elderly care [9] mentions that SARs in elderly care are mainly categorized in two types: Service robots and Companion Robots, although some robots can be categorized to both types. Companion robots can be used in therapy sessions. Therefore, for therapeutic robots for dementia patients only the companion robots are considered, the service robots are discussed in subsection 2.3.3.

The intention of companion robots is to create a bond between user and robot, similar to the way people bond with their pets. Therefore, these robots often have an animal-like appearance. Generating companionship with the robots yields in health benefits and improves the psychological well-being of patients [9].

A widely used therapy robot is the seal robot Paro. In [10] and [11] Paro is used in intervention sessions with a therapist or research assistant. During the sessions, patients are given Paro and their interactions with Paro are recorded. Both studies show promising results. Marti et al. in [10] show patients engage with Paro and have the feeling they have to take care of it. The patients seem less stressful when interacting with Paro. Finally, Paro enables the therapist to establish communication with the patient, something that can be difficult without the pet robot. Moyle et al. in [11] show varying results, some patients respond positively towards Paro, whereas others have a negative attitude towards it. Furthermore, individuals respond differently in different sessions. Therefore, the study concludes that a single approach does not improve the quality of life of all patients and that different approaches at different times are necessary for an individual patient. Sabanovic et al. in [12] uses Paro in group sessions to investigate the effect of Paro on the person holding it (primary interactor) and on the people surrounding it (non-primary interactor). The results of the study show that Paro does not only stimulate interactions between the robot and the patient, but also contributes in interaction between participants. Especially non-primary interactors show more activity, for example by looking at and talking to the person interacting with Paro. However, also the primary interactor shows more engagement towards the people around him or her.

Another companion robot is NeCoRo, a cat-like robot. Libin and Cohen-Mansfield in [13] compare the effects of the robotic cat on dementia patients with the effects of a regular plush cat toy. The results show that both the robotic cat and plush cat decrease the patients' agitation levels and increase the patients' pleasure and interest levels. Of the two cats, the plush toy leads to the largest reduction in agitation amongst the dementia patients. The patients prefer to hold the plush toy cat over the robotic one and holding the cat has a more calming effect than interacting with it from a distance. The robotic cat leads to a more significant increase in pleasure and interest compared to the plush toy cat. In terms of engagement, the robotic cat has a small positive effect on the engagement of the patients but its influence is not significant.

Companion robots are not the only type of SARs that are used in therapy for people with dementia. Tapus et al. in [14] use a SAR in musical therapy for dementia patients. The aim of the robot is to keep the patients engaged in a game called *Name That Tune*, where the patients have to recognize a song, press the button corresponding to the song, mention the song's name and sing along. The robot provides assistance by explaining the game to the patients and by giving hints and encouraging comments. Furthermore, the robot can change the difficulty of the game to the performance of the user to keep him or her engaged. The results show that the performance of the patients, measured by the number of correct answers and the time patients take to give a response, has improved during the experiment. Moreover, caregivers have observed the patients to become more engaged with the robot and in the game throughout a series of sessions. This shows that SARs can be used in cognitive therapy to prevent deterioration of the abilities of people with dementia.

Socially Assistive Robots in Autism Spectrum Disorder Therapy
Besides dementia therapy, SARs are also used in therapy for children with Autism Spectrum Disorder (ASD). Common characteristics for children with autism are: difficulty developing social skills, a preference for repetitive behaviors and having stereotypical interests. Therefore, robotic therapy can be beneficial to children with ASD. Robotic interaction is more predictable than interaction with a human therapist [15]. Furthermore, children are more engaged with robot therapists, especially children with autism. Bekele et al. in [16] conduct an experiment on joint attention skills of children with ASD. In the experiment children, aged between 2 and 5 years old, must coordinate their attention between a social partner and their environment with the help of attention prompts, for example gestures or audiovisual stimuli. The social partners are realized by a humanoid robot and a real human therapist. The experiment also includes a control group of children without ASD as a reference for the effectiveness of the system. The results show that both groups look significantly more to the robotic therapist than to the human therapist. However, the difference between looking at the robot and looking at the human is larger for the group of children with autism. Furthermore, children require more prompts to shift their attention from the robot to the target than from the human therapist to the target. This can be a sign of a large interest in the robot. For these two reasons, it is concluded that SARs have a positive effect on the engagement of children in autism therapy.

Clabaugh et al. in [17] focus not only on developing the social skills of children with autism, but also their educational skills. It conducts a long-term experiment where a SAR helps young children, aged between 3 and 7 years old, suffering from ASD with mathematical problems. The SAR is able to adapt its feedback and the difficulty of the problems to personalize the learning process for every individual child. The results show that every child, apart from one child who has not completed the experiment, has improved his or her mathematical skills significantly. Furthermore, the system is successful in adapting the feedback and difficulty to the individual users.

Socially Assistive Robots in Post-Stroke Rehabilitation Therapy
Finally, SARs can be used in post-stroke rehabilitation sessions. In contrast to regular AR's in rehabilitation therapy, SARs often do not help the patient physically but mentally, with encouragement for example. Matarić et al. in [18] perform a preliminary research on the effectiveness of SARs in post-stroke rehabilitation therapy. In the experiment patients are asked to perform two tasks: A magazine shelving task and a voluntary task that requires arm movement. During the tasks, the robot provides the patients with feedback through sound effects, either with a synthesized voice or a human voice. The results show that the robot therapist is well received among the patients and influenced their motivation to exercise positively. In [19] Tapus and Matarić use a SAR for post-stroke rehabilitation patients, where the type of encouragement is based on the level of extraversion of the patients. A more detailed discussion of this research is given in subsection 2.4.1.

2.3.2. Socially Assistive Robots as Entertainment
Another application of SARs is entertainment. Various toys which are able to interact with their user are developed for the entertainment of people [6].

First of all, there is the doll My Real Baby, designed by iRobot. The doll is equipped with actuators in its face, allowing it to have facial expressions. It can generate sounds using a speaker and senses with pressure sensors and an accelerometer. The interactive behavior of the doll is based on the measurements from these sensors. Finally, the doll develops itself like real babies do. Shortly after purchase it can only produce short infant sounds, later on it makes more advanced and lengthy sounds. Other socially interactive dolls are Amazing Babies, My Dream Baby and Miracle Moves Baby.

Secondly, pet-like SARs are used for entertainment. The companion robots from subsection 2.3.1 can also be used for entertainment of non-dementia patients. Fujita and Kitano describe the design of AIBO in [20]. AIBO is a socially interactive dog designed for entertainment, although it is also used in dementia care [21]. The strength of pet-like SARs is that they provide entertainment for people in various ways, e.g., peoply may have fun interacting with the robots and also by observing robot while it plays with a ball or while it learns to walk. Some key features of AIBO are its ability to chase bright colored objects and object avoidance, which are made possible by infrared-sensors and a vision chip enabling color segmentation. Furthermore, AIBO is equipped with pressure sensors and accelerometers to detect petting and falling. Other companion robots which also can be used as entertainment are Poo-Chi, I-Cybie, Me and My Shadow and NeCoRo, the cat-like robot mentioned before.

2.3.3. Socially Assistive Robots as Service Robots

A third application for SARs is service. Service robots assist the user with a specific task. Example tasks of service robots are welding for industrial service robots and household chores for service robots in health care. Often these tasks can be performed with general AR's. However, AR's generally take over the full task, whereas SARs can be used to only give mental feedback with the task. The user still completes the task on his or her own, which gives a feeling of independence. Furthermore, social interactions with the robot make it more user friendly, through which the user feels more comfortable with the robot [6].

Moro et al. in [22] perform research on a SAR that helps people with cognitive impairments, such as dementia, in daily life activities. The robot is taught behaviors necessary to assist patients in daily life. Example behaviors are inviting the patient to make tea, engage the patient in case he or she is distracted or correct the patient in case it performs an incorrect step. Afterwards, the robot must learn to apply the right behavior at the right time. The robot successfully learns the behaviors for tea making and when to apply them based on user models. In future research, real-life user studies are performed to see whether the robot functions well on real patients and additional activities are investigated.

In [23] the research behind a SAR named Pearl is considered. Pearl is a SAR serving as assistant in a nursing home. The robot reminds the residents of events and guides them through the nursing home. Moreover, the robot has a speech recognition module enabling it to inform the user about the weather, time or tv-shows when requested. In an experiment the robot is implemented in a real nursing home. The residents are uniformly positive about Pearl, it succeeds in reminding the elderly of scheduled appointments, in guiding the residents around the facility when walking assistance is required and in interacting with the users.

The research on CERO, a robot for assistance in fetch and carry tasks, is discussed in [24]. The robot can help people with a motion impairment to fetch or deliver objects from specified locations. The paper considers CERO to be a SAR as the user can operate the robot by means of speech and CERO can give feedback about the understanding of a task using gestures and spoken feedback. However, the interactions between CERO and the user remain shallow and are only on the topic of the fetch and carry task. Moreover, CERO for some fetching tasks CERO relies on people in its environment to place the object its transportation tray. Therefore, CERO is not considered a SAR in this literature review but an ordinary AR that is operated by voice.

2.4. Personalization and Adaptation in Socially Assistive Robots

In general, the researches discussed in section 2.3 show that people respond positively towards SARs. Also SARs that are applied in dementia care are usually well received among the patients. However, Moyle et al. in [11] conclude that there are differences in the responses between patients and that individuals respond differently at different times. Therefore, one approach is not the optimal solution for all patients. For the SAR to be effective for all patients, it must be able to personalize its behavior with respect to an individual patient and adapt its behavior to variations in the emotional status of this patient.

In literature personalization and adaptation are often interchanged. This literature review considers personalization as generating a robot behavior in line with the general preferences and capabilities of a single user. Personalization forms the basis of the robot behavior. Adaptation then adapts this behavior to variations in the user's preferences and developments in the user's capabilities. Variations in the preferences of the user can be caused by mood changes, developments in the user's capabilities relate to progression the user makes in a certain task. The following sections cover the state-of-the-art research on personalization and adaptation of SAR behavior.

2.4.1. Personalization in Socially Assistive Robots

In [19] the behavior of a SAR is personalized based on the personality of a particular user. The SAR is used in post-stroke rehabilitation therapy, where it encourages the patient with verbal cues, its activity level and its proxemics, i.e., the distance it maintains towards the patient. Based on the level of extraversion of the patient, the robot changes the type of feedback it gives. For example, for extravert patients the robot gives comments like 'You can do it!' on a high pitch and volume. Whereas for introvert patients the robot gives comments like 'I know it's hard, but remember it's for your own good' on a low pitch and volume. Furthermore, the SAR can change its distance to the user and its activity level through the speed and amount of movements it makes. The robot displays higher activity levels towards extrovert people. Results of the experiment show that people prefer the robot behavior matching their personality instead of the opposite behavior.

2.4.2. Adaptation in Socially Assitive Robots

Adaptation of SAR behavior is covered in [14], where the difficulty of the game *Name That Tune* is changed based on the performance of the user. The robot can assist the patient in three ways: Not giving a hint (difficult), reminding to push a button (medium) and telling which button to push (easy). The SAR must adapt the difficulty of the game towards the development of the user. This development is measured by the number of correct answers and the time the user takes to provide the answers. Results show that the SAR successfully adapts the game difficulty to the progress of the users. Initially, a user starts on the easy level where the robot tells which button to choose. However, the SAR increases the difficulty as the user improves. When a new level turns out to be too difficult for the user, the SAR returns to the previous level. Eventually, the user performs consistently on the most difficult level, where no hints are given.

In [16] children with autism have to perform a joint attention task. In this experiment the SAR inlcudes different prompt levels to shift the attention of the child towards the target. The system can adapt the prompt level depending on whether the previous level was successful in shifting the attention. In case the child does not shift his or her attention towards the target, the prompt level is increased. The different prompts that the SAR can use are head shifting towards the target, pointing towards the target, giving vocal cues to shift the child's attention to the target and playing sounds, displaying pictures or videos at the target location.

2.4.3. Both Personalization and Adaptation in Socially Assistive Robots

Personalization and adaptation are used together in [25] to keep a user engaged in a sequencing task. User models are used for personalization of the SAR behavior and human feedback is used for adaptation of the SAR behavior. The user models contain information on which vocal feedback to give (encouraging or challenging) and on when to increase the difficulty level of the sequencing task depending the level of engagement of the user. The user models are created for different user skills under the different difficulty levels. Real-life performance data, collected in performance and engagement measurements from a group of 69 computer science students, is used for the creation of the user models. In future work these user models are used for the personalization of the SAR behavior. Before the sequencing task, the system performs a user skill assessment, in which the task performance and engagement of the user under different difficulty levels is determined to select the corresponding user model. The user model forms the personalized basis behavior of the SAR. Throughout the active phase of the sequencing task, the SAR adapts its behavior using feedback from the user. This adaptation process is achieved with a method called interactive reinforcement learning, which is covered in chapter 5.

Clabaugh et al. in [17] describe the design of a SAR providing assistance to children with autism during various mathematical games. Similar to [25], the robot can adjust the difficulty of the games and the level of feedback it gives. The experiment is held over the course of a month and 10 different mathematical games are used in the experiment. In the beginning the system has to find the right level of challenge and feedback for the user in the different kinds of games. It learns the right levels by observing the number of mistakes and the number of help requests that the user makes in the different kinds of games. Once an initial behavior was established for the child, the system had to adapt the difficulty of the game and the level of feedback to the progress that the child made in the different games.

3
Fuzzy Logic Control

The purpose of the research is to develop a decision making module for a SAR that can personalize its behavior to the long-term personality traits of dementia patients and that can adapt this behavior to short-term variations in the patient's emotional status and the environmental status. The following chapters discuss different techniques from control theory and Artificial Intelligence (AI), that are suitable for personalized and adaptive behavior in SARs. This chapter considers fuzzy logic control, a rule-based control method that relies on a set of if-then rules for the decision making. The reasoning principles of fuzzy logic control are similar to the reasoning process of humans. Ambiguities in the reasoning and thinking procedures mainly caused by linguistic terms people use to describe real-life phenomena form challenges in classic control methods. However, fuzzy logic control uses fuzzy sets to handle these ambiguities [26]. This chapter explains the main idea of fuzzy logic control and discusses various implementations of fuzzy logic control in the field of SARs.

3.1. Main idea of Fuzzy Logic Control

Fuzzy logic control is a rule-based control method, relying on fuzzy logic. This section explains the fuzzy process consisting of:

1. Fuzzy sets
2. Fuzzification
3. Inference
4. Defuzzification

3.1.1. Fuzzy Sets

Unlike classical control methods, fuzzy logic control can handle ambiguous data: where classical controllers rely on crisp data, such as numerical values, fuzzy logic control can also handle ambiguous data, like linguistic terms [26]. For example, consider a controller for controlling the room temperature, classic rule-based controllers use rules in the following form: *If the temperature rises above 25° set the motor of the air-conditioning to 100 rotations per minute.* However, a rule in a fuzzy logic controller looks like: *If the temperature is high, rotate the motor of the air-conditioning fast.* The rule in the fuzzy logic controller is more like the way people think and reason. This makes it way more ambiguous than the classical controller, which uses crisp values for actuation. Fuzzy logic control handles these ambiguities using fuzzy sets [26].

In ordinary sets, an element either belongs to or does not belong to a set, there is no situation in between. Fuzzy sets allow an element to belong to a set to a certain degree that can vary from 0 to 1. Bai and Wang in [26] give an example of a group of faculty members to clarify the difference between ordinary sets and fuzzy sets. The group of faculty members consists of 10 people aged between 28 and 61. This group is displayed in Figure 3.1 with their age indicated on the x-axis. From the group of faculty members, a set is created with young faculty members. For an ordinary set with young faculty members, a threshold age must be set to determine whether someone is young or not. For example, people aged below 40 are considered young, i.e., only the first three members belong to the set of young faculty members. However, this is different for a fuzzy set, which allows partial membership to the set. Therefore, if the set of young faculty members is defined as a fuzzy set, then the people who are aged below 40 can be full members of the set with a membership degree of 1, but people above 40 can still be a partial member of the set. The membership of the faculty members decreases as they are older. Hence, faculty member x_4, who is 42 years old, still has a membership to the set young faculty members with a degree of, e.g., 70% (see Figure 3.1, which shows the membership function of the sets for young faculty members). Faculty member x_5, who is 49 years old, is also still part of the set, but with a membership of 30%. This example shows that the transition of belonging or not belonging to fuzzy sets is smoother than the transition for ordinary sets.

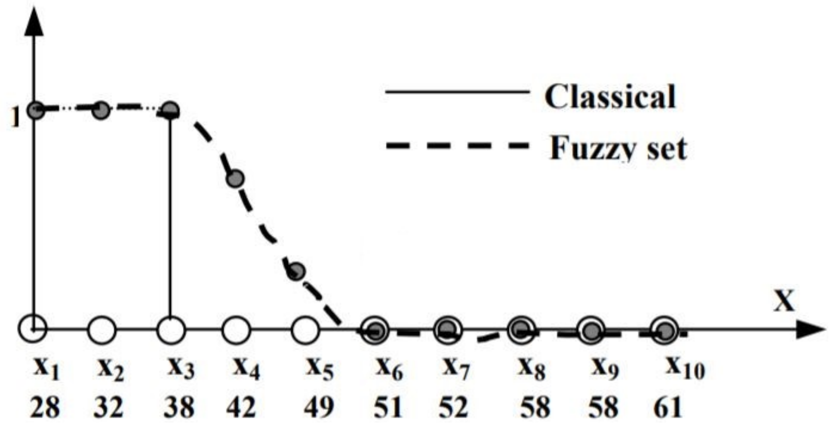

Figure 3.1: Group of faculty members and the membership functions corresponding to a classical vs. a fuzzy set [26]

The extent to which an element is part of a fuzzy set is described by membership functions. The designer of the controller sets the shape of the membership function of a fuzzy set. Example shapes for the membership functions are given in Figure 3.2 [27].

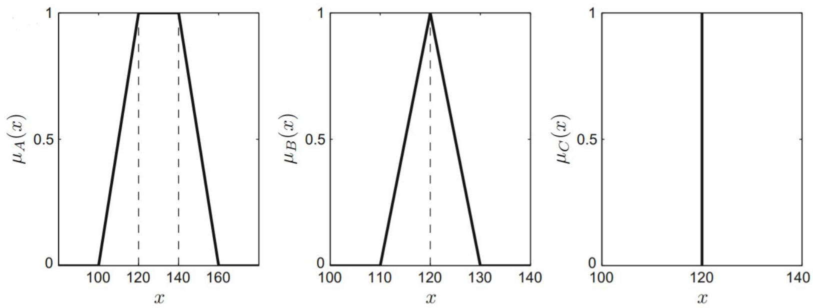

Figure 3.2: Common shapes for a member function [27]

Similar to ordinary sets, operations such as complement, union and intersection can also be extended and applied to fuzzy sets. Generally speaking, the complement of a set is its opposite set. Any element that does not belong to the set, belongs to its complement. This is visualized in Figure 3.3. When having a fuzzy set, A, its complement, $\bar{A}$, is given by (3.1) [27]:

$$\mu_{\bar{A}}(x) = 1 - \mu_A(x) \tag{3.1}$$

Where $\mu_A(x)$ is the membership degree of the element x to the fuzzy set A.

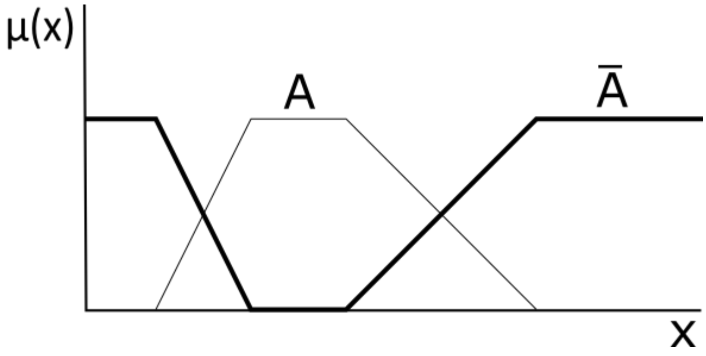

Figure 3.3: The complement, $\bar{A}$ of fuzzy set, A

The intersection of two sets in general concludes the elements that are present in both sets. For two fuzzy sets the intersection is the minimum of the membership degrees of every element in these two sets. The intersection, $C = A \cap B$, is illustrated in Figure 3.4 and is given by (3.2) [27]:

$$\mu_C(x) = \min\{\mu_A(x),\ \mu_B(x)\} \tag{3.2}$$

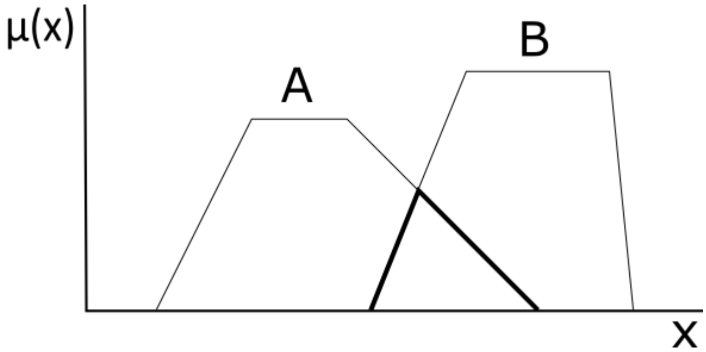

Figure 3.4: The intersection of fuzzy sets A and B

The union of two sets is in general the collection of all elements in these sets. For two fuzzy sets the union is the maximum of the membership degrees of all elements in these two sets. The union, $C = A \cup B$, is illustrated in Figure 3.5 and is given by (3.3) [27]:

$$\mu_C(x) = \max\{\mu_A(x),\ \mu_B(x)\} \tag{3.3}$$

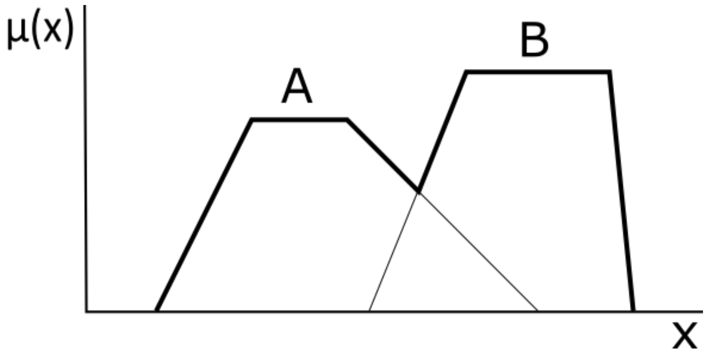

Figure 3.5: The union of fuzzy sets A and B

When having fuzzy sets in multiple dimensions, the operations projection and cylindrical extension become relevant. Projection is a reduction in the number of dimensions of a fuzzy set, cylindrical extension is increasing the dimensions of a fuzzy set. Take a fuzzy set, A, in domain $X \times Y$ with $X = \{x_1, x_2\}$ and $Y = \{y_1, y_2\}$. A can be defined as in (3.4):

$$A = \{\mu_1/(x_1, y_1), \mu_2/(x_1, y_2), \mu_3/(x_2, y_1), \mu_4/(x_2, y_2)\} \tag{3.4}$$

In this representation, μ_1 corresponds to the membership degree of the set on point (x_1, y_1), μ_2 corresponds to the membership degree of the set on point (x_1, y_2), etc. The projection of a fuzzy set in a certain domain is performed by taking the maximum of the set's membership degree at the points belonging to this domain. In the example given by (3.4), the membership degrees μ_1 and μ_2 correspond to x_1, the membership degrees μ_1 and μ_3 correspond to y_1, etc. The projections of fuzzy set, A, in, respectively, domains X and Y are given by (3.5) and (3.6) [27] and are illustrated in Figure 3.7:

$$\text{proj}_X(A) = \{\text{max}(\mu_1, \mu_2)/x_1, \text{max}(\mu_3, \mu_4)/x_2\} \tag{3.5}$$
$$\text{proj}_Y(A) = \{\text{max}(\mu_1, \mu_3)/y_1, \text{max}(\mu_2, \mu_4)/y_2\} \tag{3.6}$$

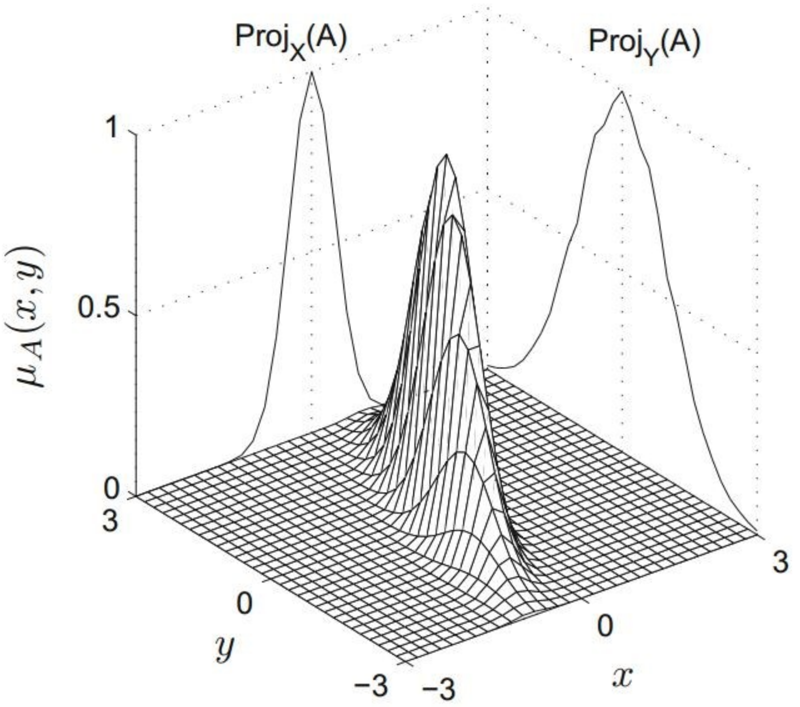

Figure 3.6: Projection of fuzzy set A in domains X and Y [27]

Cylindrical extension increases the dimensions of a fuzzy set. Consider the fuzzy set, A, in domain, X, with $X = \{x_1, x_2\}$ as given in (3.7). The extension of A to the domain Y with $Y = \{y_1, y_2\}$ is performed with (3.8) [27]:

$$A = \{\mu_1/(x_1), \mu_2/(x_2)\} \tag{3.7}$$
$$\text{ext}(A) = \{\mu_1/(x_1, y_1), \mu_1/(x_1, y_2), \mu_2/(x_2, y_1), \mu_2/(x_2, y_2)\} \tag{3.8}$$

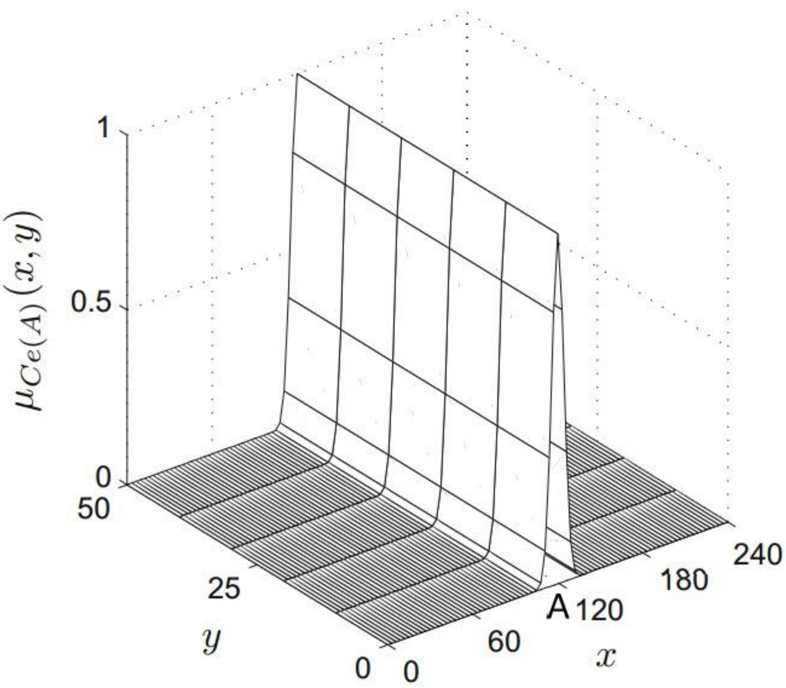

Figure 3.7: Extension of fuzzy set A to domain Y [27]

3.1.2. Fuzzification

Fuzzy sets form the basis for fuzzy logic control. However, further steps are required to make fuzzy sets useful for control applications. Fuzzification is the first step in the process of fuzzy logic control. It is the process of converting real-life data into fuzzy variables and it consists of representing the input variables by linguistic terms and the derivation of the membership functions [26].

Consider the example of room temperature control. The input variable to the controller is the room temperature in this case. If the controller classifies the temperature in three categories, it requires three linguistic terms: Low, Medium and High for example. A possible distribution of the membership functions for these three categories is shown in Figure 3.8. It can be seen that the categories may overlap, which means that some temperatures can belong to multiple fuzzy sets. For example, a temperature of 17° belongs to the set low and to the set medium, these overlaps may lead to smooth transitions between the two fuzzy concepts low and medium [26]. Examples of commonly used membership functions are triangular, trapezoidal and singleton membership functions displayed in Figure 3.2. However, the membership function can also be designed uniquely for a specific controller. The choice of membership function depends mainly on the application of the controller. For instance, triangular and trapezoidal functions are suitable for systems that require large dynamic variations in a short time period, whereas bell shaped functions are better when high control accuracy is required [26].

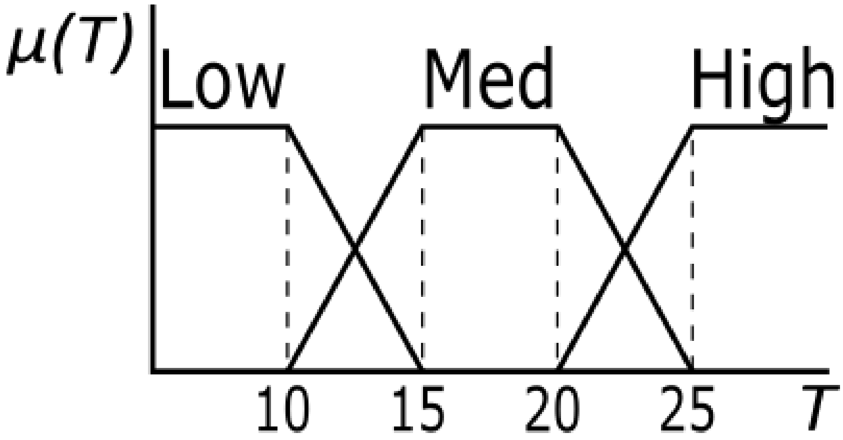

Figure 3.8: Example of fuzzification of the variable temperature in three linguistic terms

To summarize, in the fuzzification step, the crisp input variables should be converted into fuzzy sets (which represent linguistic terms). Each fuzzy set corresponds to a membership function that mathematically represents the resulting fuzzy input variable.

3.1.3. Inference

Fuzzy inference requires fuzzy inputs and a rule base consisting of logical if-then statements that may involve fuzzy sets. After the fuzzification procedure, the fuzzy output data is determined based on the fuzzy rules and the fuzzy input data via fuzzy inference [26].

The first part of the inference process is setting up the rules of the fuzzy rule base. These rules are simple if-then rules, which show the relation between the input and output sets. The rules are described with the linguistic terms corresponding to the input data. The use of linguistic terms makes fuzzy logic control similar to the way humans think and reason [26].

When the inference rules are set up, the output fuzzy sets can be determined from the input fuzzy sets. One approach to do this is the Mamdani inference, also called the max-min inference. This approach is summarized in three steps [28], shown in Figure 3.9. The figure displays a model with three rules:

1. If x is A_1, then y is B_1
2. If x is A_2, then y is B_2
3. If x is A_3, then y is B_3

Here x is the input variable, A_1, A_2 and A_3 are the fuzzy sets corresponding to the input linguistic terms, y is the output variable and B_1, B_2 and B_3 are the fuzzy sets corresponding to the output linguistic terms. For the given example, the input data, x, is represented by the fuzzy set A'.

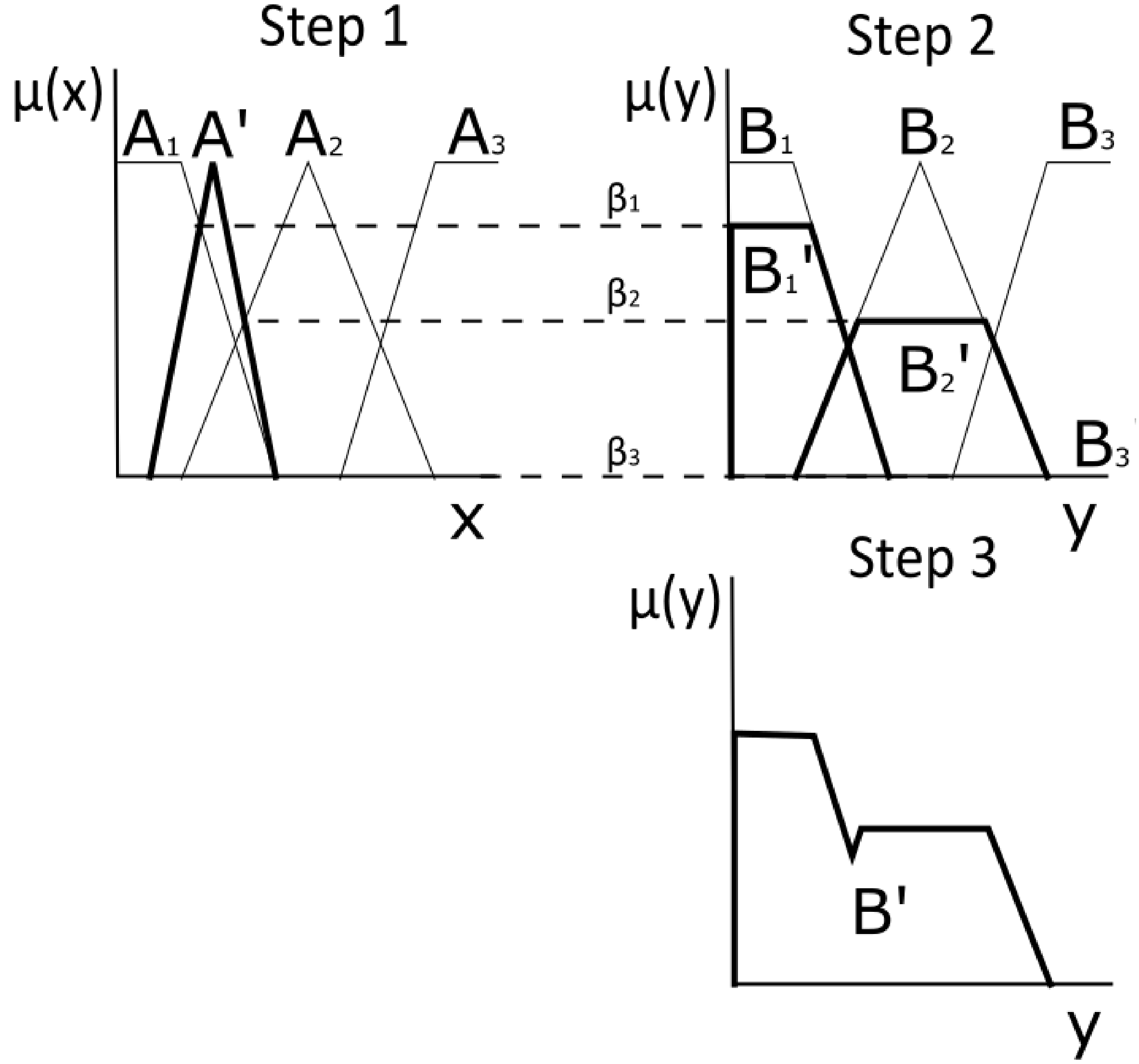

Figure 3.9: The Mamdani inference process

The operations from subsection 3.1.1 are used in the inference process. The first step is to determine to what degree the individual rules are fulfilled. The fulfillment, β, is calculated according to (3.9):

$$\beta_i = \max\{\mu_{A'}(x),\, \mu_{A_i}(x)\} \tag{3.9}$$

The i in β_i and $\mu_{A_i}(x)$ stands for each individual rule, where i for the given example can be 1, 2 or 3. The top left picture in Figure 3.9 shows the calculation of the fulfillment per rule. For example, rule 3 has a fulfillment of 0. Step two of the inference process is to derive the fuzzy output sets per rule as a result of their fulfillment. This is shown in the top right of Figure 3.9 and is done via (3.10):

$$\mu_{B_i'}(y) = \min\{\beta_i,\, \mu_{B_i}(y)\}\forall y \tag{3.10}$$

That is for every individual rule the minimum of the fulfillment and output membership function should be computed. Finally, from the output sets per rule (note that we refer to the sets, whereas the computations are performed on the corresponding membership functions) the membership function corresponding to the overall fuzzy output set can be determined, which is shown in the bottom left of Figure 3.9 and done via (3.11):

$$\mu_{B'}(y) = \max_{\forall i \in \{1,2,3\}}\{\mu_{B_i'}(y)\} \tag{3.11}$$

3.1.4. Defuzzification

With the fuzzy output set determined in the inference process, the only thing left to do is converting the fuzzy output set back to a crisp value for the actuators of the system. This is done in the defuzzification step. Defuzzification can be performed with various methods, where the commonly used methods are the center of gravity method (COG) and the mean of maxima (MOM) [26], shown in Figure 3.10.

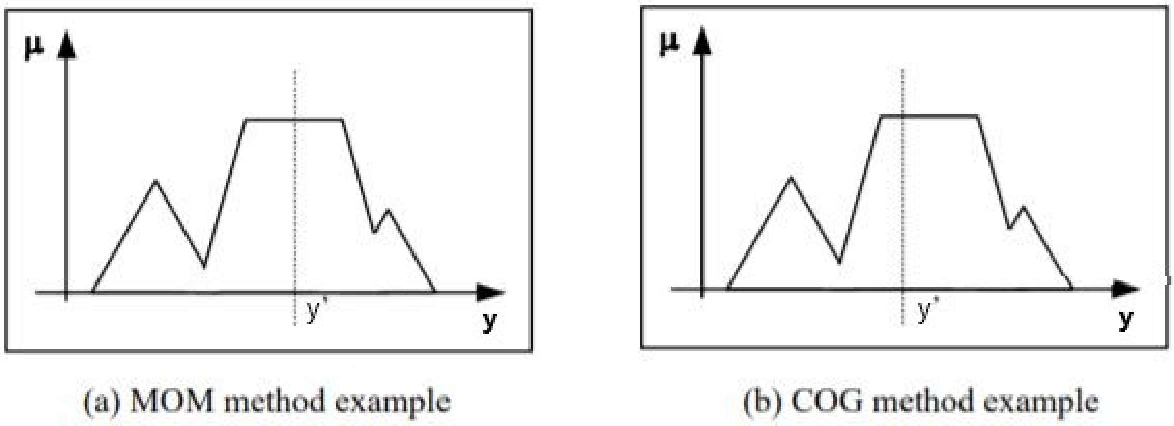

Figure 3.10: Commonly used methods for defuzzification [26]

The COG method calculates a weighted average based on the membership degrees of the output set over the elements in the output fuzzy set. The crisp value is calculated with (3.12) [26]:

$$y' = \text{COG}(B') = \frac{\sum_{i=1}^{n} \mu_{B'}(y_i) \cdot y_i}{\sum_{i=1}^{n} \mu_{B'}(y_i)} \tag{3.12}$$

Here n is the number of elements y_i in the domain Y. In case of a continuous domain Y, the integral of the membership should be used. The MOM method calculates the mean value of the elements with the highest membership degree. The MOM method can be described by (3.13) [26]:

$$y' = \text{MOM}(B') = \frac{\sum_{i=1}^{n} \arg\max_{y_i} \left(\mu_{B'}(y_i)\right)}{n} \tag{3.13}$$

In this equation only the elements with the maximum membership are considered. In the equation y_i are the elements with maximum membership and n is the number of elements with the maximum membership. A drawback of the MOM method is that it does not represent the complete shape of the output fuzzy set. Hence, different fuzzy outputs can produce the same crisp value when the maximum elements are the same [26].

3.2. Fuzzy Logic Control in Socially Assistive Robots

Little research is performed on fuzzy logic controllers as the sole control element in SARs. Therefore, in this section we have also included research that combines fuzzy logic control with other control methods.

One application of fuzzy logic control in the field of SARs is emotion expression, which has been discussed in [29]. In this paper fuzzy logic control is used to generate facial expressions, allowing the robots to express emotions like fear, surprise and anger. The robot displays reactive behavior on people approaching it. Depending on the distance and on the approaching speed of the person, the robot feels a different mixture of the three emotions and shows a different facial expression. For example, when a person blocks the SAR by standing still in front of it, the SAR shows an angry face. The SAR uses two sets of fuzzy rules to express its emotions. One set of rules is used to map the events (the distance of the person and the approaching speed) to the emotions the SAR feels. The second set of rules maps these emotions to facial expressions. The rules of this set contain information on how to actuate the motors in the robot's face according to which extent a certain or a mixture of emotions should be displayed. Enabling a SAR to express emotions is beneficial for the interaction between a user and a robot. Fuzzy logic control is a suitable approach for the emotional expression as it has an easy implementation and the design is easy to understand even for non-experts. Moreover, it allows for smooth transitions of emotions, which makes the robot behave almost like a human. Finally, fuzzy logic control allows the robot to express a mixture of emotions if this is required.

A combination of neural networks and fuzzy logic control is investigated in [30]. The two control methods are combined to determine appropriate proxemics, i.e., the distance the robot maintains with respect to the user. The proxemics are an important feature to make a person feel comfortable when interacting with the robot. The SAR must adapt the proxemics to the user's personality and current activity. People with different personalities differ in the proxemics they prefer to maintain. Furthermore, the preferred proxemics of a person depend on the activity that this person is performing. Therefore, the robot in this experiment adapts its proxemics to the personality of the user and on which activity the user is performing. The decision making module of the SAR exists of two layers. The first layer is used to assess which type of activity the user is performing. This layer relies on a Naive Bayes classifier to classify the user's activity based on sensor data captured with a smart watch worn by the user. The four activities considered in the experiment are lying, sitting, standing and walking. The second layer adapts the proxemics of the robot to the Big Five personality traits of the user and the activity that has been selected in the previous layer. The adaptation process relies on a combination of fuzzy logic control and artificial neural networks. It uses fuzzy rules, where the output of the rules contains parameters that can be trained with a neural network. An example rule of the adaptation layer has the following structure:

$$\text{If } x \text{ is } A_1 \text{ and } y \text{ is } B_1, \text{ then } f_1 = a_1 \cdot x + b_1 \cdot y + c_1$$

In this rule a_1, b_1 and c_1 are parameters that can be changed. These parameters are updated with a neural network, which uses training input data and desired output data to select the right parameters. The input data x and y can be the personality traits of the Big Five personality inventory and one of the four activities which has been identified in the classification layer. Thus, the system in this paper is not an ordinary fuzzy logic controller but a fuzzy logic controller trained with a neural network.

4 Artificial Neural Networks

Another intelligent control method that can be used for personalized and adaptive SAR behavior is control with artificial neural networks. Fuzzy Logic Control requires the designer to describe the physics of the system with a set of if-then rules, whereas artificial neural networks are based on a model free approach. More specifically, no knowledge of the underlying physics of the system is required for a decision making module relying on artificial neural networks. In this case, instead of the rules given to the system by the designer, the artificial neural network is trained based on a set of input-output data that has been gathered from the controlled system. This chapter explains the working principle behind artificial neural networks and how they are used in SARs.

4.1. Working Principle of Artificial Neural Networks

The aim of artificial neural networks is to mimic the functioning of the brain in a simplified fashion. Artificial neural networks can be used to control complex and non-linear systems without requiring prior knowledge of the exact physics of the system. Different types of artificial neural networks exist, but they all rely on the same principles. The artificial neural network is built with a set of neurons, where each neuron receives an input value, processes this input with a function and produces an output value. Every neuron is connected to at least one other neuron, meaning that the output of the first neuron is the input to the consecutive neuron. The connection between every pair of neurons is specified by a weight, which shows the importance of the connection within the artificial neural network. The artificial neural network learns to approximate the desired output data by changing the weights until satisfactory outputs are produced [31].

4.1.1. Artificial Neural Network Structure

Many artificial neural network types exist. However, the most typical form of artificial neural networks is the feed forward artificial neural network [32]. A feed forward artificial network with two hidden layers is illustrated in Figure 4.1.

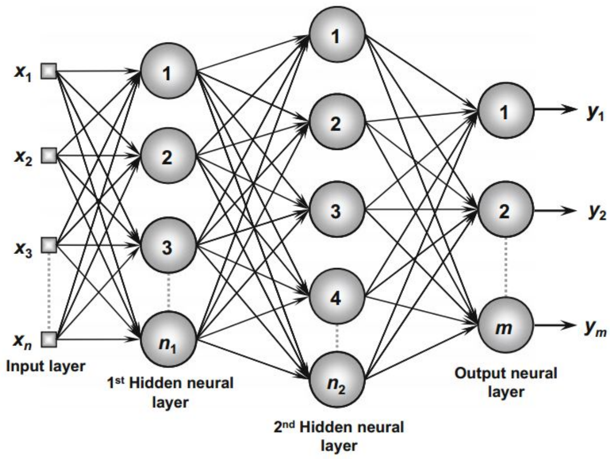

Figure 4.1: Structure of a feed forward artificial neural network [33]

In feed forward artificial neural networks, the neurons are ordered in layers. The first layer of the artificial neural network is the input layer, the final layer is the output layer and all the layers in between are called hidden layers. The number of neurons in the input and output layers correspond with the number of inputs and outputs of the system respectively, whereas the number of neurons in the hidden layers is chosen by the designer of the artificial neural network. The neurons of two consecutive layers are fully connected, these connections are defined by a weight, w_{ij}, where i is a neuron in a layer and j is a neuron in the consecutive layer. Data is fed through the layers, where the neurons transform this data with an activation function. The input to the activation functions of the neurons are the weighted sum of the outputs of the neurons in the previous layer. This is summarized in (4.1) [31].

$$x_j = \sigma_j \left(\sum_i w_{ij} x_i \right) \tag{4.1}$$

In this equation x_j is the output of a neuron j in a layer. σ_j is the activation function corresponding to the neurons in this layer. The sum in the equation contains w_{ij}, the weight of the connection between a neuron j and a neuron i from the previous layer, multiplied with x_i, the outcome of the neuron in the previous layer. Different functions can be used as activation function depending on the application of the artificial neural network, where typical activation functions are: the linear, sigmoid, tanh, ReLu and Gaussian activation functions [34].

4.1.2. Training an Artificial Neural Network

Artificial neural networks must be trained to be able to produce outputs that are expected from the artificial neural network. The artificial neural network is trained by updating the weights of the different connections between neurons. More important connections are given a higher weight than less important connections. One approach to train artificial neural networks is back-propagation. For the training process training data is required. This training data consists of expected input data to the system and the corresponding desired output data [33].

The back-propagation algorithm includes two stages: forward-propagation and back-propagation. In the forward-propagation stage the output of the artificial neural network with the current weights is determined. The training input data is fed through the artificial neural network and the output is calculated by transforming the data with the activation functions of the neurons. The produced output of the artificial neural network is compared with the desired outputs from the training data and an error is computed with the squared error function in (4.2) [33].

$$E = \frac{1}{2} \sum_{j=1}^{n} \left(y_j - y_j^{\mathrm{d}} \right)^2 \tag{4.2}$$

In this equation n is the number of neurons in the output layer, y_j is the output of a neuron in the output layer and y_d is the corresponding desired output of this neuron. Then, in the back-propagation stage, the squared error is minimized by updating the weights of the artificial neural network. This is achieved by feeding the computed error back into the artificial neural network and determining its gradient with respect to the connection weights. The weights are then updated according to (4.3) [33] [31].

$$w_{ij}(t+1) = w_{ij}(t) - \eta \cdot \frac{\partial E}{\partial w_{ij}} \tag{4.3}$$

Here η is a learning ratio, which the designer can use to tune the learning process of the artificial neural network. The back-propagation method is based on the steepest descent minimization, a first-order gradient method [31]. However, other methods for training the artificial neural network are also available. For instance, training artificial neural networks can also be done with the Levenberg-Marquardt algorithm, a second order gradient method [35].

4.2. Artificial Neural Networks in Socially Assistive Robots

A widely used application of artificial neural networks is recognition in computer vision [32]. This is also the most common application of artificial neural network in SARs. Therefore, artificial neural networks are mainly applied in the data analysis stage of SARs instead of in the control stage if SARs. This section discusses the applications of artificial neural networks in the field of SARs

Bera et al. in [36] and Garcia et al.in [37] use artificial neural networks to determine the emotional state of the user. Bera et al. in [36] use a person's walking trajectory and his or her facial expression to determine the emotional state. The emotional states are measured in four categories: angry, happy, sad and neutral based on the pleasure and arousal dimensions. Two video streams are used to determine the emotional state, one from a fixed overhead camera and one from a camera on the SAR. The overhead camera collects the trajectory parameters of pedestrians, whereas the robot camera detects the facial features. The trajectory features are mapped to an emotional state using linear regression from a user study. The facial features are fed to a convolutional artificial neural network, which is trained with an emotion dataset, to compute the corresponding emotion. The computed emotions can than be used for a socially aware SAR. Garcia et al. in [37] also uses a convolutional artificial network to map facial expressions to emotional states. The artificial neural network is trained with an emotional faces database and can detect seven different emotions: angry, disgust, fear, happy, neutral, sad and surprised. The artificial neural network focuses on the shape of the mouth, eyes and eyebrows to detect the emotional states. These features are highlighted by running the images through Gabor filters, which are powerful filters for edge detection. After training the artificial neural network it has been tested on the humanoid robot Nao. The results of the real-time test of recognizing emotions with Nao are significantly lower than from the classification of the training data. This is caused by the fact that the conditions in of the faces in the training data are optimal and in the real-life test this is not the case. The ability to recognize emotional states enables a SAR to communicate more effectively with humans and perhaps adjust the SAR's behavior to the emotional state of the user.

Martinez-Martin and Cazorla in [38] use an artificial neural network to recognize physical exercises performed by elderly. The output of the artificial neural network is a skeleton that visualizes the limb positions of the patient. The SAR reminds the elderly of scheduled exercise sessions. In these sessions it explains which exercises to perform and monitors the patient's execution of the exercises. The robot gives the elderly feedback on their performance. Furthermore, their performance is recorded such that a therapist can analyze it and follow the patient's progress.

Moreover, as discussed in section 3.2, artificial neural networks are combined with fuzzy logic control in [30]. Here fuzzy logic control is used as the main controller for the proxemics of a SAR, where the fuzzy rules depend on adaptive parameters. Artificial neural networks are used to train these parameters and improve the functionality of the fuzzy logic controller.

5
Reinforcement Learning

A final method for personalization and adaptation that is considered in this literature review is reinforcement learning. Similar to artificial neural networks, reinforcement learning is a model free approach. However, where artificial neural networks require training data to learn the expected behavior, reinforcement learning does not. This chapter starts with an explanation of the basic working principles of reinforcement learning. This is followed by the description of different forms of reinforcement learning. Finally, examples of reinforcement learning in SARs are discussed.

5.1. Working Principle of Reinforcement Learning

Reinforcement learning relies on the interactions of an agent (for instance, an agent that should make decisions about taking some actions in an unknown environment) with the environment and learning which actions have a positive result when trying to achieve a certain goal. This is similar to the way people learn throughout their lifetime. For example, when a child touches a hot pan, it hurts and the child knows not to do this in the future. In reinforcement learning rewards are given to an agent when it performs an action. Depending on the quality of the action with respect to the desired goal, the given reward is high or low. The agent learns by updating its behavior such that it can maximize the rewards it receives. Thus, in the learning process the agent is not told what behavior is expected from the agent, but the agent should discover this by itself [39].

Some crucial concepts in reinforcement learning are: agent and environment, policy, reward function and the value function. The agent is the decision maker, which interacts with its environment. The interaction between the agent and the environment is displayed in Figure 5.1.

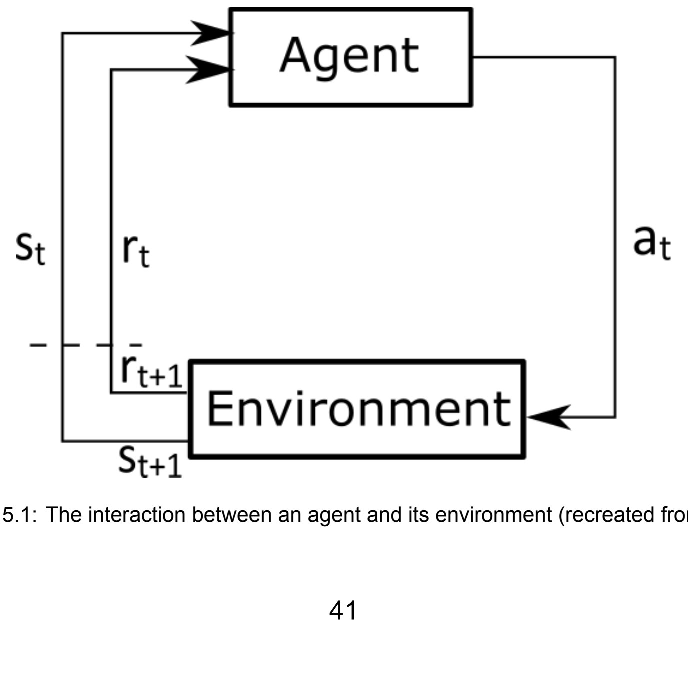

Figure 5.1: The interaction between an agent and its environment (recreated from [39])

The agent makes its decisions based on a policy. The agent performs different actions depending on the the state of the environment. For example, consider a mobile robot that must avoid obstacles in its environment. When this robot senses that the path in front of it is clear, it continues to drive straight ahead. However, when it notices that it approaches an obstacle, it makes a turn to avoid this obstacle. The reward function determines the reward that the agent receives for a given action. The reward function is set up by the designer of the reinforcement learning system. Based on the rewards, the agent knows whether an action is desirable or not. In the case of the mobile robot, a small positive reward can be given to the robot every time step that it has not collided with an obstacle and a large negative reward can be given in case of a collision. The reward function only provides short-term feedback on whether an action is good or bad. However, the goal of the agent is to maximize the rewards over the entire length of operation. Hence, the value function is introduced. The value function assigns a value to different states, the higher this value the more desirable the state. The agent must find actions that lead to the states with higher values. The value function can be seen as the expected sum of future rewards and indicates which actions are desirable in the long run. It can be the case that an action has a low reward but leads to a state with a high value. Then, despite the low immediate reward, this action is still desirable in the long run [39].

Initially, the agent has no information about the environment. Therefore, the agent starts to interact with the environment to gather information in episodes. The agent learns by taking actions, receiving rewards and updating the value functions. However, as the agent gathers more information a challenge of balancing the exploitation and exploration of the environment is raised. Exploiting the environment means taking actions that have resulted in high rewards in the past. Exploiting is in line with the goal of maximizing the agent's reward. However, it can be the case that there are states with high rewards which the agent has not visited yet. Therefore, the agent also must explore the environment to discover new states with high rewards. The designer of the agent must find a good balance between exploring and exploiting the environment to allow the agent to learn the optimal solution to the problem. This balance can be implemented in the policy. Three possible policies are:

Random: a random action is selected from the set of possible actions

Greedy: the agent takes the action with the highest value function

ϵ-Greedy: in ϵ percent of the cases, the random policy is used. The other times the greedy policy is used

The random policy has full exploration, the greedy policy has full exploitation. The ϵ-greedy policy is suitable to balance exploration and exploitation [39].

5.2. Forms of Reinforcement Learning

Different approaches to reinforcement learning exist. This section discusses the following forms of reinforcement learning:

1. Q-learning
2. Deep Reinforcement Learning
3. Policy-Gradient Reinforcement Learning

5.2.1. Q-learning

Q-learning is applied to discrete systems with a limited number of states and actions. Instead of the value function, Q-learning uses a value called the Q-value to determine which action to take. The Q-value does not assign a value to each individual state but to every state-action combination, these values are stored in a table. When the agent is in a certain state it looks up the Q-values for this state and selects an action according to its policy. The exact algorithm works as follows [39]:

1. Initialize $Q(s,a)$
2. Repeat for a specific number of episodes
   (a) Determine starting state, s
   (b) Repeat for a specific number of steps in the episode
      i. Determine corresponding action, a, based on the policy derived from Q and s
      ii. Take action a
      iii. Observe the reward, r, and the new state, s'
      iv. Update Q-value according to: $Q(s,a) = Q(s,a) + \alpha\left[r + \gamma \text{max}_u Q(s',u) - Q(s,a)\right]$
      v. Update s to s'
   (c) Stop episode if target is reached or the maximum number of steps is exceeded

In the Q-value update step $\text{max}_u Q(s',u)$ means taking the maximum Q-value for the new state. Here u means every possible action that can be taken in the new state, the action which results in the highest Q-value should be taken. α is the learning rate of the reinforcement learning system, a high learning rate leads to larger updates of the Q-values. γ is a discount factor on future rewards incorporated in the Q-values. α and γ are set by the designer of the reinforcement learning system. The discount factor allows the designer of the reinforcement learning system to account for uncertainties about whether these future rewards are received. Thus, in summary, in the Q-learning algorithm the agent explores the environment in episodes. The episode lasts for a predefined number of time steps or until a terminating state is reached. Every time step the agent takes an action according to the policy, which can be a greedy or ϵ-greedy policy for example. Next it observes the reward and new state. Then it updates the Q-value for the state-action pair using the received reward and maximum Q-value of the new state. Finally, the state is updated and the process is repeated [39].

The main advantage of Q-learning is that the method is easy to implement. However, Q-learning is limited in its applications, i.e., it can only be used for systems with discrete state-action spaces. Furthermore, when the state-action space becomes too large the system requires a lot of memory and the performance of the system reduces time-wise.

5.2.2. Deep Reinforcement Learning

When the state space becomes large, Q-tables take too much time and memory. This requires a more advanced form of reinforcement learning: Deep Reinforcement Learning. Deep reinforcement learning combines classic reinforcement learning with artificial neural networks. The deep version of Q-learning is called Deep Q-Networks (DQN) [40].

DQN's do not use a table to determine the Q-values, but they calculate the Q-value using artificial neural networks. Two artificial neural networks are used in the DQN method: A prediction and a target artificial neural network. The predictor artificial neural network predicts the Q-value for the current state, whereas the target artificial neural network is used to give a target Q-value required to update the prediction artificial neural network. The steps of the algorithm are similar to the original Q-learning method apart from the Q-value update. The deep learning method works as follows [41]:

1. Initialize predictor artificial neural network, Q, with weights θ
2. Initialize target artificial neural network, $\hat{Q}$, with weights $\hat{\theta} = \theta$
3. Repeat for a specific number of episodes
   (a) Determine starting state, s
   (b) Repeat for a specific number of steps in the episode
      i. Determine action, a, based on the policy and the Q-values determined by the prediction artificial neural network, Q
      ii. Take action a
      iii. Observe the reward, r, and the new state, s'

iv. Calculate the maximum Q-value of the new state using the target artificial neural network $\hat{Q}$
v. Update the weights, θ, of Q based on the following error: $\left(r + \gamma \max_u \hat{Q}(s', u) - Q(s, a)\right)^2$
vi. Update s to s'
vii. After a predefined number of steps reset weights $\hat{\theta}$ equal to θ, such that $\hat{Q} = Q$

(c) Stop episode if target is reached or the maximum number of steps is exceeded

In the update step of the weights $\max_u \hat{Q}(s', u)$ means taking the maximum Q-value of the outcome of the target artificial neural network for the new state. Here u means every possible action that can be taken in the new state, the action which results in the highest Q-value should be taken. γ is a discount factor on future rewards incorporated in the Q-values. The discount factor allows the designer of the reinforcement learning system to account for uncertainties about whether these future rewards are received. γ is set by the designer of the reinforcement learning system. Furthermore, the designer must set a number of steps after which the weights of the target artificial neural network are updated. This update is simply copying the current weights of the prediction artificial neural network. The target artificial neural network increases stability in the learning process. Furthermore, a replay memory can be used to increase stability of the learning process [41]. When using a replay memory the artificial neural network is not directly trained by interacting with the environment, which means that only the most recent experience is used for training. Instead, the interactions (states, actions, rewards and new states) of the agent are stored in a large table, the artificial neural network then trains by taking random batches from this table. Learning from a memory replay has the advantage that the state action pairs can be used more than once for the training process. Furthermore, the convergence properties of the training process are improved.

An advantage of DQN is that it can handle large state-action spaces, because Q-values are approximated with artificial neural networks instead of a Q-table. However, similar to Q-learing, in their original form, DQN's are also limited to discrete state-action spaces [40].

5.2.3. Policy Gradient Reinforcement Learning

Another form of reinforcement learning that is discussed in this literature review is Policy Gradient Reinforcement Learning. Policy Gradient methods do not rely on value functions or Q-values to select the actions, but they use a parameterized policy to select an action. Rewards and values can still be used to train the policy parameters. The parameters of the policy are described by θ and the policy is defined by (5.1). It describes the probability that action, a, is taken when the environment is in state, s, and the policy has parameters, θ [39].

$$\pi(a|s, \theta) = \Pr\left\{a_{t=a|s_t=s,\theta_t=\theta}\right\} \tag{5.1}$$

The policy is updated with a gradient ascent method based on the gradient of the parameters with respect to a performance measure, $J(\theta)$. This update is described by (5.2) [39].

$$\theta_{t+1} = \theta_t + \alpha \nabla J(\theta_t) \tag{5.2}$$

Where α is the learning rate of the system. A common algorithm for policy gradient reinforcement learning is REINFORCE. The REINFORCE algorithm uses the expected sum of rewards as performance measure, $J(\theta)$. When performing the Policy Gradient Theorem described in [39], the policy parameter update for the REINFORCE algorithm is equal to (5.3).

$$\theta_{t+1} = \theta_t + \alpha \cdot G_t \nabla_\theta \ln \pi\left(a_t|s_t, \theta\right) \tag{5.3}$$

In this equation G_t is the t-step return which is calculated by summing the discounted rewards of the future steps: $G_t = r_1 + \gamma \cdot r_2 + \gamma^2 \cdot r_3 + ... + \gamma^{t-1} \cdot r^t$, where γ is a discount factor set by the designer, t is the time step until which the return is considered and $r_1, r_2,$ etc. are the rewards received at the corresponding time steps.

Policy gradient reinforcement learning has the advantage that it can handle continuous systems and stochastic policies. Furthermore, the updates in policy gradient methods are smoother than the updates in value-based methods [39]. However, policy gradient methods have the pitfall of getting stuck in a local optimum [40].

5.3. Reinforcement Learning in Socially Assistive Robots

Reinforcement Learning is a widely used method for adaptation in SAR systems. It allows the SAR to adapt its behavior to the preferences of the user without the necessity of training data.

Q-learning is used to adapt the assistance provided by a SAR during a game of memory (finding pairs of identical cards) in [8]. A SAR provides assistance to the user when he or she loses the attention to the game and gives the user positive feedback. The SAR must show different behaviors depending on the user's states and the game's state. The user's states are the user's gaze and speech. A human researcher helps the SAR to categorize the user's gaze and speech in terms of whether he or she and how bad he or she requires assistance. For example, if the user searches all cards and gives a comment like: *"I don't know where the matching card is anymore"*, it is an indication for the human researcher to send an assistance request to the SAR. The researcher must categorize the severity of the request in three levels. The SAR uses the severity of a help request together with the game's state, i.e., the number of identical pairs that remain to be found to determine which behavior to display. The robot can provide feedback to the user in different ways, such as gazing in the direction of the right card, facial expressions to show that the user is near the right card, head gestures to steer the user to the right card or by using speech either to motivate the user if he or she is distracted or to tell the answer. In the reinforcement learning process the SAR learns to select the right type of feedback for the combinations of the game's state and the severity of the help request. The reward function of the reinforcement learning system prioritizes the use of the more subtle types of feedback. For example letting the SAR gaze in the direction of the right card is preferred over letting the SAR tell which is the right card, because this increases the difficulty for the user. Therefore, the rewards for the SAR are higher if it successfully helps the user with a more subtle type of assistance (gazing to the right card) than when it successfully helps the user with a very direct type of feedback (telling the right answer).

Tsiakas et al. in [25] and Clabaugh et al. in [17] use reinforcement learning to change the difficulty of the task and the type of feedback a SAR gives to its user. Tsiakas et al. in [25] use a SAR to help people with cognitive impairments with a sequencing task (the user must repeat a sequence of As, Bs and Cs). Clabaugh et al. in [17] use a SAR to assist children with autism spectrum disorder in a set of mathematical games. However, the concept of providing adaptive feedback and changing the difficulty of the task is the same for the two experiments. The goal of having adaptive feedback and task difficulty is to keep the user engaged. If the task is too easy, the user might get bored. If the task is too difficult, the user might give up. Therefore, it is important to change the difficulty of the task to fit the capabilities of the user and still remain challenging enough for the user. In both papers the performance of the user forms the basis of the reward function. Higher rewards are given when the user successfully completes more difficult tasks. Therefore, the SAR aims to assist the user such that he or she can complete the most difficult tasks.

In [22] Q-learning is used for a SAR that should learn to assist elderly with daily activities. In this paper, the robot learns to help patients with tea making. In the first stage of the research the robot is taught different behaviors using a method called learning from demonstration. Example behaviors of the SAR are inviting the patient to make tea, giving instructions on how to perform a step or correcting the patient if a step is performed incorrectly. Q-learning is used to learn the SAR to apply the right behavior at the right time, because the most effective behavior of the robot depends on the user's state. The user's states consist of the user's functioning states and the user's activity states. The user's functioning states consider the following five mental states: focused, distracted, having a memory lapse, showing misjudgement and being apathetic. The user's activity state consider the actions that can be performed by the patient: successfully completing a step, being idle, repeating a step, conducting a step incorrectly or declining to continue the activity. The rewards given to the SAR depend on the user's execution of the different steps in the process of tea making. Negative rewards are given to the SAR when the user repeats a previous step, performs an incorrect step or declines to perform the step. Positive rewards are given to the SAR when the user successfully completes the right step necessary for tea making.

Tapus and Matarić in [19] use policy gradient reinforcement learning to adapt the behavior of a SAR to the personality of a post-stroke rehabilitation patient. The SAR motivates the patient to participate in rehabilitation exercises, such as drawing, lifting books, moving pencils and turning pages of a newspaper. The SAR adapts its behavior according to the level of extraversion of the patients. The adaptive variables in the robot behavior are: Proxemics, moving speed, the type of feedback (encouraging or nurturing) and how it brings this feedback to the patient (i.e., with high pitch and high volume or with low pitch and low volume). The effectiveness of the robot's behavior is measured and optimized by means of the number of tasks completed by the patient and the time it takes for the patient to complete the task. The results of the experiment show that people prefer a robot behavior that matches their own personality (whereas in this experiment, from the various personality traits extrovert/introvert has been considered).

6
Proposed Solution

The previous chapters have given a general overview on the concept of SARs and methods to achieve personalization and adaptation in the behavior of these robots. This chapter dives deeper into the problem of developing a decision making module for a SAR to improve the quality of life of elderly with dementia by intervening in their daily routines.

6.1. Considerations

The applications for SARs, discussed in section 2.3, can be categorized in applications with short-term and long-term use. In this division, short-term use is related to situations where the user is exposed to the robot for a limited amount of time, for example in therapy sessions. Long-term use of the SAR relates to service and entertainment robots. This research focuses on a SAR to improve the quality of life of elderly with dementia. Therefore, it is considered to be a service robot, which is used for longer periods of time. Example tasks of the robot are:

- Taking the user for a walk when the weather is nice
- Play music if the user feels sad
- Remind the user of group activities if these are planned for the day

As concluded by [11] people vary in their responses to a SAR and individuals respond differently to SARs at different times. Therefore, a single constant behavior is not the optimal solution for effective interactions between a SAR and a human. Instead, the robot must display personalized and adaptive behavior, to account for differences between different people and to account for variations in an individual's attitude towards a SAR. Currently, SARs with personalized and adaptive behavior are mainly used for short-term interactions. From the experiments included in section 2.4, [17] sessions over the course of a month is the longest lasting experiment. However, this Master thesis research focuses on personalization and adaptation of SAR behavior for a SAR applied in long-term dementia care, e.g., as a home assistant that can stay with the person at home for long time periods.

Taking this into consideration, there are two sides to the problem. First of all, a baseline behavior must be set that matches the general preferences of the user. This is considered as personalization of the SAR behavior. Tapus and Matarić in [19] show that people prefer SARs whose behaviors match their own personality. Therefore, the personalization of the proposed SAR for dementia care is based on the personality of the user. The SAR establishes a baseline behavior on the long-term personality traits of the dementia patients.

Secondly, the robot must adapt the basis behaviour to variations in the patient's mental status and to the environment. These adjustments in the robot's behavior are considered the adaptation side of the robot. If the user is in a bad mood, he or she might prefer not to have long interactions or when the weather is bad, going for a walk might not be a suitable activity. Through the adaptive behavior, the SAR becomes a more similar to a human caregiver who takes the current mental status of the patient into consideration when providing care to the patient.

6.2. Solution

The proposed solution to achieve personalized and adaptive SAR behavior is to combine fuzzy logic control and reinforcement learning in a modular approach. Fuzzy logic control is used to personalize the robot's behavior based on the personality of the user. Reinforcement learning is used to adapt the base behavior to handle fluctuations in the patient's mental state and the environmental changes. This section discusses the approaches to the personalization and the adaptation of the SAR behavior and explains the reasoning behind decisions that have been made.

Two options have been considered for combining the two control techniques. First of all, a hierarchical approach has been considered. In the hierarchical approach fuzzy logic control is used as the main control method and reinforcement learning is used to tune the fuzzy rules in this system. Ye et al. in [42] use this approach for obstacle avoidance in a mobile robot. However, no literature is found from this approach in the field of SARs. The second option for combining fuzzy logic control and reinforcement learning is a modular approach, which is the proposed solution for this thesis. In the modular approach fuzzy logic control and reinforcement learning are used in different stages of the decision making module of the SAR. An advantage of this method is that it does not require modification of the two control mechanisms, but they can be used in their original form. Moreover, the modular approach allows to take out the fuzzy logic control or reinforcement learning module and implement it in a different research with minimal adjustments. This is convenient as the research team also performs research on the use of SARs for children with autism spectrum disorder. The modular approach is selected over the hierarchical approach for these two reasons.

6.2.1. Personalization with Fuzzy Logic Control

Personalization of the SAR behavior is achieved with fuzzy logic control. The fuzzy logic controller uses a rule-based approach to match the SAR's behavior to match the personality of the user. Rules can be set up to define the main characteristics of the robot's behavior depending on the user's personality traits. Tapus and Matarić in [19] use the extraversion trait for personalization, this is a trait from the Eysenck Personality Inventory [43]. The extraversion trait is related to excitement and arousal [19]. Therefore, this trait is already visible on a short-term basis. However, the aim of this research is to create a personalized and adaptive SAR to assist dementia patients in their daily lives for extended periods of time. Therefore, other personality traits become relevant too. The proposed SAR system personalizes its behavior according to the Big Five personality traits [44]:

Extraversion: Extraversion measures the level of sociability and excitability. People with high levels of extraversion are more sociable, assertive and active. They pursue excitement and challenge. People scoring low on the extraversion trait are more reserved and cautious.

Agreeableness: Agreeableness relates to someone's intentions. Individuals high in agreeableness are modest, cooperative and trustworthy. They are concerned about the feeling of others and show interest in them. People low on agreeableness are suspicious, irritable and competitive. Sometimes people low on agreeableness can even be manipulative.

Openness: Openness describes a person's imagination and acceptance for new experiences. People scoring high on openness are open-minded, imaginative and tolerant. They are curious to try out new things and engage in new ideas. People low on openness are more down-to-earth and conventional.

Conscientiousness: Conscientiousness is concerned with someone's feeling of responsibility. Conscientious people are more careful, organized and scrupulous. They pay attention to details and aim to finish tasks successfully. People with low levels of conscientiousness are irresponsible and disorganized.

Neuroticism: Neuroticism marks someone's emotional stability. People with high neuroticism are anxious, insecure and depressed. Often they are upset more easily and sad. People scoring low on neuroticism are more laid back, emotionally stable and can handle stress very well.

Extraversion is the measure of sociability of the user. Therefore, the higher the level of extraversion of the user, the more interactions the robot can have with him or her. Instead of only interacting with the user if necessity is high, the SAR can also have chats about daily things like the weather. Furthermore, the ways the robot interacts with the person may be based on the level of extraversion of the user. As an example, Tapus and Matarić in [19] show that more extravert people prefer challenging comments on high pitch and volume, whereas introvert people prefer nurturing comments on low pitch and volume. For this research also the type of activities are important. Introvert people prefer to spent time alone instead of group activities. Therefore, the SAR can suggest activities like reading a book to introvert people, whereas it can suggest to participate in group activities of the caring home for extravert people.

Agreeableness affects the cooperativeness of the user. People who score low on agreeableness are focused more on their own well being. The SAR can use this trait in the way it motivates people. For agreeable people the SAR should bring the activities as if they are more of a team effort. For people who show low levels of agreeableness, the SAR must focus on the benefits for the patient when motivating the patient to participate in an activity. Furthermore, people low in agreeableness prefer to work independently. Therefore, a less agreeable patients might not want to participate if the SAR tells him or her which activities to do. Instead the SAR can give multiple suggestions and let the patient decide on its own, this way the person feels like it was his or her own plan to do the activity. These two things might increase the likelihood of less agreeable users to participate in the activities. Furthermore, people low on agreeableness are competitive, so the SAR can use this trait to motivate the user to engage in the activities by creating or proposing activities that involve some levels of competition.

Openness is related to people's view on trying new things. This can be useful to determine which activities to propose to the user. People scoring high in openness are more open to new experiences. Therefore, the robot can suggest new activities to the user every once in a while. Moreover, open people are generally more creative. Therefore, the SAR can suggest more creative activities, such as painting, to further motivate or engage the user. People with low levels of openness are less eager to engage in new activities. Thus, the SAR should suggest completely new activities less often to users scoring low in openness. Instead the SAR should stick to activities to which the user has responded positively in the past.

Conscientiousness determines to what extent someone likes structure and feels responsibility. In case the user has high levels of conscientiousness, he or she prefers to have a fixed schedule. The robot can take this into account by communicating activities upfront and planning them for specific time intervals. People low in conscientiousness do not like fixed time schedules. Therefore, the robot should be more spontaneous for these types of users. Instead of having a schedule of when to do certain activities, the SAR could suggest improvised activities.

Neuroticism is related to a person's emotional stability and anxiousness. The SAR should take into account the level of neuroticism of the person it interacts with in its way of interacting with the user. For users with high levels of neuroticism, the robot should be careful not to distress the user. For instance, the SAR should have comforting comments and should try to resolve the user's insecurities, this can make neurotic patients feel less anxious. People scoring low on neuroticism do not feel anxious or insecure. Therefore, the SAR does not require special behavior for these patients.

Fuzzy logic control is selected for the personalization of the SAR's behavior for several reasons. First of all, fuzzy logic control is a rule-based approach, which creates the possibility to set up logical if-then rules for each personality trait, which is beneficial as the different personality traits influence the robot's behavior in different ways. Furthermore, research performed on human personalities, e.g, the authors in [44] provide expert knowledge on people's preferences corresponding to the different personality traits. This expert knowledge can be used in the rules of the fuzzy system. Implementing expert knowledge in the fuzzy rules during the design phase prevents the system from needing an extensive learning phase to match the right behavior with the user's personality.

A second reason to use fuzzy logic control over the other intelligent control methods is that fuzzy logic can handle partial membership to sets. The personality traits are measured along a spectrum, someone's personality is not defined by having certain traits but by the extent someone displays the traits. Taking extraversion into consideration as an example, someone is not considered fully extravert or fully introvert. However, someone displays characteristics from extraversion or introversion and is considered extravert or introvert depending on to which extent he or she displays these characteristics. Fuzzy logic control can deal with such partial memberships, making it a suitable method for personalizing the SAR behavior to the user's personality.

A final reason to use fuzzy logic control for personalization of the SAR's behavior is that fuzzy logic control is able to work with ambiguous variables like linguistic terms. Therefore, setting up the rules for the decision making module is a straightforward task. The personality traits and the corresponding behavioral elements can be described with words. This makes the control rules easy to understand even for people without technical knowledge about the robot itself.

6.2.2. Adaptation with Reinforcement Learning

Reinforcement learning is used to achieve adaptation in the robot's behavior. Behavior adaptation enables the SAR to adjust its behavior to changing circumstances. For example, the user might prefer cheerful music when he or she is in a good mood. However, if the user feels sad he or she might want to listen to calm music. The SAR should learn which preferences the user has in different circumstances and should hence display a behavior that matches these preferences.

The reinforcement learning method that is planned for the adaptation process is Q-learning. Inputs to the reinforcement learning module are the user's states and the environmental states. The user's states can include the user's emotional status (happy, sad, bored, etc.), activity level (tired or energized) and attentiveness (distracted or paying attention). Environmental states are elements that influence the activities that can be considered for a user, such as the weather condition or the time of day. Moreover, the examples for the outputs of an adaptive decision making system for the SAR can be to suggest an activity for the user, to initiate a conversation with the user or to entertain the user by playing music.

If it turns out that the total number of input and output state combinations becomes too large or the input and output states are difficult to describe by discrete variables, deep reinforcement learning can be selected over Q-learning. However, solving the adaptation problem is attempted with Q-learning first as its implementation is easier than deep reinforcement learning, which requires a neural network. Policy gradient reinforcement learning is not selected as it has the possibility to converge to a local optimum.

Reinforcement learning is selected for the behavior adaptation as it is able to tackle the adaptivity problem for each individual user and it does not require large amounts of training data. Fuzzy logic control is suitable to set the baseline behavior of the SAR as this can be captured with rules to link the behavior to the different personality traits. However, adaptation is used to make the behavior specific to individual users. Each user has different preferences in different situations, which makes it impossible to set up rules that lead to a solution that fits all users. For this reason, rule-based approaches like fuzzy logic control, although suited for the personalization problem, are not suited for the adaptation problem. Moreover, artificial neural networks are not used to solve the adaptation problem in this thesis, because they require lots of training data to get to a possible solution. Furthermore, the obtained solution is based on the preferences of the dementia patients who have been included in the dataset. Therefore, the SAR learns adaptations that generally work in certain situations and for certain people. However, it can be that a user responds differently than the patients included in the dataset, and hence, the adaptation process may be inefficient for this user. Besides this problem, collecting sufficient training data from dementia patients is out of the scope of this thesis research. Therefore, reinforcement learning is selected for this Master thesis as the most suitable approach for the behavior adaptation of the SAR.

6.3. Implementation

The created decision making module is going to be eveluated via user models, which are computer simulations of example user personality scores with possible interactions with the SAR, and on people from the research team. Testing the personalized and adaptive SAR behavior on real dementia patients is out of the scope of this thesis research, but can be done in future research. The user models considered for the proposed research exhibit different scores on the Big Five personality traits, which has an effect on the personalization of the SAR's behavior. Moreover, these models exhibit responses (selected from a determined set of potential responses) to the robot's actions and based on the different environment states, where these different responses affect the adaptation of the SAR's behavior. With these carefully designed user models and scenarios, the developed decision making system for the SAR can efficiently be assessed for a variety of user identities. Furthermore, the user models and scenarios allow us to simulate long-term useinteraction scenarios of the SAR with different potential users in short periods of time, since a series of successive user and environment states can be easily simulated. Besides the user models, the system is also assessed via real people from the research team to see if the personalization and adapatation of the behaviors of the SAR also work well with real human beings. Before implementing the experiments, the participants are asked to fill in a personality questionnaire to determine their scores on the Big Five personality traits for the personalization of the SAR's behavior. A limitation of testing the SAR's decision making system with real people is that it takes an extensive period of time. The adaptive side of the SAR focuses on long-term use, so the test subject should have a lot of interactions with the robot to see if the behavior adaptation is successful. After the interactions, the people are asked to complete a questionnaire on their findings of the robot behavior. The content of this questionnaire is yet to be determined. However, the questions in the questionnaire should fulfill the criteria that a proper and relevant assessment of the interactions between the SAR and the participant can be established.

The developed decision making module is going to be implemented on a real humanoid robot owned by the research team: the Nao robot [45]. The Nao robot is a small human-like robot, illustrated in Figure 6.1. The robot is equipped with:

- Two 2D camera's for object recognition and facial detection
- Four microphones and speakers for speech
- Speech recognition in 20 languages to understand the user
- 25 degrees of freedom for movement
- Seven touch sensors and an inertial measurement unit for orientation

Furthermore, the Nao robot is equipped with NAOqi, a special software framework that allows to program the robot. NAOqi is a cross-platform and cross-language framework, thus it works on different operating systems and accepts the programming languages Python and C++. Therefore, the controller is written in Python and is implemented on the robot via the NAOqi framework.

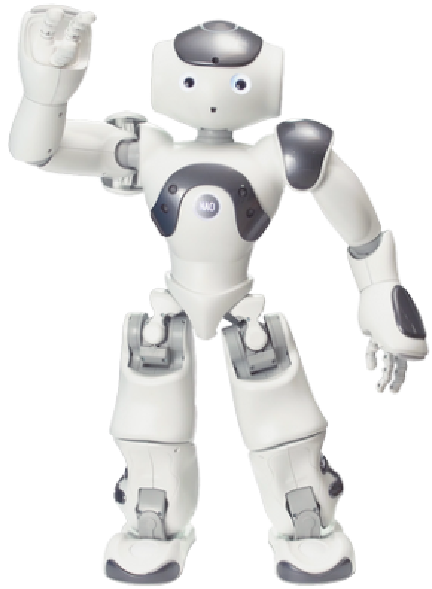

Figure 6.1: Appearance of the Nao humanoid robot [45]

In the design of the decision making module, the face detection of the Nao robot can be used to recognize the user. Moreover, in [46] Nao is used to detect emotional states of children, which can be useful and inspirational for the behavior adaptation in this Master thesis project. Nao's speech recognition can be used to understand the feedback provided by the user. The Nao robot is a suitable robot for the assistance of dementia patients. Martin et al. in [47] perform an acceptance test among a group of dementia patients with a Nao robot, where the results of this test show that most patients respond positively to Nao and consider it as a child. For some applications a humanoid robot is not a fitting robot type. Ricks and Colton in [48] conclude that children with autism are more interested in simplified abstract forms, so non-humanoid robots might be a better option for applications that address people with autism spectrum disorder. Assisting dementia patients in daily life is chosen as the application of the SAR, as Nao's human like appearance might be more effective for dementia patients than for example children with autism.

7 Conclusions

An ageing society leads to a growth in the number of people suffering from dementia. This development puts an increasing pressure on health care workers in caring homes. Consequently, the care-givers are limited in the personalization they can give in the care, which reduces the quality of life of patients with dementia. A potential solution to the problem is to provide personalized and adaptive care with Socially Assistive Robots. However, from the performed literature study, it turns out that the research performed on personalized and adaptive SAR behavior is mainly limited to short-term interactions and usage. Therefore, the following research question has been formulated for this Master thesis project:

How can a decision making module be developed for a Socially Assistive Robot by means of techniques from control theory and Artificial Intelligence, such that the robot personalizes its decision making towards long-term personality traits of dementia patients and exhibits adaptive behavior to handle changes in the needs and emotional status of patients in its interactions?

The intended SAR decision making module uses two different control methods for the personalization and adaptation of the robot. First of all, the robot personalizes its behavior to the long-term personality traits of the user via fuzzy logic control, where rules are set up for the different traits of the Big Five personality model. The adaptation of the behavior adjusts the robot's interactions to the changes in the user's mental state and in the environmental states. The designed system is going to be assessed with user model simulations and with real human participants from the research team. The target of the SAR is to aid dementia patients in their daily routine by providing means of social interactions, suggesting activities or giving reminders of important events with the aim of increasing the quality of life of the patients.

Bibliography

[1] United Nations, Department of Economic and Social Affairs, Population Division, *World Population Ageing 2019*. New York: United Nations, 2020.

[2] C. Ferri, M. Prince, C. Brayne, H. Brodaty, L. Fratiglioni, M. Ganguli, K. Hall, K. Hasegawa, H. Hendrie, Y. Huang, A. Jorm, P. Menezes, E. Rimmer, and M. Scazufca, "Global prevalence of dementia: A delphi consensus study," *Lancet*, vol. 366, no. 9503, pp. 2112–2117, 2006.

[3] L. Mierlo, H. van der Roest, F. Meiland, and R.-M. Dröes, "Personalized dementia care: Proven effectiveness of psychosocial interventions in subgroups," *Ageing research reviews*, vol. 9, no. 2, pp. 163–183, 2009.

[4] T. Epp, "Person-centred dementia care: A vision to be refined," *The Canadian Alzheimer Review*, vol. 5, no. 3, pp. 14–18, 2003.

[5] D. Feil-Seifer and M. J. Mataric, "Defining socially assistive robotics," in *9th International Conference on Rehabilitation Robotics*, (Chigago, United States of America), pp. 465–468, IEEE, 2005.

[6] T. Fong, I. Nourbakhsh, and K. Dautenhahn, "A survey of socially interactive robots," *Robotics and Autonomous Systems*, vol. 42, no. 3-4, pp. 143–166, 2003.

[7] C. Cherry, *On human communication : a review, a survey, and a criticism*. Cambridge, United States of America: The MIT Press, 3rd ed.

[8] J. Hemminghaus and S. Kopp, "Towards adaptive social behavior generation for assistive robots using reinforcement learning," in *12th ACM/IEEE International Conference on Human-Robot Interaction*, (Vienna, Austria), pp. 332–340, IEEE, 2017.

[9] J. Broekens, M. Heerink, and H. Rosendal, "Assistive social robots in elderly care: A review," *Gerontechnology*, vol. 8, no. 2, pp. 94–103, 2009.

[10] P. Marti, M. Bacigalupo, L. Giusti, C. Mennecozzi, and T. Shibata, "Socially assistive robotics in the treatment of behavioural and psychological symptoms of dementia," in *The First IEEE/RAS-EMBS International Conference on Biomedical Robotics and Biomechatronics*, (Pisa, Italy), pp. 483–488, IEEE, 2006.

[11] W. Moyle, C. Jones, J. Murfield, L. Thalib, E. Beattie, D. Shum, and B. Draper, "Using a therapeutic companion robot for dementia symptoms in long-term care: reflections from a cluster-RCT," *Aging & Mental Health*, vol. 23, no. 3, pp. 329–336, 2017.

[12] S. Sabanovic, C. Bennett, W.-L. Chang, and L. Huber, "PARO robot affects diverse interaction modalities in group sensory therapy for older adults with dementia," in *13th IEEE International Conference on Rehabilitation Robotics*, (Seattle, United States of America), pp. 1–6, IEEE, 2013.

[13] A. Libin and J. Cohen-Mansfield, "Therapeutic robocat for nursing home residents with dementia: Preliminary inquiry," *American journal of Alzheimer's disease and other dementias*, vol. 19, no. 2, pp. 111–116, 2004.

[14] A. Tapus, C. Tapus, and M. J. Matarić, "The use of socially assistive robots in the design of intelligent cognitive therapies for people with dementia," in *IEEE International Conference on Rehabilitation Robotics*, (Kyoto, Japan), pp. 924–929, IEEE, 2009.

[15] C. Syriopoulou-Delli and E. Gkiolnta, "Review of assistive technology in the training of children with autism spectrum disorders," *International Journal of Developmental Disabilities*, vol. 0, no. 0, pp. 1–13, 2020.

[16] E. T. Bekele, U. Lahiri, A. R. Swanson, J. A. Crittendon, Z. E. Warren, and N. Sarkar, "A step towards developing adaptive robot-mediated intervention architecture (ARIA) for children with autism," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 21, no. 2, pp. 289–299, 2013.

[17] C. Clabaugh, K. Mahajan, S. Jain, R. Pakkar, D. Becerra, Z. Shi, E. Deng, R. Lee, G. Ragusa, and M. Matarić, "Long-term personalization of an in-home socially assistive robot for children with autism spectrum disorders," *Frontiers in Robotics and AI*, vol. 6, no. 110, pp. 1–18, 2019.

[18] M. Matarić, J. Eriksson, D. Feil-Seifer, and C. Winstein, "Socially assistive robotics for post-stroke rehabilitation," *Journal of NeuroEngineering and Rehabilitation*, vol. 4, no. 5, pp. 1–9, 2007.

[19] A. Tapus and M. Matarić, "Socially assistive robots: The link between personality, empathy, physiological signals, and task performance.," in *AAAI spring symposium: emotion, personality, and social behavior*, (Stanford, United States of America), pp. 133–140, AAAI, 2008.

[20] M. Fujita and H. Kitano, "Development of an autonomous quadruped robot for robot entertainment," *Autonomous Robots*, vol. 5, no. 1, pp. 7–18, 1998.

[21] L. Odetti, G. Anerdi, M. P. Barbieri, D. Mazzei, E. Rizza, P. Dario, G. Rodriguez, and S. Micera, "Preliminary experiments on the acceptability of animaloid companion robots by older people with early dementia," in *29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, (Lyon, France), pp. 1816–1819, IEEE, 2007.

[22] C. Moro, G. Nejat, and A. Mihailidis, "Learning and personalizing socially assistive robot behaviors to aid with activities of daily living," *ACM Transactions on Human-Robot Interaction*, vol. 7, no. 2, pp. 1–25, 2018.

[23] J. Pineau, M. Montemerlo, M. Pollack, N. Roy, and S. Thrun, "Towards robotic assistants in nursing homes: Challenges and results," *Robotics and Autonomous Systems*, vol. 42, no. 3, pp. 271–281, 2002.

[24] K. Eklundh, A. Green, and H. Hüttenrauch, "Social and collaborative aspects of interaction with a service robot," *Robotics and Autonomous Systems*, vol. 42, no. 3, pp. 223–234, 2003.

[25] K. Tsiakas, M. Abujelala, and F. Makedon, "Task engagement as personalization feedback for socially-assistive robots and cognitive training," *Technologies*, vol. 6, no. 2, pp. 1–17, 2018.

[26] Y. Bai and D. Wang, "Fundamentals of fuzzy logic control — fuzzy sets, fuzzy rules and defuzzifications," in *Advanced Fuzzy Logic Technologies in Industrial Applications*, London, United Kingdom: Springer, 2006.

[27] M. Jezewski, R. Czabanski, and J. Leski, "Introduction to fuzzy sets," in *Theory and Applications of Ordered Fuzzy Numbers: A Tribute to Professor Witold Kosiński*, Cham, Switzerland: Springer Nature, 2017.

[28] S. Bova, P. Codara, D. Maccari, and V. Marra, "A logical analysis of mamdani-type fuzzy inference, i theoretical bases," in *International Conference on Fuzzy Systems*, (Barcelona, Spain), pp. 1–8, IEEE, 2010.

[29] H. Mobahi and S. Ansari, "Fuzzy perception, emotion and expression for interactive robots," in *IEEE International Conference on Systems, Man and Cybernetics*, (Washington, United States of America), pp. 3918–3923, IEEE, 2003.

[30] A. Vitiello, G. Acampora, M. Staffa, B. Siciliano, and S. Rossi, "A neuro-fuzzy-bayesian approach for the adaptive control of robot proxemics behavior," in *IEEE International Conference on Fuzzy Systems*, (Naples, Italy), pp. 1–6, IEEE, 2017.

[31] D. Svozil, V. Kvasnicka, and J. Pospíchal, "Introduction to multi-layer feed-forward neural networks," *Chemometrics and Intelligent Laboratory Systems*, vol. 39, no. 1, pp. 43–62, 1997.

[32] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, vol. 1. Cambridge, United States of America: MIT Press, 2016.

[33] I. N. da Silva, D. H. Spatti, R. A. Flauzino, L. H. B. Liboni, and S. F. dos Reis Alves, *Artificial Neural Networks: A Practical Course*. Cham, Switzerland: Springer Nature, 1st ed., 2017.

[34] S. Sharma, S. Sharma, and A. Athaiya, "Activation functions in neural networks," *International Journal of Engineering Applied Sciences and Technology*, vol. 04, no. 12, pp. 310–316, 2020.

[35] H. Yu and B. M. Wilamowski, "Levenberg–marquardt training," in *The industrial electronics handbook: Intelligent Systems*, vol. 5, pp. 12–1 – 12–16, Boca Raton, United States of America: CRC Press, 2011.

[36] A. Bera, T. Randhavane, R. Prinja, K. Kapsaskis, A. Wang, K. Gray, and D. Manocha, "How are you feeling? Multimodal emotion learning for socially-assistive robot navigation," in *15th IEEE International Conference on Automatic Face and Gesture Recognition*, (Buenos Aires, Argentina), 2020.

[37] A. Ruiz-Garcia, M. Elshaw, A. Altahhan, and V. Palade, "A hybrid deep learning neural approach for emotion recognition from facial expressions for socially assistive robots," *Neural Computing and Applications*, vol. 29, no. 7, pp. 359–373, 2018.

[38] E. Martinez-Martin and M. Cazorla, "A socially assistive robot for elderly exercise promotion," *IEEE Access*, vol. 7, pp. 75515–75529, 2019.

[39] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, United States of America: A Bradford Book, 2018.

[40] V. François-Lavet, P. Henderson, R. Islam, M. G. Bellemare, and J. Pineau, "An introduction to deep reinforcement learning," *Foundations and Trends in Machine Learning*, vol. 11, no. 3-4, p. 219–354, 2018.

[41] Y. Li, "Deep reinforcement learning," *arXiv preprint arXiv:1810.06339*, vol. abs/1810.06339, 2018.

[42] C. Ye, N. H. C. Yung, and D. Wang, "A fuzzy controller with supervised learning assisted reinforcement learning algorithm for obstacle avoidance," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 33, no. 1, pp. 17–27, 2003.

[43] H. J. Eysenck, *The structure of human personality*. Methuen London, 1970.

[44] S. Roccas, L. Sagiv, S. Schwartz, and A. Knafo-Noam, "The big five personality factors and personal values," *Personality and Social Psychology Bulletin*, vol. 28, no. 6, pp. 789–801, 2002.

[45] Softbank Robotics, "NAO^6." https://www.softbankrobotics.com/emea/en/nao. Accessed: 22-02-2021.

[46] A. Lopez-Rincon, "Emotion recognition using facial expressions in children using the nao robot," in *2019 International Conference on Electronics, Communications and Computers (CONIELECOMP)*, pp. 146–153, 2019.

[47] F. Martín, C. Durán, J. Plaza, G. Dago, R. Mejías, S. Espinosa, M. Valentí-Soler, and P. Martinez-Martin, "Robots in therapy for dementia patients," *Journal of Physical Agents*, vol. 7, no. 1, pp. 49–56, 2013.

[48] D. J. Ricks and M. B. Colton, "Trends and considerations in robot-assisted autism therapy," in *IEEE International Conference on Robotics and Automation*, (Anchorage, United States of America), pp. 4354–4359, IEEE, 2010.