Personalized Human-Robot Cognitive Interaction via a Novel Fuzzy Logic Control and Learning-Based Paradigm

Munster, Marcel; Jamshidnejad, Anahita

DOI
10.1109/ACCESS.2025.3584194

Publication date
2025

Document Version
Final published version

Published in
IEEE Access

Citation (APA)
Munster, M., & Jamshidnejad, A. (2025). Personalized Human-Robot Cognitive Interaction via a Novel Fuzzy Logic Control and Learning-Based Paradigm. *IEEE Access*, *13*, 112568-112593. https://doi.org/10.1109/ACCESS.2025.3584194

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 20 May 2025, accepted 25 June 2025, date of publication 30 June 2025, date of current version 8 July 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3584194

RESEARCH ARTICLE

Personalized Human–Robot Cognitive Interaction via a Novel Fuzzy Logic Control and Learning-Based Paradigm

MARCEL MUNSTER AND ANAHITA JAMSHIDNEJAD
Department of Control and Operations, Delft University of Technology, 2629 HS Delft, The Netherlands

Corresponding author: Anahita Jamshidnejad (a.jamshidnejad@tudelft.nl)

This work was supported in part by the Netherlands Organization for Scientific Research (NWO)-Open Competition Domain Science XS Project "Creative Social Robots with Evolutionary AI for Dementia Therapy," under Grant OCENW.XS2.055; and in part by the TU Delft AI Laboratories and Talent Program.

This work involved human subjects in its research. Approval of all ethical and experimental procedures and protocols was granted by the Human Research Ethics Committee of TU Delft under Approval 1952.

ABSTRACT Socially assistive robotics is an emerging field that, through effective human-robot cognitive interactions (HRCIs), offers potential solutions for personalized care, education, and entertainment. For improved impact and for acceptance by humans, socially assistive robots (SARs) should autonomously personalize and adapt their behavior to, respectively, the personality and the changes in the states-of-mind of people they interact with. Despite extensive research on the ethical, societal, and psychological aspects of SARs, bridging systems-and-control-based methods and socially assistive robotics for developing control approaches that automate the personalization and adaptation of HRCIs remains under-attended. We propose the first systematic and generalizable paradigm for personalization and adaptation of the social interactive behaviors of SARs, combining two highly promising modeling and decision making approaches, namely fuzzy logic control (FLC) and reinforcement learning (RL). By replicating the rule-based decision making of humans, FLC provides a highly effective personalization mechanism and warm-starts the RL algorithm, which takes care of adapting the behaviors of SARs to the dynamics of people's state-of-mind. Fuzzy logic is also used to develop two consecutive processes inside the RL-based adaptation module that, from the emotional responses of humans, estimate their state-of-mind and assign a reward to the most recent action of the SAR. Our extensive experiments for validation of this combined paradigm and for comparing it with conventional RL methods show meaningful improvements in the criteria that assess the personalization, convergence of learning, and performance accuracy of the proposed steering system for SARs.

INDEX TERMS Socially assistive robots, human–robot cognitive interaction, learning-based decision making, fuzzy logic control, personalization and adaptability of social robots.

I. INTRODUCTION

An emerging field in robotics that introduces potential solutions for personalized care, education, and entertainment concerns socially assistive robots (SARs) [1], [2], [3], [4]. SARs are expected to improve the engagement and reduce the stress level of people who deal with cognitive disorders or impairments (e.g., anxiety, depression, dementia, autism) during their therapy sessions, and to contribute to the development of smooth interactions and communication between the

TABLE 1. Frequently used abbreviations in the paper.

Abbreviation	Full term
SAR	Socially assistive robots
RL	Reinforcement learning
FLC	Fuzzy logic control
FIS	Fuzzy inference system
HRCI	Human-robot cognitive interaction

The associate editor coordinating the review of this manuscript and approving it for publication was Yiming Tang.

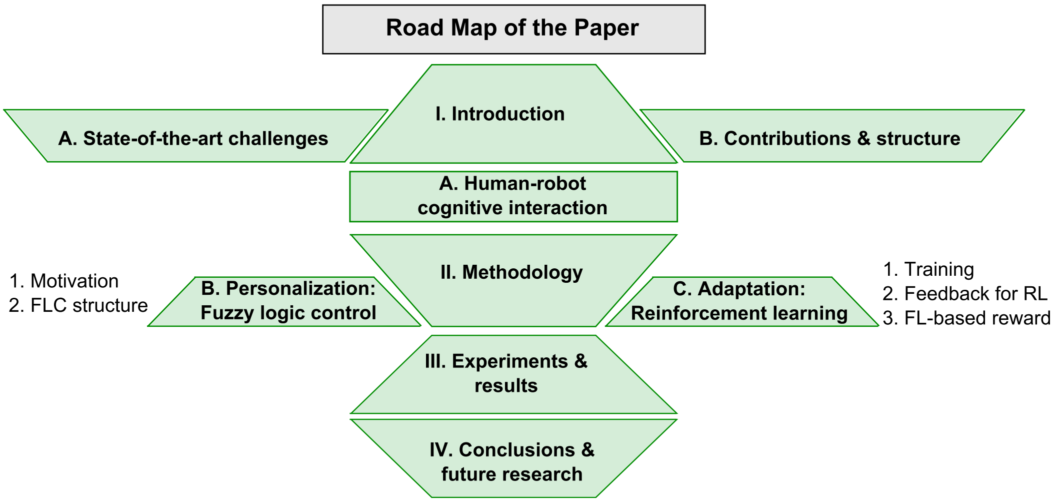

Road Map of the Paper
I. Introduction
A. State-of-the-art challenges
B. Contributions & structure
A. Human-robot cognitive interaction
II. Methodology
1. Motivation
2. FLC structure
B. Personalization: Fuzzy logic control
C. Adaptation: Reinforcement learning
1. Training
2. Feedback for RL
3. FL-based reward
III. Experiments & results
IV. Conclusions & future research

FIGURE 1. Road map of the paper.

therapists and patients [5], [6], [7]. SARs, however, are not meant solely for controlled therapeutic sessions. Ideally, they should accompany and assist people, including those with special needs, in their daily life activities [8], [9].

A. PROBLEM STATEMENT

Two key concepts, *personalizability* and *adaptability*, play a crucial role for SARs in sustaining meaningful human-robot cognitive interactions (HRCIs) with their users. Personalizability is the ability of tailoring the interactive social behavior of a SAR to the particular needs and preferences of each individual that interacts with the SAR. Adaptability allows the interactive behavior of a SAR to properly change with regards to the environmental/external variations that impact the HRCIs and the evolving states-of-mind of the human.

a: PROPOSED SOLUTION

Both personalizability and adaptability are crucial for SARs, in order to be accepted by humans and to succeed in engaging them within long-term social interactions [10], [11], [12], [13], [14]. Thus, control systems that are designed to steer the interactive social behavior of SARs should provide both capabilities [15]. This requires integrated behavioral steering methods aware of context-relevant states-of-mind of the users and their inter-personal variations. In other words, autonomy, reliability, and effectiveness of decision making by SARs demands (1) incorporating the implicit and explicit feedback of the users, (2) learning from interactions, and (3) handling diverse environmental conditions. These cannot be achieved by a solely learning-based, or intuitive, or control-theoretic method, but by a systematic integration of them.

We adopt a human-centered engineering approach. This involves identifying the needs and goals of user groups, using these to define the objectives and constraints in the development of HRCIs, and evaluating the resulting systems first via simulations based on human data collected through surveys. The step after this, as is detailed in Section IV, is to evaluate the developed systems through real-life experiments with human participants.

B. BACKGROUND AND STATE-OF-THE-ART CHALLENGES

Moyle et al. [16] report responses of dementia patients to SARs, especially accepting and maintaining interactions with them. Their results stress the importance of *personalization* and *adaptation* of the behavior of SARs for effective interactions with humans. This is further supported in [12] and [17]. In [17] the skill level of the participants, i.e., children with learning disabilities, is the basis for the SAR to *adapt* the difficulty level of its interactive tasks and the type of the feedback it provides for the children. In [12], the vocal content, activity level, and proxemics of the SAR are personalized based on the extroversion/introversion of the patients in a post-stroke rehabilitation therapy session. The results show that the participants prefer interactions by the SAR that resemble their own personality.

SARs are still very new to the societies, and research on SARs involves crucial psychological, societal, and ethical aspects to be addressed. A correlated topic is the development of systems-and-control-based paradigms that systematically steer the social behavior of SARs towards desired psychological, societal, and ethical impacts. However, despite extensive research on ethics of SARs and on variables that influence their societal and psychological impacts, incorporating or adopting systematic control methods to steer the social behavior of SARs remains under-attended [18], [19], [20]. Thus, the focus of this paper is on the development of a systematic, personalizable and adaptable, control paradigm for SARs.

Reinforcement learning is the most common method used in the last decade for SARs [20], [21]. A main disadvantage of RL, however, is its requirement for extensive trial-and-error-based interactions with participants during the learning phase that may negatively impact the interactions [22], [23]. In order to train an RL algorithm, it is common for human-robot interactions (see, e.g., [9], [17]) to use computer-based simulated participants. However, the need for further learning via trail-and-error for personalization to each real participant and for adaptation with respect to the varying state-of-mind and environmental conditions of the person remains unchanged. Conventionally, RL-based methods lack principle-based models that allow for generalizability and that represent the learned policies in relation to the dynamics and characteristics of the system that should be controlled by these policies. This implies that non-negligible changes in the states or dynamics of the system that RL steers, the (exhaustive) learning procedure should be conducted again. Finally, RL-based methods learn/evolve their policy according to the rewards that they receive per action generated via the candidate policy during the learning procedure. Mathematical formulation of a relevant reward function, however, is an open challenge for RL.

Three categories of learning-based methods exist: supervised learning, unsupervised learning, and reinforcement learning (RL). Supervised and unsupervised learning methods use labeled and unlabeled data, respectively, for training [24], [25], whereas RL learns an optimal policy based on the outcome of iterative interactions with a system. More specifically, RL bases its learning on three main elements: actions, states, and rewards per realized state for the selected actions. Learning methods commonly use feedback (e.g., the estimation error or the performance degradation) during the training in order to adjust their parameters accordingly. The feedback received in supervised learning is direct, through the labeled data, while unsupervised learning receives no explicit feedback. This poses a challenge on each method, i.e., the requirement for having access to sufficiently large labeled datasets in supervised learning, and the difficulty to interpret the results and to validate the trained system in unsupervised learning. The feedback on a candidate policy in RL is via the reward from the system. RL has proven to effectively learn policies that guarantee successful interactions with complex environments, where other learning methods would struggle. Accordingly, RL has been extensively used in interactions that need personalization, especially for SARs.

Static rule bases may also be used to steer the behavior of SARs, but due to a lack of systematic adaptability, their use remains limited to very simple and high-level interactions with humans, e.g., to explain the rules of a game for people, to use encouraging speech while a user plays a game or solves a puzzle, and to adjust the difficulty level of the game/puzzle based, solely, on the user's performance, e.g., the number of the correct answers so far (see, e.g., [26]).

A common use of SARs is for assisting people with autism spectrum disorder [27], [28]. In a recent paper [6], an adaptive, personalizable control system is introduced for a socially assistive drone that autonomously and interactively performs dance movement therapy (an interactive therapeutic method) for people with autism. The control framework uses fuzzy logic control and real-time image processing to let the drone interact with participants, such that their level of engagement in the therapy and their performance are incorporated in the decision making and are maintained at desired levels. In parallel, the control system adapts the rules in the fuzzy rule bases, according to the data it collects during the interaction sessions per participant, in order to personalize the interactions of the drone with the participant. A main advantage of fuzzy logic is its capability in replicating the rule-based decision making of humans (e.g., expert therapists) with affordable online, on-board computations. The results of the real-life experiments presented in [6] that involved a tiny quadcopter (a Parrot Bebop drone) and volunteer participants (without autism) proved the effectiveness of fuzzy logic control and the importance of personalizing and adapting the interactions in engaging the participants for longer terms.

Recent advances in fuzzy logic (see, e.g., [29]) allow to leverage this theory for effective modeling and control of dynamic variables (i.e., variables with a memory). This is particularly crucial since states-of-mind of humans are mathematically modeled as dynamic variables [18].

No systematic framework has yet been proposed for steering the social behavior of SARs that incorporates both personalization with respect to the personality traits of humans and adaptation to the evolution of their states-of-mind and environment. Moreover, the methods that have been proposed for the adaptation are mainly based on learning from the data captured from humans performing particular, simplistic tasks (e.g., solving simple math questions or puzzles). The state-of-the-art personalization approaches for SARs also simply mirror the personality of the users, which is not necessarily preferred for all people and in all interactive contexts.

C. MAIN OBJECTIVE AND CONTRIBUTIONS

We propose a novel integrated decision making paradigm for SARs to achieve the following **overarching objective**:

Allowing the social interactive behavior of SARs to be personalized to individual humans despite their different personalities and to be adapted to the changes in the state-of-mind of humans during the interactions, based on a systematic, generalizable approach

Our main contributions, leading to this objective, include:

1) The first **systematic** and **generalizable** paradigm for steering the social interactive behaviors of SARs that allows for simultaneous **personalization** and **adaptation**
2) A novel combination of fuzzy logic and reinforcement learning (RL), resulting in an RL-based adaptation module that, in addition to **precise action selection** (with regards to real preferences of humans) for SARs, is significantly more **efficient in learning** than conventional RL-based methods

3) Validation of the proposed paradigm through extensive computer simulations and simulated participants, designed based on data gathered from real humans via two different online surveys, designed for this research

D. RATIONALE OF THE PROPOSED METHODS

- Unlike state-of-the-art approaches where the SAR imitates the introversion level of a human in their interactions, we allow the SAR to interact with the human based on 3 relevant personality traits of that person (as identified by psychological tests).
- The personality-aware social interactive behavior of the SAR is not simplified to mimicry behavior, but is determined via a fuzzy inference system, based on heuristics and common sense of humans in social interactions.
- SARs are brought to a new level, significantly enhancing their adaptability, impact, and acceptability, through introducing a fuzzy-logic-based *dynamic cognitive model* that assesses and incorporates into the adaptation procedure the evolution of the states-of-mind of the human due to the interactions and environmental inputs.
- The learning procedure is *warm-started by the outputs of the personalization fuzzy inference system*, thus it requires significantly less number of exhausting trial-and-error-based interactions with humans.
- Next to the personalization procedure and the cognitive model used for adaptation, fuzzy logic, inspired by and enriched with heuristics and common knowledge of humans, is used to generate the rewards for the RL algorithm during the learning phase. This solves the lack of proper mathematical functions for the reward, especially in the context of human-robot cognitive interactions.

E. STRUCTURE OF THE PAPER

The rest of the paper is organized as it follows: Section II provides the main motivations for the methodologies used and the details of the proposed paradigm. Section III explains the setup and implementation of the experiments and presents and discusses the results. Finally, Section IV concludes the paper and proposes topics for future research. To guide the reader smoothly throughout the paper, a road map has been illustrated in Figure 1. Moreover, Table 1 and Appendix F represent the abbreviations used in the paper.

II. PROPOSED INTEGRATED CONTROL PARADIGM FOR SARS

Developing effective autonomous behavioral steering systems for SARs is a twofold problem: (1) The SAR should *personalize* its social interactive behavior to specific characteristics of the human it interacts with. (2) The SAR should *adapt* its social interactive behavior to the changes in the states-of-mind of the human. In broader contexts, next to the impact of the interactions on the evolution of the state-of-mind, the influence of the environmental factors relevant for the interactions may also be considered (see Figure 2). Our idea for obtaining such a steering system for SARs is based on a novel integration of fuzzy logic control (FLC) and RL, as is illustrated in Figure 3. The next sections motivate and discuss the details. Moreover, in addition to defining the mathematical notations when they appear in the text, Appendix A represents these notations and their definitions at once.

A. FRAMING THE HUMAN-ROBOT COGNITIVE INTERACTIONS

In this paper, we mainly focus on the cognitive, rather than physical, interaction of SARs and humans. Thus, the steering system of the SAR controls the cues and behavioral parameters that relate mainly to the social behavior of the SAR, in particular the following 7 elements/parameters:

1) Amount of speech
2) Volume of speech
3) Number of gestures
4) Type of interactive comments (e.g., energetic/cautious)
5) Type of motivating comments (e.g., cooperative/challenging)
6) Type of feedback to the human (e.g., realistic/nurturing)
7) Proxemics

Examples of interactive, motivating, and feedback statements that the SAR may use include:

- Energetic comment: *"I've got an amazing idea. Let's play this fun board game!"*
- Cautious comment: *"What about playing this board game? It could be fun!"*
- Cooperative comment: *"Let's try together to walk another round around the block!"*
- Challenging comment: *"I bet you cannot walk another round around the block! Want to prove me wrong?"*
- Realistic feedback: *"This time the exercises did not go as expected. Let's try different exercises next week!"*.
- Nurturing feedback: *"You will for sure do better next week with these aerobic exercises!"*

In the course of the interactions, the SAR regularly perceives the state-of-mind of the patient (for details see Section II-C3), and keeps on adapting its decisions considering the outcomes of the interactions, particularly the influence on the state-of-mind of the patient (see Section II-C4 for details).

B. FUZZY LOGIC CONTROL FOR PERSONALIZATION

The significance of considering the personality of the users in improving the outcomes of the HRCIs has been acknowledged, particularly in care domains [30], [31], [32]. In fact, any technology, including HRCI, that involves understanding, prediction, and synthesis of human behavior will benefit from methodologies that are capable of dealing with differences in human personalities [33]. However, the number of literature that systematically incorporate the

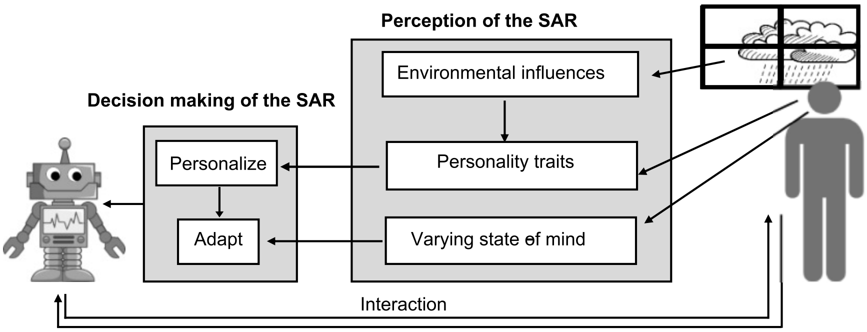


FIGURE 2. The decision making system of the SAR personalizes and adapts the actions of the robot according to the characteristics and sate-of-mind of the human, considering also the environmental factors that influence the human-SAR interactions.

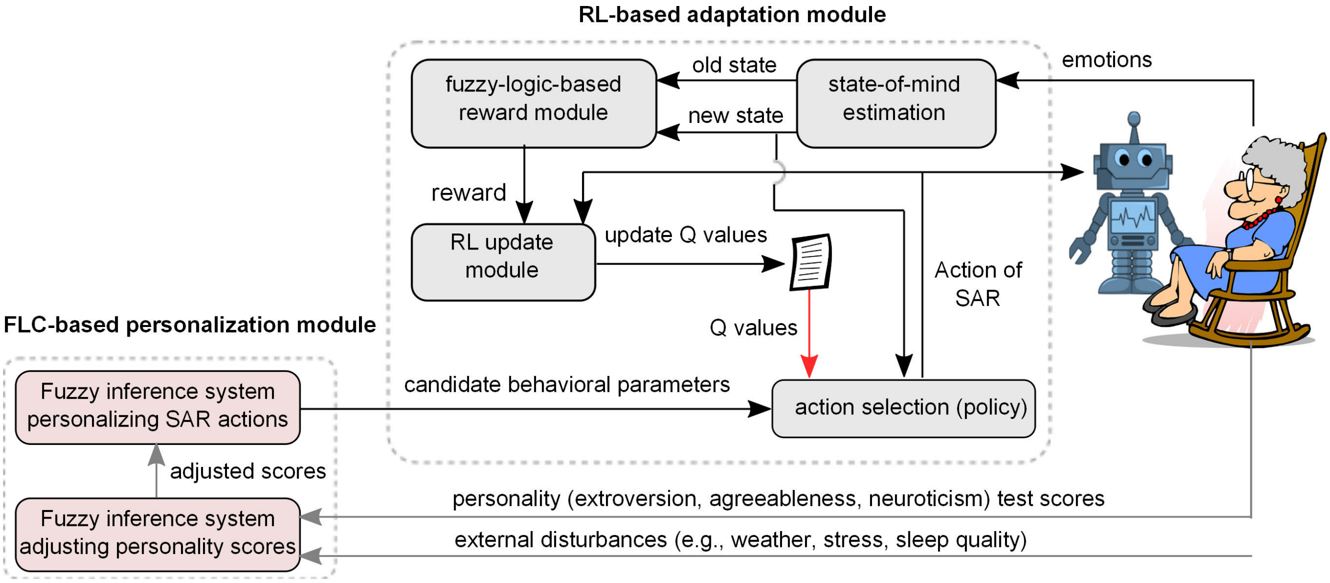

RL-based adaptation module
fuzzy-logic-based reward module
old state
new state
state-of-mind estimation
emotions
reward
RL update module
update Q values
Q values
Action of SAR
FLC-based personalization module
Fuzzy inference system personalizing SAR actions
candidate behavioral parameters
action selection (policy)
adjusted scores
Fuzzy inference system adjusting personality scores
personality (extroversion, agreeableness, neuroticism) test scores
external disturbances (e.g., weather, stress, sleep quality)

FIGURE 3. Different elements proposed for the decision making of SARs(FLC and RL stand for fuzzy logic control and reinforcement learning, respectively).

influence of the personality traits of humans into the decision making of SARs is very narrow [34], [35].

The scarce literature that considers this (see, e.g., [12], [36], [37]) mainly focuses on *matching* the personality of the SAR and the human, particularly for one trait, i.e., extroversion.[1] Research, however, shows that agreeableness[2] and neuroticism,[3] together with extroversion, play the main role in cognitive and social interactions of humans, especially for sensitive groups, e.g., for patients with dementia [35], [38], [39]. Moreover, the claim that congruence between the personalities (for example of clients and therapists, whose social interactions with their clients may inspire the design of the steering systems of SARs) could result in more impactful interactions (e.g., enhancing desired therapeutic outcomes) has been tested for very specific cases (see, e.g., [36]) where the SAR interacts with very specific, thus non-representative, user groups (e.g., undergraduate students or community members of a specific age range with no diagnosis of cognitive impairments [40]). In-depth research on the impact of the personality congruence, however, rebuts any positive impact of such congruence, especially for agreeableness and neuroticism, on the outcome of therapeutic interactions [41].

Therefore, instead of mirroring the same personality traits, within the personalization module of the SAR we directly implement the knowledge (based on heuristics and common sense) of an average human in their daily social interactions. In particular, next to extroversion, we consider agreeableness and neuroticism. The scores, ranged between 0 and 100, for each of these personality traits can be known via the Big Five personality test [42]. The light red boxes in Figure 3 build the personalization module, which receives the scores for the personality traits of the person and generates the corresponding behavioral (communication or cue) parameters for the SAR.

[1] High levels of extroversion are indicative of higher sociability, assertiveness, and activity.

[2] High levels of agreeableness are indicative of higher modesty, cooperativeness, trustworthiness, and concern about the feelings of others.

[3] High levels of neuroticism are indicative of higher levels of anxiety, insecurity, and depression.

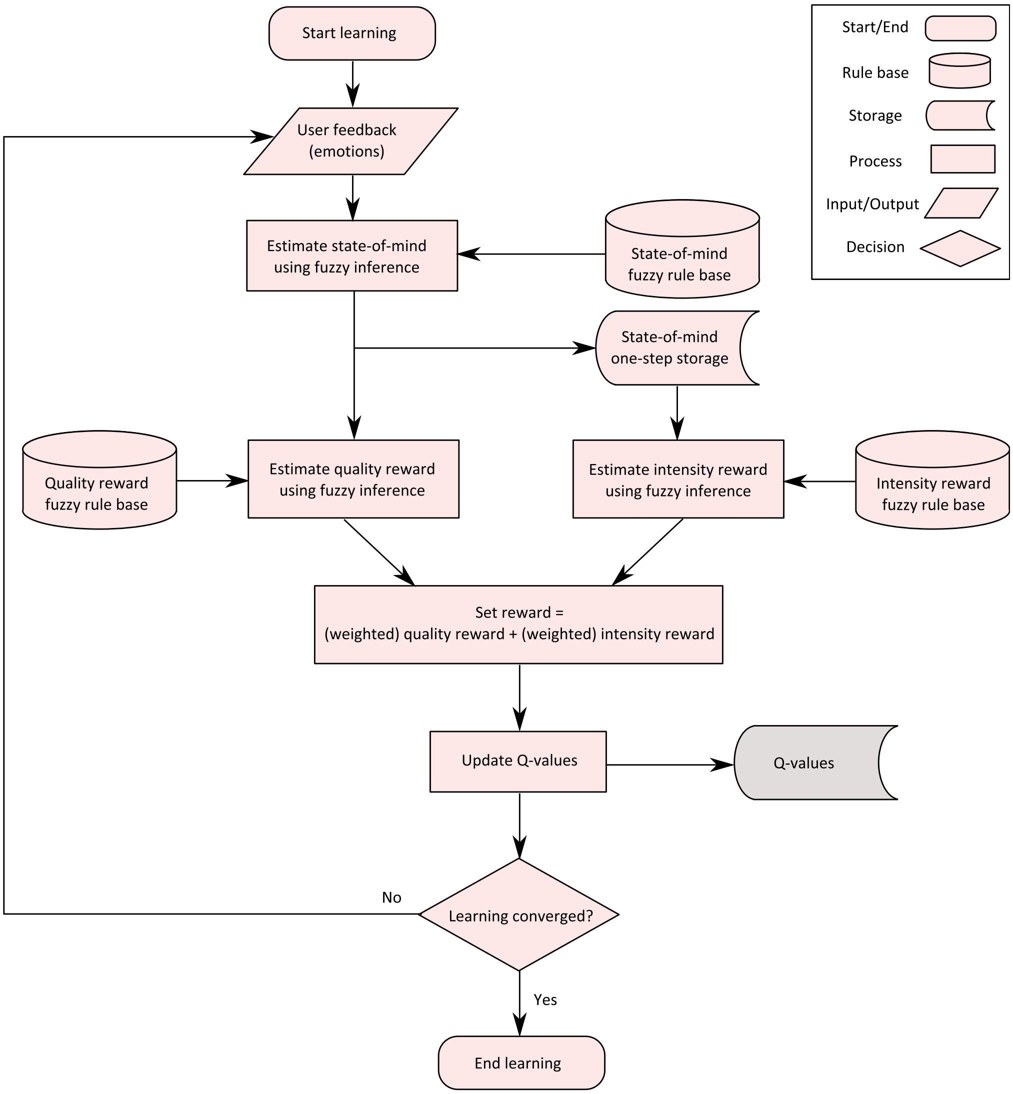

Start learning
User feedback (emotions)
Estimate state-of-mind using fuzzy inference
State-of-mind fuzzy rule base
State-of-mind one-step storage
Quality reward fuzzy rule base
Estimate quality reward using fuzzy inference
Estimate intensity reward using fuzzy inference
Intensity reward fuzzy rule base
Set reward = (weighted) quality reward + (weighted) intensity reward
Update Q-values
Q-values
Learning converged?
No
Yes
End learning
Start/End
Rule base
Storage
Process
Input/Output
Decision

FIGURE 4. Flowchart representing the learning procedure for the Q-learning adaptation module (explanations are included in Section II-C2): During the learning procedure, Q-values are updated (see Appendix C for details) and are stored in a database. Estimation of the state-of-mind based on the user's feedback, i.e., emotions, is discussed in Section II-C3. The procedure of estimating the reward, using fuzzy inference systems, is given in detail in Algorithm 1 in Section II-C4.

1) MOTIVATION FOR USING FUZZY LOGIC CONTROL

Fuzzy logic control (FLC), a rule-based method that mimics reasoning of humans and that effectively deals with linguistic variables [43], [44], [45], [46], [47], is used in this paper to personalize the decisions of the SAR based on the personality traits of humans. For more details regarding FLC see Appendix B.

FLC was selected for the personalization of the SAR, due to the following main reasons:

1) There are no mathematical models that explicitly relate the personality scores with the behavior that people prefer to experience from others in social interactions. This relationship, however, can properly be captured via fuzzy if-then rules, based on existing human knowledge (see, e.g., [38] for some behaviors that, according to the Big Five personality traits, people with each personality trait prefer to experience in social interactions).
2) Personality traits and variables that relate to them are described across spectra, and associating precise crisp quantities to them is either impossible or erroneous. Personality traits and their related variables are by nature fuzzy [46], so can most properly be captured by inference systems that apply fuzzy sets and fuzzy operations. FLC is built upon such inference systems and handles partial memberships of fuzzy variables that relate to the personality traits to fuzzy sets that represent the corresponding quantitative terms (e.g., relatively high).

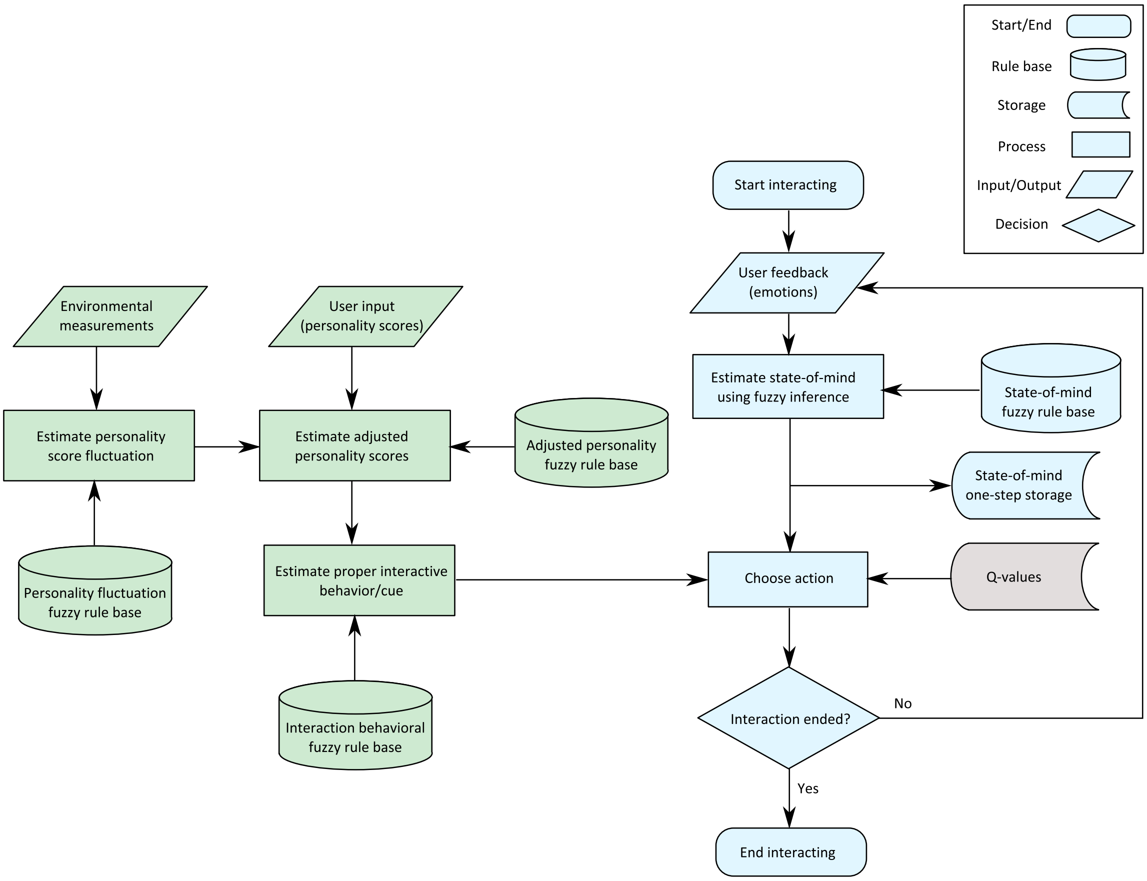

Start/End
Rule base
Storage
Process
Input/Output
Decision
Start interacting
User feedback (emotions)
Estimate state-of-mind using fuzzy inference
State-of-mind fuzzy rule base
State-of-mind one-step storage
Environmental measurements
User input (personality scores)
Estimate personality score fluctuation
Estimate adjusted personality scores
Adjusted personality fuzzy rule base
Personality fluctuation fuzzy rule base
Estimate proper interactive behavior/cue
Choose action
Q-values
Interaction behavioral fuzzy rule base
Interaction ended?
No
Yes
End interacting

FIGURE 5. Flowchart illustrating the interactions between the FLC-based personalization module (green part of the flowchart) and the RL-based adaptation module (blue part of the flowchart) for generating a suited interactive behavior for the SAR: The formulation of the three fuzzy rules that are incorporated in the FLC-based personalization module is discussed in Section II-B2. The gray database that includes the Q-values is generated through the learning procedure illustrated in Figure 4.

3) Humans properly personalize their behavior in the course of interaction with each other [10]. Rule-based approaches based on fuzzy logic, e.g., FLC, are the closest methods to the reasoning and decision making of humans [48], [49]. Thus, we hypothesize that FLC will provide SARs with personalization capabilities similar to those that humans deploy in social interactions.

Remark 1: Methods other than FLC, e.g., neural networks [50] and RL [51], have been used to personalize the behavior of SARs to humans. These methods, however, require extensive training with large datasets. Thus, during the training there are risks for reduced performance, lost effectiveness, and thus withdrawal of the users. On the contrary, the rule bases of FLC-based methods are built upon the existing expert knowledge and from the start if HRCIs, incorporate human heuristics in the interactions.

2) STRUCTURE OF THE PERSONALIZATION FLC MODULE

Querengässer and Schindler in [52] showed that the scores of the personality tests are influenced by the state-of-mind of the participants of these tests. This implies that to consider fixed personality trait scores for humans is neither realistic, nor is it effective for sustaining long-term HRCIs. Therefore, two fuzzy inference systems (FISs), as is shown in Figure 3 and is represented via the flowchart in Figure 5, are considered for the personalization. The general formulations of the fuzzy rules that are used by the SAR to, respectively, adjust the personality trait scores and personalize its social behaviors based on these scores for each human include:

Formulation of the fuzzy rules for the personalization

Personlization FIS 1

If *external variable 1* is V_{1,v_1} **and** … **and** *external variable* n^{ext} is $V_{n^{\text{ext}},v_{n^{\text{ext}}}}$, **then** *score fluctuation* for personality trait τ is F_{τ,f_τ}. $\qquad \tau = 1, 2, 3$

Personlization FIS 2

If *personality trait* τ is P_{τ,p_τ} and *fluctuation score* for personality trait τ is F_{τ,f_τ}, **then** *adjusted personality trait* τ is P_{τ,a_τ}. $\qquad \tau = 1, 2, 3$

Personlization FIS 3

If *adjusted personality trait 1* is P_{1,a_1} **and** *adjusted personality trait 2* is P_{2,a_2} **and** *adjusted personality trait 3* is P_{3,a_3}, **then** *interactive behavior* (*cue parameter*) is B_{m_1} (is C_{m_2}).

In the first set of rules, which correspond to FIS 1 (see the top plot in Figure 6 as an example for FIS 1), *external*

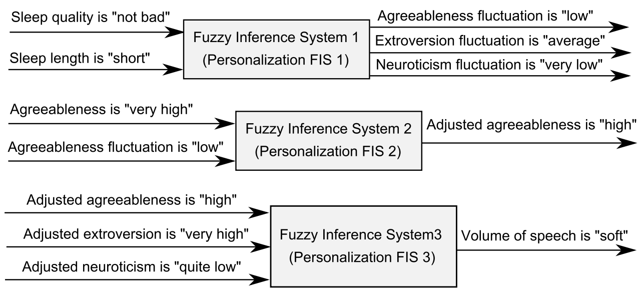

Sleep quality is "not bad"
Sleep length is "short"
Fuzzy Inference System 1
(Personalization FIS 1)
Agreeableness fluctuation is "low"
Extroversion fluctuation is "average"
Neuroticism fluctuation is "very low"
Agreeableness is "very high"
Agreeableness fluctuation is "low"
Fuzzy Inference System 2
(Personalization FIS 2)
Adjusted agreeableness is "high"
Adjusted agreeableness is "high"
Adjusted extroversion is "very high"
Adjusted neuroticism is "quite low"
Fuzzy Inference System3
(Personalization FIS 3)
Volume of speech is "soft"

FIGURE 6. **Fuzzy inference systems that adjust the scores for thepersonality traits of the users depending on the external factors that impact the state-of-mind of the users,and that determine candidate behaviors and cues for the SAR in the course of HRCIs.**

variable 1, …, *external variable* n^{ext} refer to the external factors that impact the state-of-mind, thus the personality trait scores, of the person. For instance, these may be the quality and the length of the sleep of the person, or, based on the context, include the weather conditions, e.g., the precipitation and the temperature. In addition, V_{1,v_1}, …, $V_{n^{\text{ext}},v_{n^{\text{ext}}}}$ (for $v_i = 1, \ldots, N_i^{\text{v}}$ and $i = 1, \ldots, n^{\text{ext}}$) are fuzzy sets corresponding to the linguistic terms that are used to categorize these external variables. In these definitions, N_1^{v}, …, $N_{n^{\text{ext}}}^{\text{v}}$ are the number of those categories. Similarly, F_{τ,f_τ} for $f_\tau = 1, \ldots, N_\tau^{\text{f}}$, is a fuzzy set representing the linguistic term that describes the fluctuations in the score of personality trait τ, with N_τ^{f} the total number of these terms. Note that $\tau = 1, 2, 3$ refers to the three personality traits extroversion, agreeableness, and neuroticism.

In the second set of rules, which correspond to FIS 2 (see the middle plot in Figure 6 as an example for FIS 2), the three personality traits are adjusted based on their qualification obtained from the Big Five Personality test and based on the fluctuations due to the external factors obtained from the first set of rules. Note that P_{τ,p_τ} is the fuzzy set that describes personality trait τ, where $p_\tau = 1, \ldots, N_\tau^{\text{p}}$ and for $\tau = 1, 2, 3$ the number of these fuzzy sets are, respectively, $N_1^{\text{p}}, N_2^{\text{p}}, N_3^{\text{p}}$. For instance, when the personality traits are all categorized as *low*, *medium*, or *high*, then $N_1^{\text{p}} = N_2^{\text{p}} = N_3^{\text{p}} = 3$.

In the third set of rules, which correspond to FIS 3 (see the bottom plot in Figure 6 as an example for FIS 3), the three adjusted personality trait qualities, i.e., P_{1,a_1}, P_{2,a_2}, P_{3,a_3} where $a_i \in \{1, \ldots, N_i^{\text{p}}\}$ for $i = 1, 2, 3$, are used to generate the candidate interactive behavior or cues of the SAR. Moreover, B_{m_1} for $m_1 = 1, \ldots, N^{\text{b}}$ is a fuzzy set modeling the linguistic terms that describe the nature of a social interactive behavior by the SAR, i.e., an interactive comment, a motivating comment, a realistic feedback, or a nurturing feedback (see Section II-A), and N^{b} is the number of different terms that describe these comments. Finally, C_{m_2} for $m_2 = 1, \ldots, N^{\text{c}}$ is a fuzzy set modeling the linguistic terms that describe the parameters corresponding to the SAR cues (e.g., *medium* for the *volume of speech*) and N^{c} is the total number of these terms. The resulting personalization FIS whose rule bases are composed of fuzzy rules with the given formulations is a Mamdani inference system [53].

C. RL FOR ADAPTATION

While personalization of the SAR decisions with respect to the personality traits of a patient is expected to improve the quality of HRCIs, and to sustain the engagement of the human and thus, the interactions for longer terms [34], [35], such a personalization (i.e., considering an optimal action per situation for a specific personality trait) by itself is not sufficient for effective/engaging long-term interactions, especially when these interactions follow specific, e.g., therapeutic, goals. This is further supported by the proven essence and impact of the creativity and variation in the responses of SARs for sustainable interactions in HRCI [7], [10]. In other words, providing the same response over and over, even if personalized to the human, is not effective in the long term. Therefore, the SAR should adapt its behaviors to the state-of-mind of the person. We use RL in order to adapt the social interactions of SARs to different states-of-mind of people.

1) MOTIVATION FOR USING RL

The main motivation for selecting RL is the following: While personalization of the SAR behaviors via establishing general rules per personality trait and by implementing FLC (see Section II-B for details) appreciates the distinction among various categories of a personality trait (e.g., very extrovert and moderately extrovert), this does not incorporate the variations per individual in each category. For instance, highly introvert people may in general prefer to have a short conversation with a SAR without exhibiting high excitements through the volume of the speech. When feeling sad (i.e., a particular state-of-mind), however, one highly introvert individual may be more sensitive about the volume of the speech, whereas another highly introvert individual may prefer to use less statements in a conversation. Such a difference, which specifies the importance or weight of each fuzzy rule in the personalization module (see Section II-B), can be learned using RL in the course of the interactions with each person.

2) TRAINING THE RL MODULE

The RL module operates in a loop (see Figure 3 and the flowchart represented in Figure 4): The SAR interacts with the human and per interaction step receives the emotions of the human as feedback, which it uses (as is detailed in Section II-C3) to determine how the state-of-mind of the person has evolved in the last interaction step. This evolution of the state-of-mind is the basis for determining the reward, which reflects the effectiveness of the social interactive behaviors of the SAR. **A crucial aspect of the proposed paradigm is that the RL module is warm-started by the action that is proposed for the SAR via the FLC-based personalization module** (see Figure 3). This is hypothesized to reduce the number of the learning iterations and thus, the number of the trial-and-errors.

In the RL framework, the state-of-mind of the human is considered as the state variable s_k of the RL module and the change, with respect to the candidate action proposed by the FLC-based personalization module, regarding any of the behavioral elements of the SAR (where the 7 behavioral elements considered for the SAR in this paper have been indicated in Section II-A) is the action a_k of the RL. The integer k specifies the interaction step. The main question to answer is that *during the training phase of the RL module, how, and based on what rationale, does the SAR select an action, a_k, at interaction step k when the state-of-mind of the human is s_k?*

Since the number of the sate-action pairs in the given HRCI context is finite, the Q-learning algorithm with an ϵ-greedy approach [51] can be used by the RL module. For details regarding this approach, we refer the readers to Appendix C. After sufficient number of human-SAR interactions, when the RL algorithm converges (which implies that the training has sufficiently been performed), the trained RL module will directly be used to steer the social interactive behavior of the SAR.

3) FEEDBACK FOR THE RL MODULE DURING THE TRAINING PHASE

The RL adaptation module receives receives the emotional responses of users (collected via self-report, video/image analysis, or multi-modal sensors) as feedback during the HRCIs with SARs.

The level of engagement of people in social interactions reflects their state-of-mind, which is a consequence of all the active emotions that they experience and that bind them to other individuals, places, and activities [54]. Compared to the abstract concept of state-of-mind, i.e., the overall cognitive state of a person, providing qualified assessments for different emotions is more intuitive and straightforward for humans, especially when precision is crucial. Moreover, there are automated algorithms in literature (see, e.g., [55]) that detect the emotions of a human from, e.g., their facial expressions. Thus, we use the basic emotions, according to [56] (i.e., surprise, joy, sadness, fear, anger, disgust, trust, anticipation) as the feedback for RL. Hence, in order to learn about the impact of its social interactive behaviors, the SAR keeps track of the emotions of the human, and adapt its interactive behaviors accordingly to maintain the HRCIs.

Remark 2: In case real-time detection of the emotions from the images/videos that capture the facial/body expressions of humans is challenging or impossible (e.g., for people in advanced stages of dementia or with the Parkinson's disease [57]) an additional FLC module may be used that estimates the state-of-mind from other measurable variables. Such an FLC module can be expanded to include the impact of the environmental and external circumstances (e.g., the weather condition, the stress level, the quality of sleep; see, e.g., [58], [59] on how these variables can be obtained) on the state-of-mind. This, however, is out of the scope of this paper.

In general, emotions, which are the main user feedback received by the integrated behavioral control paradigm, are experienced and assessed by humans in fuzzy terms, which lack clear distinctions. For instance, joy and sadness, although opposite, may at the same time be experienced by a person. For proper incorporation of this fuzziness into the analysis and computations for the decision making of SARs, the emotions are represented as fuzzy variables, which allow the overlapping emotions that may simultaneously exist (e.g., joy and sadness) to be represented via fuzzy membership functions with overlaps.

The fuzzy scores for the emotions provided by a user will be injected into a FIS that estimates the current state-of-mind of the human, according to a fuzzy rule base (see Figure 3). The corresponding fuzzy rules are generally given by:

Fuzzy rule for estimating the state-of-mind

If *emotion 1* is E_{1,i_1} **and** … **and** *emotion* n^{emot} is $E_{n^{\text{emot}},i_{n^{\text{emot}}}}$, **then** *state-of-mind* is S_ℓ.

where *emotion 1*, …, *emotion* n^{emot} are n^{emot} different emotions, especially those that are correlated (e.g., surprise and trust), and $E_{1,i_1}, \ldots, E_{n^{\text{emot}},i_{n^{\text{emot}}}}$ for $i_1 = 1, \ldots, N_1^{\text{e}}$, …, $i_{n^{\text{emot}}} = 1, \ldots, N_{n^{\text{emot}}}^{\text{e}}$ are fuzzy sets corresponding to the linguistic terms that categorize these emotions, with N_1^{e}, …, $N_{n^{\text{emot}}}^{\text{e}}$ the number of these categories. Moreover, S_ℓ for $\ell = 1, \ldots, N^{\text{s}}$ is a fuzzy set that represents the linguistic term that describes the state-of-mind, with N^{s} the total number of these linguistic terms.

Next, another FIS is used to assign a reward to the action of the SAR that has resulted in this state-of-mind for the human.

4) FUZZY-LOGIC-BASED REWARD MODULE

One of the fundamental challenges in the implementation of RL for real-life problems is to choose a suitable and relevant reward function for the given application [60], [61]. In particular, for our application of RL, i.e., steering the social interactive behavior of SARs, there are no mathematical functions that explicitly describe how emotions, state-of-mind, and criteria of success in social interactions are related. However, such relationships can be formulated intuitively, using linguistic fuzzy rules. Thus, we propose a FIS to compute the rewards based on both the quality and the intensity of the changes, with respect to the previous interaction step, in the state-of-mind.

For instance, if the state-of-mind of a human is improved (is worsened) after interacting with the SAR, the corresponding decision of the SAR is rewarded (is punished). This implies rewarding the SAR based on the influence it has on the quality of the changes in the state-of-mind of the person. Moreover, an intense change in the state-of-mind of the person is rewarded/punished more significantly. For example, if the state-of-mind transitions from very sad to happy, the corresponding reward is more significant than when the state-of-mind transitions from happy to very happy.

Algorithm 1 Reward Estimation Procedure for the RL Module, Based on the Emotions (feedback) Captured From Humans

```
Variables, functions, and parameters
    $k$: Counter for the interaction steps
    $k^{\text{int}}$: Last interaction step
    $\mathbb{E}_k$: Set of all emotions of the human relevant for the HRCI, given as fuzzy values at interaction step $k$
    $\tilde{S}_k$: Sate-of-mind of the human, given as a fuzzy value, for interaction step $k$
    $\text{FIS}_{\text{SoM}}(\cdot)$: Fuzzy inference system estimating the state-of-mind of the human based on the emotions
    $\text{FIS}_{\text{QoR}}(\cdot)$: Fuzzy inference system estimating the quality component of the reward for the RL module, based on the current state-of-mind of the human
    $\text{FIS}_{\text{IoR}}(\cdot,\cdot)$: Fuzzy inference system estimating the intensity component of the reward for the RL module, based on the current and previous states-of-mind of the human
    $w_1$ and $w_2$: Given weight parameters
    $\rho_k$: Reward of the RL module for interaction step $k$
    $\tilde{Q}_k$: Fuzzy quality component of the reward estimated for the RL module at interaction step $k$
    $\tilde{I}_k$: Fuzzy intensity component of the reward estimated for the RL module at interaction step $k$
    Defuzzify$(\cdot)$: Operator that defuzzifies fuzzy inputs
Reward estimation process
 1: $k \leftarrow 1$
 2: while $k < k^{\text{int}}$ do
 3:     Collect $\mathbb{E}_k$ from the human
 4:     $\tilde{S}_k \leftarrow \text{FIS}_{\text{SoM}}(\mathbb{E}_k)$
 5:     $\tilde{Q}_k \leftarrow \text{FIS}_{\text{QoR}}(\tilde{S}_k)$
 6:     if $k > 1$ then
 7:         $\tilde{I}_k \leftarrow \text{FIS}_{\text{IoR}}(\tilde{S}_{k-1}, \tilde{S}_k)$
 8:     else
 9:         $k \leftarrow k + 1$
10:         Go to line 2
11:     end if
12:     $\rho_k \leftarrow w_1\,\text{Defuzzify}(\tilde{Q}_k) + w_2\,\text{Defuzzify}(\tilde{I}_k)$
13:     $k \leftarrow k + 1$
14: end while
```

This implies that the rewarding is also based on the intensity of the influence of the SAR on the state-of-mind of the person. Correspondingly, when the initial state-of-mind of the person is s_k and the SAR takes action a_k, which results in state-of-mind $s_{k+1}(a_k)$, the reward $r\left(s_k, a_k\right)$ is given by:

$$r\left(s_k, a_k\right) = w_1 r^{\text{quality}}\left(s_{k+1}(a_k)\right) + w_2 r^{\text{intensity}}\left(s_k, s_{k+1}(a_k)\right) \tag{1}$$

where $r^{\text{quality}}\left(s_{k+1}(a_k)\right)$ and $r^{\text{intensity}}\left(s_k, s_{k+1}(a_k)\right)$ are the partial rewards based on, respectively, the quality of the new state-of-mind and the intensity of the change in the state-of-mind, and w_1 and w_2 are weights. The fuzzy rules for determining these two components are generally given by:

Fuzzy rules for determining the reward

If *new state-of-mind* is S_i, **then** *quality component of the reward* is Q_ℓ.

If *new state-of-mind* is S_i **and** *old state-of-mind* is O_j, **then** *intensity component of the reward* is I_m.

with S_i and O_j the linguistic terms that describe the new and old states-of-mind, Q_ℓ and I_m the linguistic terms that describe the quality and intensity components of the reward, $i, j = 1, \ldots, N^{\text{s}}$, $\ell = 1, \ldots, N^{\text{q}}$, $m = 1, \ldots, N^{\text{i}}$, and N^{q} and N^{i} the number of the linguistic categories for, respectively, the quality and intensity components of the reward. The fuzzy-logic-based reward module uses a Mamdani FIS with the center of gravity method for defuzzification.

The procedure of estimating the rewards for the RL module, using FISs, has been summarized in Algorithm 1.

Remark 3: User feedback is incorporated into the proposed behavioral control paradigm for SARs as follows: During RL training, user emotions serve as primary feedback. A FIS converts these emotions into states-of-mind, which then inform another FIS to estimate rewards for the RL module. This enables the RL module to learn actions suited to various external and internal states for users. Once learning converges, the FLC-based personalization module processes user feedback, allowing the trained RL module to select actions (based on Q-values) that best match the current state-of-mind of the user (see the red arrow in Figure 3).

By using fuzzy logic at both the feedback interpretation and reward shaping stages, the system ensures interpretability, flexibility, and resilience to uncertainty, all being crucial for real-world human-robot interaction.

III. EXPERIMENTS AND RESULTS

A road map that clarifies the body of this section has been illustrated in Figure 7. In order to assess both the FLC-based personalization module (see Section II-B) and the RL-based adaptation module (see Section II-C) within the proposed decision making paradigm for steering the social behavior of SARs, a set of experiments were designed and executed.

A. EXPERIMENT DESIGN

This section explains the design of the experiments, as well as the choice and processing of the data of the participants.

1) EXPERIMENTAL STAGES

The experiments were composed of three stages:

Stage one. We asked 20 volunteer participants to fill in an online survey designed for the purpose of these experiments. Their responses were used to identify the parameters of the fuzzy membership functions used by the inference system of the personalization module, and to evaluate the performance of the identified FLC-based personalization module.

Stage two. We asked the same participants to fill in a second online survey. The data gathered from the responses was used to model an extensive number of virtual, computer-based participants and their cognitive responses in interactions with

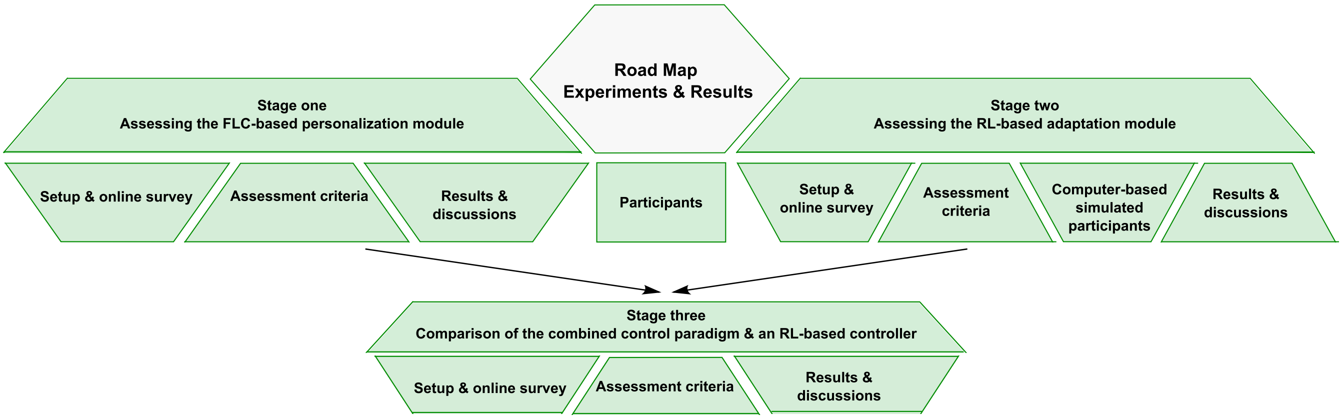


FIGURE 7. Road map of the experiments and results section.

SARs. These simulated participants were then used both to train and to evaluate the RL-based adaptation module.

Stage three. We compared the performance of the proposed combined paradigm and a conventional RL-based steering system that is common for SARs. We used the computer-based simulated participants for the training and evaluation of these two steering systems.

2) PARTICIPANTS

This research introduces very novel methods to steer SARs. Accordingly, due to ethical and safety concerns, preliminary proofs-of-concept are required before involving humans, especially vulnerable or sensitive groups who are the main target groups that will benefit from SARs. Moreover, running the experiments, which involve trial-and-error-based training, identification, and validation phases, with real participants may in general be frustrating, potentially harmful, or boring, such that the reliability of the results is impacted. With computer-based simulated participants, however, a large number of interactions can be conducted and analyzed in a reasonable time. Furthermore, a wide variety of interactive scenarios, including different potential responses of the simulated participants to the decisions of a SAR that uses the proposed adaptation approach, can be simulated and assessed. Obtaining such level of variety is overly difficult, if not impossible, by considering even large groups of people.

At this stage of the research, we thus used data from participants who do not belong to vulnerable, sensitive groups with cognitive impairments, and computer-based simulated participants for modeling long-term interactions with a SAR that uses the proposed FLC-based and RL-based personalization and adaptation approaches. Nevertheless, these mechanisms have been developed according to generalized theories given in Sections II-B and II-C. Therefore, if they prove effective for human-SAR interactions with rules and policies that have been identified and trained based on the cognitive needs of the participants without cognitive impairments, by adjusting the rule base and the policies according to the cognitive needs of participants with cognitive impairments (e.g., dementia patients), the approach is expected to work properly.

B. IMPLEMENTATION

Next, we provide details on the implementations, the deduction of the required information based on the data gathered from our participants, and the processing procedures of the data in order to generate the personalization FLC module. We also present the corresponding results and discussions.

1) STAGE ONE: ASSESSING THE FLC-BASED PERSONALIZATION MODULE

The FLC-based personalization module (see Section II-B and Figure 3) computes, based on the scores of the Big Five personality test for a particular participant, the initial values for the 7 behavioral parameters (amount of speech, volume of speech, number of gestures, relative number of interactive (energetic vs cautious) comments, relative number of motivating (cooperative vs challenging) comments, relative number of realistic and nurturing feedback, and the proxemics) that a SAR uses in interactions with that person.

a: SETUP & FIRST ONLINE SURVEY

Appendix D gives detailed information on the rule bases of the FISs that were used to adjust the scores of the personality traits and to personalize the behavioral elements of the SARs based on these scores (see Table 5 designed based on [12], [38], [62], and Figures 23 and 24 in Appendix D). Fine-tuning the personality scores of the participants with respect to their environmental variations (e.g., the weather conditions, the level of stress, the quality of sleep) requires extensive interactions that expose the participants to different circumstances, such that there is enough data that reflects the variations in the preferences of the participants with respect to these external factors. We thus excluded this from the experiments.

An online anonymous survey was designed and shared with the participants, who were asked to first fill in the Big Five personality test [42] and enter their scores in the survey (see Figure 25 in Appendix D). Then, they were asked to provide their answers to 10 questions that assessed, on a scale from 1 to 5 (with a step of 0.1), their preferences in various social interactive scenarios (see Figure 26 in Appendix D for an example question from the survey). The complete survey is accessible online via [63]. Out of the 20 sets of results obtained from the survey, 14 datasets were used to identify the membership functions of the FLC-based personalization module, and 6 datasets were used to validate these functions.

b: ASSESSMENT CRITERIA

The criteria used to train and validate the personalization module included the root mean squared error (RMSE):

$$\mathrm{RMSE}(i) = \sqrt{\frac{\sum_{j=1}^{N^{\mathrm{part}}} \left(b_j^{\mathrm{FLC}}(i) - b_j^{\mathrm{part}}(i) \right)^2}{N^{\mathrm{part}}}} \tag{2}$$

with i an index for the behavioral element of the SAR ($i = 1$ for the amount of speech, $i = 2$ for the volume of speech, $i = 3$ for the number of gestures, $i = 4$ for the relative number of interactive (energetic/cautious) comments, $i = 5$ for the relative number of motivating (cooperative/challenging) comments, $i = 6$ for the relative number of realistic and nurturing feedback, $i = 7$ for the proxemics), $b_j^{\mathrm{FLC}}(i)$ the value of the behavioral parameter computed by the FLC-based personalization module for interaction with participant j, $b_j^{\mathrm{part}}(i)$ the preferred value of the behavioral parameter for participant j based on the results obtained from the online survey, and N^{part} the number of participants.

Additionally, the scatter index (SI) was used to account for the range over which the different values were observed:

$$\mathrm{SI}(i) = \frac{\mathrm{RMSE}(i)}{\frac{1}{N^{\mathrm{part}}} \sum_{j=1}^{N^{\mathrm{part}}} b_j^{\mathrm{FLC}}(i)} \tag{3}$$

c: RESULTS AND DISCUSSION OF THE RESULTS

The personalization module generates a candidate output for each behavioral element of the SAR based on the consequent of the fuzzy rules (see Table 5, Appendix D). Thus, first, the parameters of the fuzzy membership functions that represent the terms in the consequent of the rules (e.g., low, soft, neutral, loud, …) were identified such that the RMSE for the output of the identified FIS with respect to the training dataset (i.e., the responses that were provided by the 14 participants of the first online survey, whose data was included in the training dataset) was minimized. Figure 24 in Appendix D illustrates the identified membership functions.

The identified personalization module was then evaluated via the validation dataset (which included the responses of the other 6 participants of the first online survey). Figure 8 shows the scores, based on the Big Five personality test, for the extroversion, agreeableness, and neuroticism for these participants, denoted by P1, …, P6. The outputs for the 7 behavioral elements of the SAR, as preferred by the participants according to the identified personalization FLC module, were compared and illustrated in Figures 9-15. These outputs were scaled to the range of 1 to 5, which matches the scaling of the answers by the participants for the first online survey.

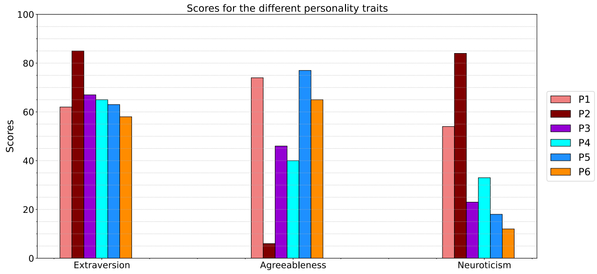


FIGURE 8. Personality scores corresponding to the validation data set.

Moreover, the corresponding values of RMSE and SI are displayed in Table 2, where the values in the second column of the table are the RMSE values within the scaled range of 1 to 5 and the values in the third column display the RMSE values within the real range of that behavioral element.

The comparative bar plots shown in Figures 9-15, as well as the values for the RMSE and the SI given in Table 2, imply that the personalization module performs satisfactorily (i.e., an RMSE $\leq$ 14.5% and an SI $\leq$ 25%) for various behavioral elements of the SAR. The only exception is for the amount of the realistic versus nurturing comments where the RMSE and the SI go up to 23.6% and 52.6%, respectively. Based on the bar plots in Figure III-B1c, the larger values for the RMSE and SI are due to the results for participant P2, who shows the largest variation in the 3 different personality traits compared to others. Since the parameters are identified for all the participants at once, it is expected that for outliers the errors are significant. Moreover, since the RMSE (see (2)) is based on the squared values of the errors for the individual participants, the outliers are penalized relatively heavily.

The results are expected to improve when situating participants in real-life scenarios instead of asking them to imagine themselves in social interactions. Imagining that, particularly to distinguish their preference regarding receiving realistic or nurturing comments, may be challenging for multiple participants. In case with a larger group of participants the RMSE and SI do not improve, instead of re-identifying the parameters of the membership functions, the formulation of the fuzzy rules in Table 5 may be adjusted according to the preferences of the participants.

2) STAGE TWO: ASSESSING THE RL-BASED ADAPTATION MODULE

The setup and implementation of the experiments for assessing the RL-based adaptation module, including the state-of-mind estimation module, the fuzzy-logic-based reward module, and the RL update module (see Section II-C and Figure 3) are explained next.

TABLE 2. The RMSE and SI values for different behavioral elements of the SAR, computed via the identified personalization FLC module and as specified by the participants of the first online survey.

Behavioral element	RMSE for the scaled range 1 to 5	RMSE for the real range	(real range)	Percentages	SI
Amount of speech	0.42	0.42	1 to 5 sentences	10.6%	14.4%
Volume of speech	0.33	1.22	50 to 65 decibels	8.14%	2.13%
Number of gestures	0.74	0.74	1 to 5	18.6%	22.9%
Cautious/Energetic	0.37	9.33	0% to 100%	9.33%	12.9%
Challenging/Cooperative	0.51	12.80	0% to 100%	12.8%	25.0%
Realistic/Nurturing	0.95	23.60	0% to 100%	23.6%	52.6%
Proxemics	0.58	10.10	50 to 120 centimeters	14.5%	13.2%

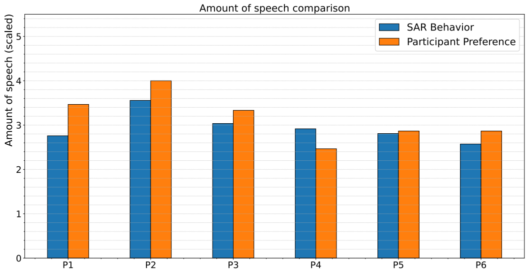


FIGURE 9. Comparing the output of the identified personalization module(shown via the orange bars) for the amount of speech of the SAR wheninteracting with the participants in the validation set with the preferences(shown via the blue bars) that have been specified by the participantsthemselves in the first online survey.

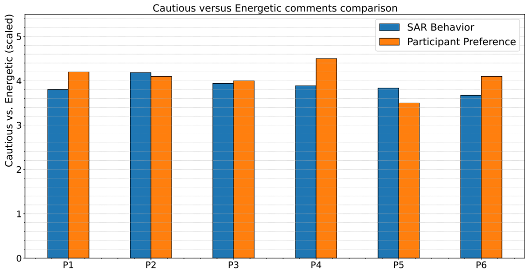


FIGURE 12. Comparing the output of the identified personalization module (shown via the orange bars)for energetic versus cautious comments by the SAR when interacting with the participants in the validation setwith the preferences (shown via the blue bars) that have been specified by the participants themselves in the first online survey.

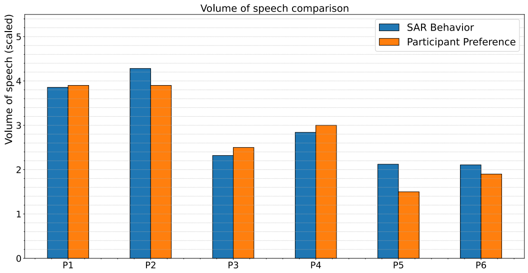


FIGURE 10. Comparing the output of the identified personalization module(shown via the orange bars) for the volume of speech of the SAR wheninteracting with the participants in the validation set with the preferences(shown via the blue bars) that have been specified by the participantsthemselves in the first online survey.

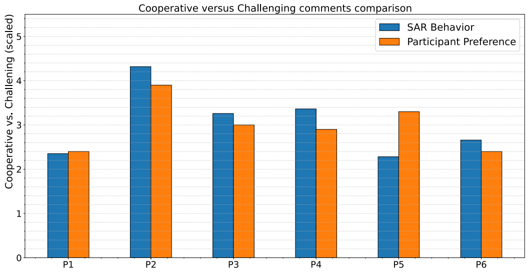


FIGURE 13. Comparing the output of the identified personalization module (shown via the orange bars)for cooperative versus challenging comments by the SAR when interacting with the participants in the validation setwith the preferences (shown via the blue bars) that have been specified by the participants themselves in the first online survey.

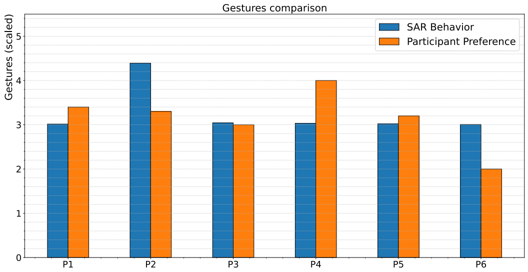


FIGURE 11. Comparing the output of the identified personalization module(shown via the orange bars) for the number of gestures of the SAR wheninteracting with the participants in the validation set with the preferences(shown via the blue bars) that have been specified by the participantsthemselves in the first online survey.

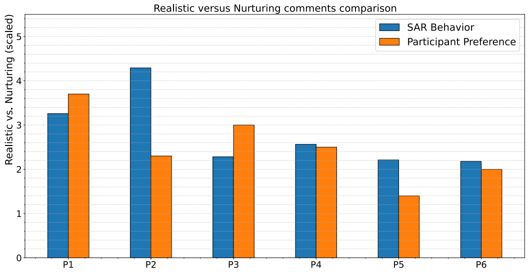


FIGURE 14. Comparing the output of the identified personalization module (shown via the orange bars)for realistic versus nurturing comments by the SAR when interacting with the participants in the validation setwith the preferences (shown via the blue bars) that have been specified by the participants themselves in the first online survey.

a: SETUP & SECOND ONLINE SURVEY

Appendix E gives detailed information on the rule bases of the FISs used to estimate the state-of-mind and the reward for the RL-based adaptation module of the SARs.

The state-of-mind estimation module uses a Mamdani FIS to determine the overall state-of-mind of a person as a fuzzy variable that adopts very good, good, neutral, bad, or very bad. The FIS receives the quantified scores (on a

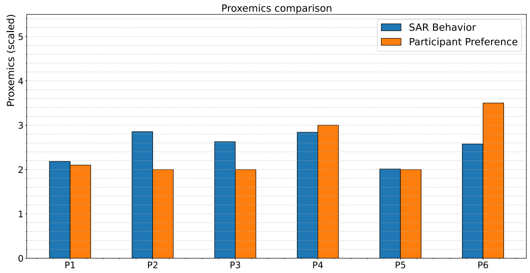


FIGURE 15. **Comparing the output of the identified personalization module (shown via the orange bars)for the proxemics of the SAR when interacting with the participants in the validation setwith the preferences (shown via the blue bars) that have been specified by the participants themselves in the first online survey.**

continuous range from 0 to 100) or the fuzzy evaluations (very high, high, low, very low) provided by a person about their emotions, including surprise, joy, sadness, fear, anger, disgust, trust, anticipation. The rules used by the FIS are given in Table 6, Appendix E: Some rules are fired via two input values regarding two different emotions, where these inputs are aggregated by an adjunct operator (logical **and**). Other rules are fired by only one input, i.e., the (fuzzy) value for one emotion or for two emotions that have been aggregated via a conjunct operator (logical **or**). The membership functions for describing the terms in the antecedent and consequent of the rules are illustrated in Figures 27 and 28 of Appendix E. If none of the fuzzy rules in Table 6 are fired, then the state-of-mind will be set to neutral.

The reward of the RL-based adaptation module is composed of an absolute (depending on the current state-of-mind of the person) and a relative (depending on the changes in the state-of-mind of the person) term. A set of rules that have been given in Table 7 of Appendix E is used to generate the value of the absolute component of the reward based on the state-of-mind of the person. For instance, if the updated state-of-mind of the person is either good or very good, a positive absolute reward of, respectively, 1.5 and 2.5 is given. This is equivalent to a simple, static Takagi-Sugeno-Kang FIS. For estimation of the relative component of the reward, a fuzzy Mamdani inference system is used with the rule base given in Table 8 of Appendix E. These rules suggest a fuzzy value for the relative component of the reward that adopts very negative, negative, neutral, positive, or very positive, based on the inputs, the previous and the current state-of-mind of the person given as a fuzzy value. Thus, the fuzzy outputs of the state-of-mind estimation module for the previous and current state-of-mind can directly be used as the input of this Mamdani FIS. The terms in the consequent of these rules are described by the membership functions that are illustrated in Figure 29 of Appendix E.

In the next section, we explain how simulated participants were generated to train and assess the RL-based adaptation module. In order to generate simulated participants that make realistic sense and that cover a wide range of personalities, we designed and conducted a second online survey, accessible via [64]. The participants of the survey were the same as for the first survey. They were asked to qualify the changes in their state-of-mind due to experiencing specific behaviors from a friend or partner in 160 different interactive scenarios. Per scenario, a hypothetical situation was pictured. The participant was asked to consider a given initial state-of-mind. A particular interaction was then described where the friend or partner would show particular behaviors (based on the interactive behavioral elements considered for the SAR). The participant was asked to qualify their state-of-mind variations according to very positively, positively, neutrally, negatively, very negatively. Next we explain how the results of the survey were used to simulate computer-based participants.

b: COMPUTER-BASED SIMULATED PARTICIPANTS

Since the RL-based adaptation module works based on the feedback (i.e., the emotions) from the participants (see Section II-C3 for details), the emotional response (i.e., the score for the eight basic emotions) to various behaviors of the SAR for the simulated participants should be modeled. Stochastic models were developed that, for any state-of-mind and change in behavioral parameters of the SAR, generate the probability that the resulting value of the emotions surprise, joy, sadness, fear, anger, disgust, trust, anticipation, for the participant would fall in [0, 25], (25, 50], (25, 75], (75, 100].

Table 9 in Appendix E shows an example of such a stochastic model, providing the emotional response of a simulated participant for the emotion surprise to the variations that the SAR makes in its volume of speech. The scores associated to the emotion per behavioral parameter that the SAR selects have been divided into four ranges, [0, 25], (25, 50], (50, 75], (75, 100]. The probabilities assigned to these ranges for each fuzzy state-of-mind (very bad, bad, neutral, good, very good) of the simulated participant represent the likelihood that a score within that range is provided in response to the behavioral parameter change (i.e., increase a lot, increase, decrease a lot, decrease). For instance, when the state-of-mind of this simulated participant is good, and the SAR selects "increase a lot" for its volume of speech, the probability that the score for the emotion surprise of the simulated participant is in the ranges [0, 25], (25, 50], (50, 75], (75, 100], is, respectively, 0%, 20%, 30%, 50%.

Using the results of the second online survey, we deduced the stochastic models that generate the same results as those provided by the participants of the online survey. We then expanded the set of the simulated participants by randomly fluctuating the parameters of these stochastic models, such that a wider range of 100 potential participants is simulated.

c: ASSESSMENT CRITERIA

The computer-based simulated participants were used to train and validate the RL-based adaptation module. We used the Q-learning algorithm (see Appendix C for details). The criteria of assessing the training and validation processes included, respectively, the convergence tolerance of the learning and the suitability of the behavior in response to

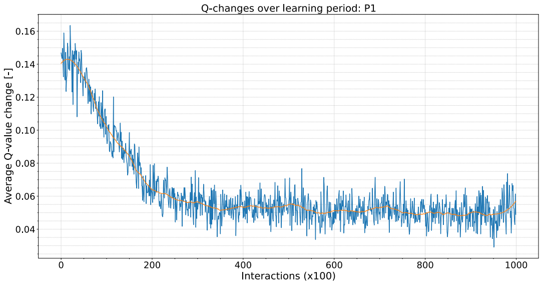


FIGURE 16. Evolution of the changes in the average Q-values during 10^5simulated interactions for a simulated participantthat was modeled according the real data from one of the participants of the second online survey.

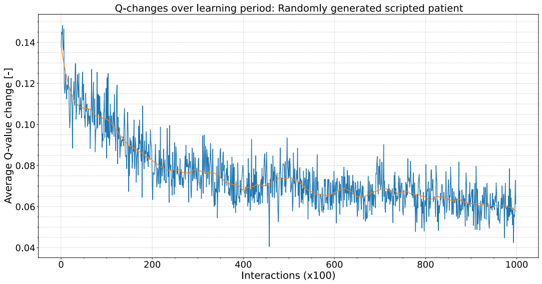


FIGURE 17. Evolution of the changes in the average Q-values during 10^5simulated interactionsfor a randomly generated simulated participant,which was modeled by small (bounded) fluctuations in the real-lifedata that has been gathered from one of the participants of the second online survey:In this case the model of the simulated participant was once more changedat interaction 5 × 10^4 in order to assess the robustness of the learning procedureto changes in the cognitive responses of a person.

different emotional states of a simulated participant by a SAR that used the trained RL-based adaptation module. For the convergence tolerance, the evolution of the changes in average Q-values during a maximum of 10^5 interactions was scanned. In case the absolute value of the evolution (i.e., the absolute difference of two consecutive average Q-values) remained below 5×10^{-5}, convergence was considered. For the performance of the trained RL-based adaptation module (i.e., suitability of the chosen behavior), action $a_{p,\sigma}$ proposed by the trained RL-based adaptation module to the SAR in interaction scenario σ with simulated participant p was compared with action $a_{p,\sigma}^{\text{proper}}$ that, according to the stochastic models of participant p, and using the fuzzy-logic-based reward module, resulted in the highest reward for participant p during scenario σ.

For example, for the participant modeled via Table 9 in Appendix E, when the current state-of-mind is very good, and the SAR selects 'increase a lot' for the volume of speech, the average expected value for surprise is $0.3(62.5) + 0.7(87.5)$. For all possible behavioral elements for the SAR, this average expected value is estimated and goes through the fuzzy-logic-based reward module to determine the corresponding reward. The action that corresponds to the highest reward is considered as the proper action $a_{p,\sigma}^{\text{proper}}$.

Since 7 behavioral elements for the SAR per state-of-mind of a participant exist, the total number of possible scenarios for all the 5 states-of-mind for a participant is 35. Therefore, the accuracy α_p^{RL} of the trained RL-based adaptation module for simulated participant p was calculated via:

$$\alpha_p^{\text{RL}} = \frac{\sum_{\sigma=1}^{35} \beta_{p,\sigma}^{\text{RL}}}{35}. \tag{4a}$$

where we have:

$$\beta_{p,\sigma}^{\text{RL}} = \begin{cases} 1 & a_{p,\sigma} = a_{p,\sigma}^{\text{proper}} \\ 0 & a_{p,\sigma} \neq a_{p,\sigma}^{\text{proper}} \end{cases} \tag{4b}$$

d: RESULTS AND DISCUSSION OF THE RESULTS

Figure 16 shows the evolution of the changes in the average Q-values during the training procedure of the RL-based adaptation module for a computer-based simulated participant, modeled based on the real data of one of the participants. The training procedure covered 10^5 interactions, where, to keep the figure readable, the average absolute change in the Q-values has been plotted per 100 interactions (see the blue curve in Figure 16). Moreover, the average overall trend of the evolution has been shown by the orange curve. The plot shows that the difference in the average Q-values for consecutive interactions decreases during the training interactions, and that the learning process eventually converges. More accurately, from interaction 3.64×10^4 onward, the slope of the orange curve remains under 5×10^{-5}. The performance accuracy of the RL-based adaptation module is 84.3%.

Figure 17 shows the evolution of the changes in the average Q-values during the learning process for one of the 100 randomly generated simulated participants. In order to assess the robustness of the learning process, the model was changed slightly at interaction 5×10^4. The plot shows that the difference in average Q-values decreases consistently, where around interaction 5×10^4, i.e., when the stochastic model of the simulated participant fluctuates, a slight increase is observed. Soon after, however, the average Q-values continue to decline consistently again. The performance accuracy of the RL-based adaptation module is 82.9%.

The RL-based adaptation module converged for all models that were based on the real participants. In average, the learning convergence occurred around the training interaction 3.4917×10^4, and the average performance accuracy of the trained RL-based adaptation module is 84.3%. For the 100 participants that were simulated by expanding the initial models of the real participants, the convergence occurred, in average, around training interaction 2.8472×10^4 before the models were changed (i.e., before interaction 5×10^4), and around training interaction 5.4060×10^4 after the models were changed. Thus, the second convergence needed less training interactions. The average performance accuracy of the trained RL-based adaptation module is 81.7%.

The slightly smaller average performance accuracy for the extended set of simulated participants may be due to the fact that (1) there is more variation in the stochastic models, and that (2) the models provide more dynamic and widespread cognitive preferences compared to models generated based on real data from the participants of the surveys.

3) STAGE THREE: COMPARING THE PROPOSED PARADIGM AND CONVENTIONAL RL-BASED STEERING METHODS FOR SARS

In this section, we compare the performance of the combined FLC-based personalization and RL-based adaptation modules for steering the social interactive behaviors of a SAR and a frequently-used approach that is based on conventional RL. The question is whether or not combining the personalization and adaptation procedures and integrating human-inspired and learning-based methods will allow the autonomous decisions of the SAR to converge faster to high-performing ones.

a: SETUP

From the computer-based simulated participants based on the real data from the participants of the surveys, 6 representative ones were selected to train and validate the steering methods. For the combined paradigm, the RL-based adaptation module was warm-started via the outputs of the FLC-based personalization module. For the conventional RL method the initialization was random.

b: ASSESSMENT CRITERIA

The number n^{converge} of the successful convergences from all the 35 scenarios possible for a participant, the convergence speed given as the number κ^{converge} of the interactions when convergence has occurred, and the performance accuracy α of the two steering approaches are compared for all combinations of the state-of-mind of the each simulated participant and the 7 behavioral elements of the SAR.

c: RESULTS AND DISCUSSION OF THE RESULTS

From the results, given in Table 3, the number n^{converge} of successful convergences for the combined paradigm was significantly larger than for the conventional RL-based method. For the combined paradigm, in only 1 case the learning did not converge. The number of the convergence interactions κ^{converge} for the combined paradigm was in average 13.6 % smaller than for the conventional RL-based method. This implies, as it was hypothesized, that the FLC-based personalization module properly initializes the RL-based adaptation module of the SAR, leading to a faster convergence.

When the combined paradigm was trained, it always had a larger performance accuracy than the conventional RL-based module. More precisely, the maximum difference in the performance accuracy was 7.9% (participant 1 in Table 3) and the average difference for all cases was 4.7%.

C. SUMMARIZED OVERVIEW OF THE RESULTS

The main findings based on the results of the case study are summarized below:

- Incorporating a fuzzy-logic-based personalization approach significantly enhances SAR decision making, yielding interactions that align with the needs and preferences of the participants:
  - Optimizing membership functions via RMSE minimization (based on data from 20 participants), resulted in a very satisfactory performance across most behavioral elements (RMSE $\leq$ 14.5%, SI $\leq$ 25%), except for “realistic vs. nurturing comments” (RMSE $=$ 23.6%, SI $=$ 52.6%) due to an outlier.
  - RMSE penalized outliers more heavily due to its squared-error formulation.
- Leveraging RL, warm-started by the fuzzy-logic-based personalization module, enables real-time adaptation to dynamics of state-of-mind:
  - Training the RL-based adaptation module with 10^5 interactions using computer-based simulated participants led to consistent convergence after, on average, 3.64×10^4 interactions.
  - The trained adaptation module achieved 84.3% accuracy for simulated participants.
  - Robustness testing with 100 randomly generated participants, showed initial learning convergence after 2.8472×10^4 interactions, with a significantly faster re-convergence (after, on average, 4.06×10^3 interactions) after model deviation.
  - The extended test group showed a slightly lower average accuracy of 81.7%, due to greater variations and extreme cases in simulated participants.
- The paradigm integrating fuzzy-logic-based personalization and RL-based adaptation modules significantly outperforms conventional RL methods used for steering the behavior of SARs. The key advantages of such integration, in particular, are:
  - Convergence rate, according to the results, was significantly higher (i.e., learning failed only once, while conventional RL faced multiple failures).
  - Learning was faster, requiring 13.6% fewer training interactions for convergence.
  - Accuracy was higher, outperforming standard RL in all cases, with a maximum accuracy gain of 7.9% and an average improvement of 4.7%.

These results confirm that the fuzzy-logic-based personalization module provides a strong initialization for the RL-based adaptation module, leading to faster and more effective learning in SAR decision making.

IV. CONCLUSION AND FUTURE RESEARCH

We proposed a novel paradigm to steer the social interactive behavior of socially assistive robots (SARs). This paradigm integrates two highly effective modeling and control methods, i.e., fuzzy logic control (FLC) and reinforcement learning (RL) in a novel way, allowing SARs to systematically personalize their behavior to various users and adapt their

TABLE 3. The results for the combined FLC-RL-based steering paradigm and the conventional RL-based steering method for social interactive behavior of SARs.

Participant	$n^{\text{convergence}}$ for combined paradigm	$n^{\text{convergence}}$ for conventional RL	$\kappa^{\text{convergence}}$ for combined paradigm	$\kappa^{\text{convergence}}$ for conventional RL	difference of convergence speeds [%]	α for combined paradigm	α for conventional RL
Participant 1	35 out of 35	35 out of 35	423	473	11.8	92.2	84.3
Participant 2	35 out of 35	34 out of 35	415	485	16.9	83.3	82.0
Participant 3	35 out of 35	34 out of 35	426	481	12.9	87.8	85.7
Participant 4	34 out of 35	34 out of 35	398	425	6.8	89.5	86.5
Participant 5	35 out of 35	35 out of 35	409	480	17.4	76.1	70.1
Participant 6	35 out of 35	35 out of 35	400	461	15.3	85.7	80.3
Average	99.5%	98.5%	412	468	13.6	85.8	81.5

interactions based on changes in their state-of-mind during the interactions. This steering system is generalizable, i.e., it has not been developed for a specific case study or interactive task and benefits from general theories that can be adopted for human-robot cognitive interactions (HRCIs) with varying goals. We ran extensive computer-based simulations for validation and comparison of the proposed paradigm with conventional RL, the most common method for steering SARs. The experiments were designed based on real-life data from extensive online surveys with 20 volunteer participants.

a: COMPARISON WITH EXISTING STUDIES

The learning procedure of the proposed RL-based adaptation module is warm-started by the outputs of the FLC-based personalization module. This significantly reduces the number of the exhausting, potentially harmful, trial-and-error-based interactions with humans, as confirmed by the results of our experiments (a reduction of almost 14%). Moreover, by including three FLC-based modules (for personalization, for estimation of the state-of-mind of the human, and for generating a reward according to the evolution of this state-of-mind), in addition to improved learning convergence, the performance accuracy of the proposed combined paradigm was improved for up to 8% compared to conventional RL.

b: IMPLICATIONS FOR REAL-LIFE APPLICATIONS

Following a human-centered engineering approach, we allowed the needs of humans to serve as the main objectives and constraints in developing the behavioral steering system of SARs, through a novel systematic integration of FLC and RL. The follow-up stage of this research will be to implement this in real-life experimental setups, based on:

- Designing a plan for long-term evaluation of the impacts of the SAR on physical and mental health, well-being, and cognitive enhancement of elderly users
- Ensuring compliance with ethical laws, norms, and standards for service/medical robots
- Developing clear and transparent consent procedures so that the participants understand the purposes and data collection policies of the experiments
- Interviewing the users, caregivers, and healthcare staff to identify specific needs of each group of users, and to tune the range of the actions of the robot accordingly
- Foreseeing and deploying essential measures to protect the privacy and confidentiality of the data that is collected during the experiments
- Scheduling introduction and supervised training sessions for the participants on how to work safely with the SAR or to seek help whenever needed
- Establishing support frameworks to address potential questions that may raise during the experiments
- Running real-time monitoring and detection methods to address unexpected risks in timely manners

c: LIMITATIONS AND FUTURE WORK

We proposed a novel, human-centered engineering approach for personalized and adaptable human-robot cognitive interactions between SARs and elderly users. While the development of such systematic methods is grounded in engineering principles, the users are the central factor in identifying, formulating, and incorporating the requirements, objectives, and constraints of the methods and evaluation metrics. Therefore, the primary proofs-of-concept provided in this paper combine requirements and inputs gathered via surveys from humans, with computer-based simulations. These will serve the next stage of the research where the proposed paradigm will be evaluated in real-world interactions with diverse populations of users. During these experiments, the methods, parameters, and rules of the FISs will be refined if needed, considering human-focused subjective and objective metrics.

A key technical challenge to be addressed for real-world implementation is the computational complexity arising from the large state-action space of the RL module and the fuzzy operations on extensive rule bases. To mitigate this, we propose several strategies.

First, as suggested in [65], incorporating expert-defined rules into the RL process through a modified actor-critic framework can significantly reduce the exploration of sub-optimal regions in the the state-action space, thereby improving sample efficiency and reducing computational demand.

Second, following modularity principles as described in [66], the architecture can be decomposed into independently manageable components, enabling localized optimiza-

TABLE 4. Mathematical notations used in the paper.

Notations in Section II	
Notation	Definition
n^{ext}	Total number of the external variables that impact the state-of-mind of a human in a given context
$V_{i,j}$	Fuzzy set that models the linguistic term that describes external variable i, which impacts the state-of-mind of the human, where j is an index that specifies which linguistic term from a set of N_i^{v} possible linguistic terms has been used
τ	Index referring to the personality trait of a human, with $\tau = 1, 2, 3$ corresponding to the personality trait, respectively, extroversion, agreeableness, and neuroticism
$P_{\tau,j}$	Fuzzy set that models the linguistic term that describes personality trait τ, where j is an index that specifies which linguistic term from a set of N_τ^{p} possible linguistic terms has been used
$F_{\tau,j}$	Fuzzy set that models the linguistic term that describes the fluctuations in the score of personality trait τ, where j is an index that specifies which linguistic term from a set of N_τ^{f} possible linguistic terms has been used
a_k	Crisp value for the change, with respect to the candidate action proposed by the FLC-based personalization module, regarding any of the 7 behavioral elements of the SAR
s_k	Crisp value for the state-of-mind of the human that is considered as the state variable of the RL module and is, thus, considered as a function of action a_k of the SAR
S_i	Fuzzy set that models the linguistic term that describes the state-of-mind of the human, where i specifies which linguistic term from a set of N^{s} possible linguistic terms has been used
$\tilde{S}_k$	Fuzzy variable that indicates which realization $\{S_1, \dots, S_{N^{\mathrm{s}}}\}$ corresponds to the state-of-mind of the human at a given interaction step k, i.e., $\tilde{S}_k$ is the fuzzy counter-part of s_k
B_i	Fuzzy set that models the linguistic term that describes the nature of a social interactive behavior by the SAR, i.e., an interactive comment, a motivating comment, a realistic feedback, or a nurturing feedback, where i is an index that specifies which linguistic term from a set of N^{b} possible linguistic terms has been used
C_i	Fuzzy set that models the linguistic term that describes the parameters corresponding to the SAR cues, where i is an index that specifies which linguistic term from a set of N^{c} possible linguistic terms has been used
$r\left(s_k, a_k\right)$	Reward estimated by the RL module, when the state-of-mind of the human is s_k and the SAR takes action a_k, which results in state-of-mind $s_{k+1}(a_k)$ for the human
$r^{\mathrm{quality}}\left(s_{k+1}(a_k)\right)$	Partial reward based on the quality of the state-of-mind of the human that has been updated for interaction step k under the impact of action a_k of the SAR
$r^{\mathrm{intensity}}\left(s_k, s_{k+1}(a_k)\right)$	Partial reward based on the the intensity of the change in the state-of-mind of the human from interaction step k to interaction step $k+1$ under the impact of action a_k of the SAR
Q_i	Fuzzy set that models the linguistic term that describes the quality component of the reward, where i specifies which linguistic term from a set of N^{q} possible linguistic terms has been used
$\tilde{Q}_k$	Fuzzy variable that indicates which realization $\{Q_1, \dots, Q_{N^{\mathrm{q}}}\}$ corresponds to the state-of-mind of the human at a given interaction step k, i.e., $\tilde{Q}_k$ is the fuzzy counterpart of $r^{\mathrm{quality}}\left(s_{k+1}(a_k)\right)$
I_i	Fuzzy set that models the linguistic term that describes the intensity component of the reward, where i specifies which linguistic term from a set of N^{i} possible linguistic terms has been used
$\tilde{I}_k$	Fuzzy variable that indicates which realization $\{I_1, \dots, I_{N^{\mathrm{i}}}\}$ corresponds to the state-of-mind of the human at a given interaction step k, i.e., $\tilde{I}_k$ is the fuzzy counterpart of $r^{\mathrm{intensity}}\left(s_k, s_{k+1}(a_k)\right)$
$E_{i,j}$	Fuzzy set that models the linguistic term that describes emotion i of the human, where in general n^{emot} number of different emotions are considered, and j is an index that specifies which linguistic term from a set of N_i^{e} possible linguistic terms has been used
O_i	Fuzzy set that models the linguistic term that describes the old (i.e., one interaction step earlier) state-of-mind of the human, where i is an index that specifies which linguistic term from a set of N^{s} possible linguistic terms has been used
Notations in Section III	
Notation	Definition
$\mathrm{RMSE}(i)$	Root mean squared error, with i an index that specifies for which of the 7 behavioral elements of the SAR the root mean squared error is estimated
$\mathrm{SI}(i)$	Scatter index, with i an index that specifies for which of the 7 behavioral elements of the SAR the scatter index is estimated
N^{part}	Number of the participants in the experiments
$b_j^{\mathrm{part}}(i)$	Value for behavioral parameter i of the SAR that, according to the results of the online survey, is preferred to by participant j
$b_j^{\mathrm{FLC}}(i)$	Value for behavioral parameter i of the SAR that is computed by the FLC-based personalization module for interacting with participant j
$a_{p,\sigma}^{\mathrm{proper}}$	Action that corresponds to the highest reward achieved in scenario σ for virtual participant indexed p in stage two of the experiments
α_p^{RL}	Accuracy of the trained RL-based adaptation module for simulated participant p in stage two of the experiments

tion and parallel processing, which in turn lowers the overall computational load on the system.

Additionally, we propose employing fuzzy rule clustering and weighting techniques [7], which allow for filtering and prioritization of relevant rules, thereby avoiding the exhaustive evaluation of the entire rule base for each FIS.

To further enhance real-time feasibility, pre-training of fuzzy membership functions and offline tuning of RL parameters can be carried out prior to deployment.

Lastly, the use of lightweight approximators [67] within the RL module can maintain learning performance while significantly reducing computational overhead.

Collectively, these approaches improve the computational efficiency of the integrated decision making paradigm, making it more practical for real-time deployment in SAR applications.

APPENDIX A MATHEMATICAL NOTATIONS

This section represents a list of the mathematical notations that are used in the paper, together with their definitions.

APPENDIX B FUZZY LOGIC CONTROL SYSTEMS

Fuzzy logic control (FLC) is one of the best choices for cases where instead of mathematical models, heuristic human-inspired reasoning is available or is preferred.

Fuzzy rules are the core of FLC-based controllers, and are described as logical if-then statements (propositions) that include linguistic terms. These terms are mathematically represented by fuzzy sets, as opposed to crisp sets.

While in classical logic, an element either belongs to or does not belong to a crisp set, in fuzzy logic, elements may partially belong to fuzzy sets. Consequently, fuzzy sets handle ambiguous data, and are thus used to mathematically model linguistic terms, which humans use in their reasoning. For instance, for a given temperature range, e.g., [0°C, 45°C], the term ''warm'' is represented by a fuzzy set, to which each temperature in the given interval belongs with a membership degree in [0, 1].

The processes used in FLC are, in general, composed of the following stages:

- Fuzzification, i.e., transforming real-life crisp input values into fuzzy values, unless the inputs are already fuzzy
- Fuzzy inference, i.e., evaluating the fuzzy input set according to the fuzzy rule base, which includes all the fuzzy rules of the FLC-based controller, and performing an inference based on fuzzy operations on fuzzy sets to provide a fuzzy output
- Defuzzification, i.e., converting the fuzzy output into a crisp output that directly steers a real system

Figure 18 represents different elements of an FLC system. In case the input to the FLC system is fuzzy (e.g., it is a linguistic term that is provided directly via a human user) this input is directly injected into the fuzzy inference system. Otherwise, any crisp (non-fuzzy) input should first be fuzzified.

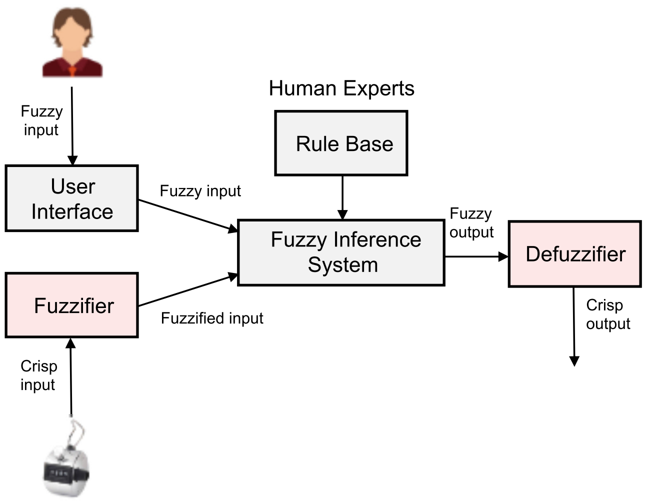


FIGURE 18. Schematic view of the two sub-modules of the fuzzy-logic-control-based decision making module.

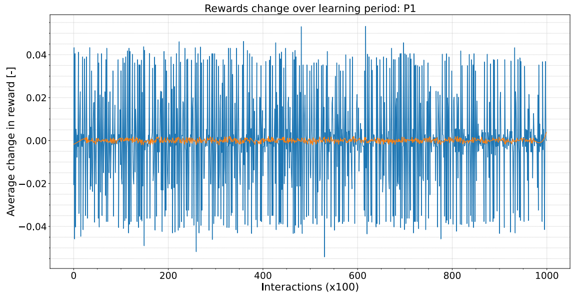


FIGURE 19. Evolution of the changes in the average reward values during 10^5simulated interactions for a simulated participantthat was modeled according the real data from one of the participants of the second online survey.

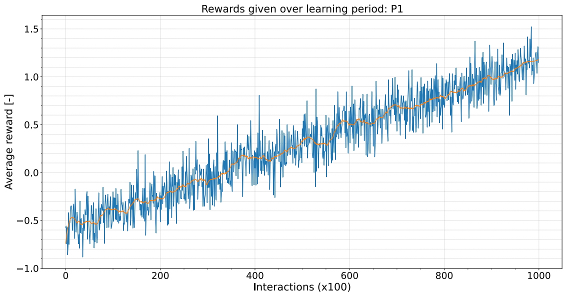


FIGURE 20. Evolution of the average reward values during 10^5simulated interactions for a simulated participantthat was modeled according the real data from one of the participants of the second online survey.

APPENDIX C Q-LEARNING AND DETAILED RESULTS FROM THE CASE STUDY

Q-learning uses a value function $Q : \mathbb{S} \times \mathbb{A} \rightarrow \mathbb{R}$ that quantifies the quality of any combination of the states and actions within, respectively, $\mathbb{S}$, i.e., the set of all states (in this paper this represents the states-of-mind for a human interacting with the SAR), and $\mathbb{A}$, i.e., the set of all actions (in this paper this represents the behavioral parameters of the SAR). Note that $\mathbb{R}$ is the set of the real numbers. This value function is numerically evaluated for various realizations of the state-action pair during the training phase. For the training iteration ℓ that corresponds to the interaction step k, the realized value of the value function $Q(\cdot, \cdot|\ell)$ for the state-action pair (s_k, a_k) is updated based on the realized reward (for more details regarding the reward in HRCIs, see Section II-C4 of the paper). The following relationship is used to update the value of this value function, for the same state-

TABLE 5. Fuzzy rule base for personalizing the behavioral elements of the SAR based on the personality trait scores.

	Antecedent 1	and/or	Antecedent 2		Consequent
If	*extroversion score is Low*	-	-	Then	*amount of speech is Low*
If	*extroversion score is Medium*	-	-	Then	*amount of speech is Medium*
If	*extroversion score is High*	-	-	Then	*amount of speech is High*
If	*extroversion score is Low*	-	-	Then	*volume of speech is Soft*
If	*extroversion score is Medium*	and	*neuroticism score is Low*	Then	*volume of speech is Medium*
If	*extroversion score is Medium*	and	*neuroticism score is Medium*	Then	*volume of speech is Medium*
If	*extroversion score is Medium*	and	*neuroticism score is High*	Then	*volume of speech is Soft*
If	*extroversion score is High*	and	*neuroticism score is Low*	Then	*volume of speech is Loud*
If	*extroversion score is High*	and	*neuroticism score is Medium*	Then	*volume of speech is Loud*
If	*extroversion score is High*	and	*neuroticism score is High*	Then	*volume of speech is Medium*
If	*extroversion score is Low*	and	*neuroticism score is Low*	Then	*number of gestures is Low*
If	*extroversion score is Low*	and	*neuroticism score is Medium*	Then	*number of gestures is Medium*
If	*extroversion score is Medium*	and	*neuroticism score is Low*	Then	*number of gestures is Medium*
If	*extroversion score is Medium*	and	*neuroticism score is Medium*	Then	*number of gestures is Medium*
If	*extroversion score is High*	or	*neuroticism score is High*	Then	*number of gestures is High*
If	*extroversion score is Low*	-	-	Then	*interactive comment is Cautious*
If	*extroversion score is Medium*	-	-	Then	*interactive comment is Neutral*
If	*extroversion score is High*	-	-	Then	*interactive comment is Energetic*
If	*agreeableness score is Low*	-	-	Then	*motivating comment is Challenging*
If	*agreeableness score is Medium*	-	-	Then	*motivating comment is Neutral*
If	*agreeableness score is High*	-	-	Then	*motivating comment is Cooperative*
If	*neuroticism score is Low*	-	-	Then	*feedback type is Realistic*
If	*neuroticism score is Medium*	-	-	Then	*feedback type is Neutral*
If	*neuroticism score is High*	-	-	Then	*feedback type is Nurturing*
If	*extroversion score is Low*	and	*agreeableness score is Low*	Then	*proxemics is High*
If	*extroversion score is Low*	and	*agreeableness score is Medium*	Then	*proxemics is High*
If	*extroversion score is Low*	and	*agreeableness score is High*	Then	*proxemics is Medium*
If	*extroversion score is Medium*	and	*agreeableness score is Low*	Then	*proxemics is High*
If	*extroversion score is Medium*	and	*agreeableness score is Medium*	Then	*proxemics is Medium*
If	*extroversion score is Medium*	and	*agreeableness score is High*	Then	*proxemics is Low*
If	*extroversion score is High*	and	*agreeableness score is Low*	Then	*proxemics is Medium*
If	*extroversion score is High*	and	*agreeableness score is Medium*	Then	*proxemics is Low*
If	*extroversion score is High*	and	*agreeableness score is High*	Then	*proxemics is Low*

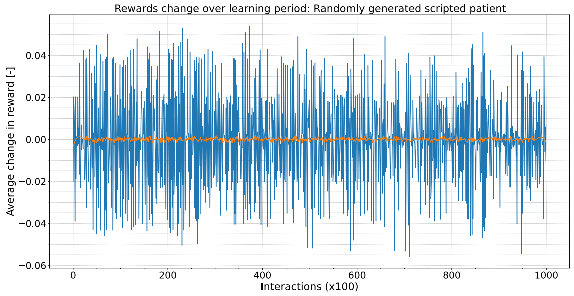


FIGURE 21. Evolution of the changes in the average reward values during 10^5 simulated interactions for one of the randomly simulated participants.

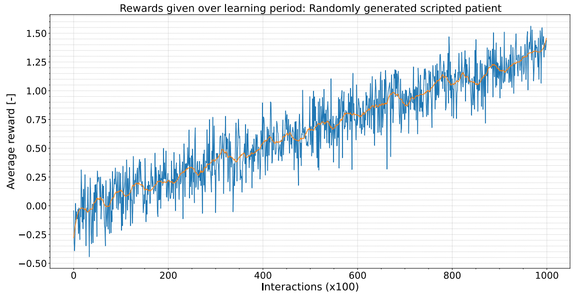


FIGURE 22. Evolution of the average reward values during 10^5 simulated interactions for one of the randomly simulated participants.

action pair, from iteration ℓ to iteration $\ell + 1$:

$$Q\left(s_k, a_k|\ell+1\right) = Q\left(s_k, a_k|\ell\right) + \lambda\Big(r\left(s_k, a_k\right) + \gamma \max_{\alpha\in\mathbb{A}} Q\left(s_{k+1}(a_k), \alpha|\ell\right) - Q\left(s_k, a_k|\ell\right)\Big) \tag{5}$$

where s_k and a_k are the values associated to the state variable and the action at interaction step k, $Q\left(s_k, a_k|\ell\right)$ is the realized value of the value function $Q(\cdot,\cdot)$ at training iteration ℓ for the state-action pair (s_k, a_k), λ is the learning rate of the algorithm, $r : \mathbb{S} \times \mathbb{A} \rightarrow \mathbb{R}$ is the reward function with $r\left(s_k, a_k\right)$ the reward that the learning system (which in this paper is the steering system of the social interactive behavior of the SAR) obtains during the training phase for selecting action a_k when the state is s_k, and γ is a discount factor. Moreover, the state that will be realized at the next interaction step, $k+1$, if action a_k is chosen at interaction step k is shown by $s_{k+1}(a_k)$, and $\max_{\alpha\in\mathbb{A}} Q\left(s_{k+1}(a_k), \alpha|\ell\right)$ evaluates the maximum expected value of function $Q(\cdot,\cdot)$ at the next interaction step, $k+1$, for all possible actions $\alpha \in \mathbb{A}$ and based on the data that is available at training iteration ℓ.

In an ϵ-greedy Q-learning approach, with a chance of ϵ a random action is chosen from $\mathbb{A}$ (this is called the exploration), and with a chance of $1-\epsilon$ an action is chosen that, based on the existing data, will result in the highest realized value for the value function $Q(\cdot,\cdot)$ at the upcoming interaction step (this is called

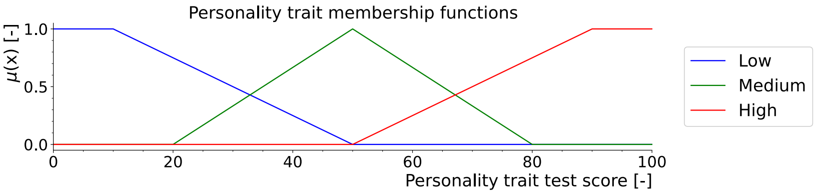


FIGURE 23. Membership functions describing the quality of the linguistic termsused in the antecedent of the fuzzy rules in Table 5.

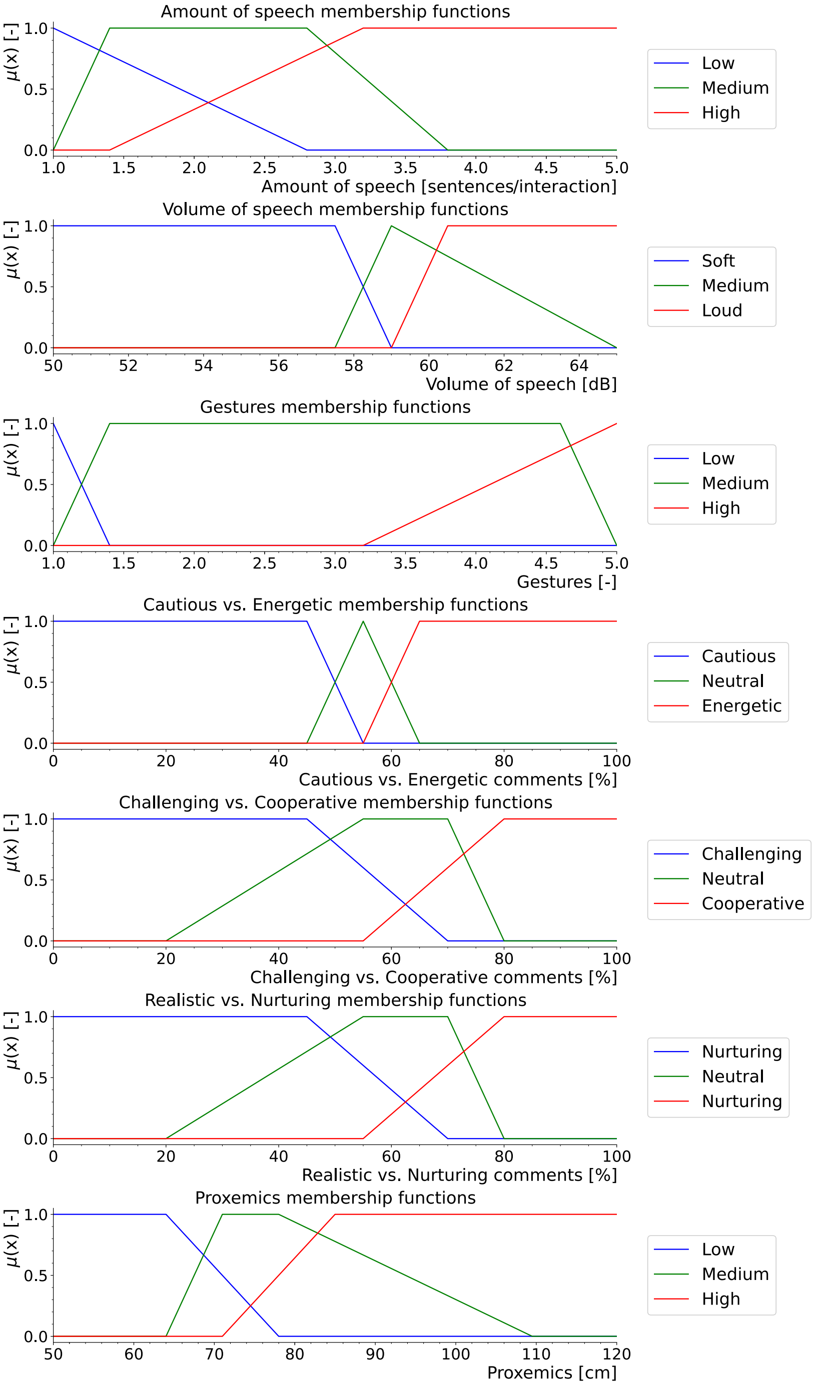
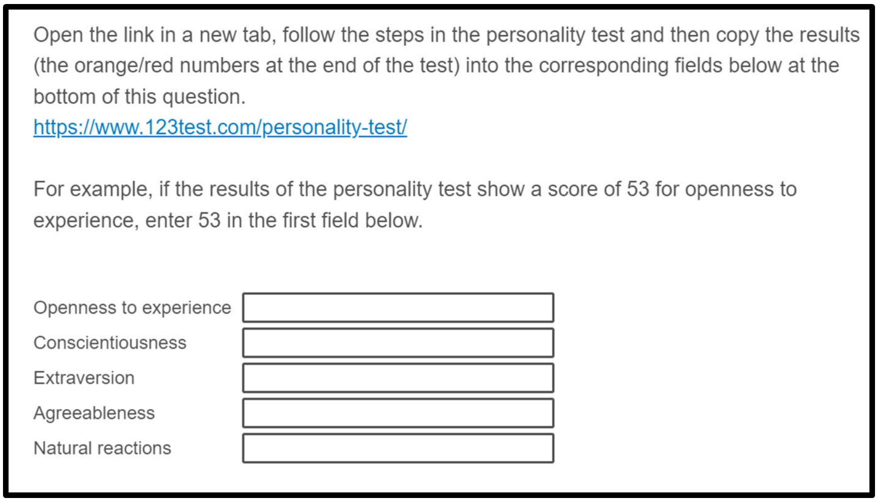


FIGURE 24. Membership functions describing the quality of the linguistic termsused in the consequent of the fuzzy rules in Table 5,after being tuned for the volunteer participants based on the results of the online survey.

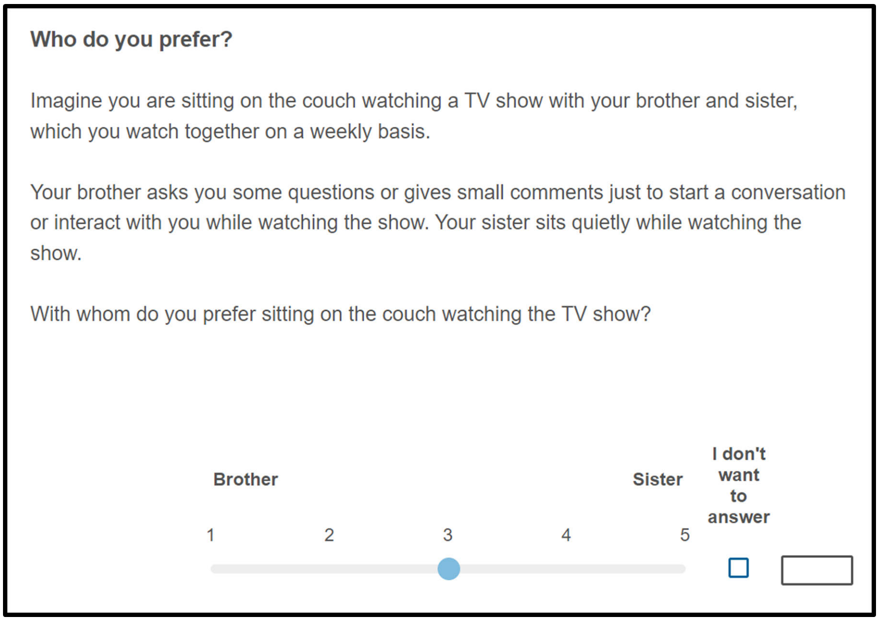


Open the link in a new tab, follow the steps in the personality test and then copy the results (the orange/red numbers at the end of the test) into the corresponding fields below at the bottom of this question.
https://www.123test.com/personality-test/

For example, if the results of the personality test show a score of 53 for openness to experience, enter 53 in the first field below.

Openness to experience □
Conscientiousness □
Extraversion □
Agreeableness □
Natural reactions □

FIGURE 25. A screenshot from the first online survey, where the participants were asked to take and report the scores of the Big Five personality test.



Who do you prefer?

Imagine you are sitting on the couch watching a TV show with your brother and sister, which you watch together on a weekly basis.

Your brother asks you some questions or gives small comments just to start a conversation or interact with you while watching the show. Your sister sits quietly while watching the show.

With whom do you prefer sitting on the couch watching the TV show?

FIGURE 26. A screenshot from an example question from the first online survey, where the participants were asked to specify their preferred behavior in given social interactions.

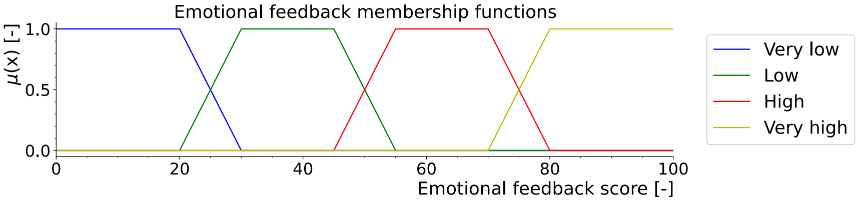


FIGURE 27. Membership functions describing the quality of the linguistic termsin the antecedent of the fuzzy rules in Table 6.

the exploitation). After sufficient number of training interactions, the RL algorithm is expected to converge, i.e., the values of function $Q(\cdot,\cdot)$ do not change significantly anymore.

Next, in Figures 19 and 21, we have presented the evolution of the reward values during the training procedure of the RL-based adaptation module for a computer-based simulated participant, modeled based on the real data of one of the

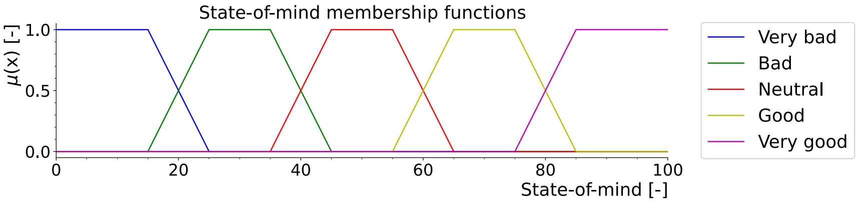


FIGURE 28. Membership functions describing the quality of the linguistic termsin the consequent of the fuzzy rules in Table 6.

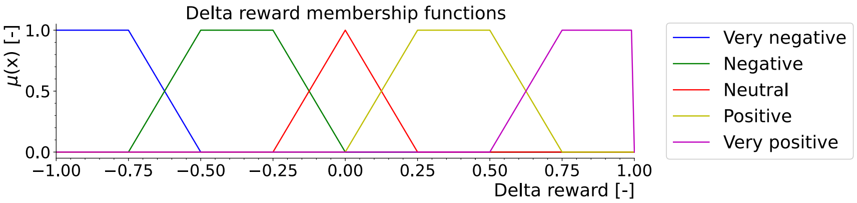


FIGURE 29. Membership functions describing the fuzzy terms in the consequent of the rulesgiven in Table 8.

TABLE 6. Fuzzy rule base for the fuzzy inference system used by the state-of-mind estimation module.

	Antecedent 1	and/or	Antecedent 2		Consequent
If	*joy is Very high*	-	-	then	*state-of-mind is Very good*
If	*joy is High*	and	*trust is Very high*	then	*state-of-mind is Very good*
If	*joy is High*	-	-	then	*state-of-mind is Good*
If	*trust is Very High*	-	-	then	*state-of-mind is Good*
If	*anticipation is High*	and	*disgust is Very low*	then	*state-of-mind is Good*
If	*surprise is High*	and	*trust is High*	then	*state-of-mind is Good*
If	*surprise is High*	and	*fear is Low*	then	*state-of-mind is Good*
If	*fear is Low*	and	*surprise is Very low*	then	*state-of-mind is Neutral*
If	*anger is Very low*	and	*anticipation is High*	then	*state-of-mind is Neutral*
If	*sadness is High*	and	*trust is Very low*	then	*state-of-mind is Bad*
If	*sadness is Very high*	-	-	then	*state-of-mind is Bad*
If	*fear is Very high*	or	*anticipation is Very low*	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	*fear is Very high*	and	*sadness is Very high*	then	*state-of-mind is Very bad*

participants as well as for one of the randomly generated participants. The changes in these average reward values almost follow a steady trend (see Figures 20 and 22). Analyzing these results next to those for the Q-values given in Figures 16 and 17 indicates that an optimal policy is being followed (as is evidenced by the stabilized Q-values), while exploration is still happening (due to an ϵ-greedy approach with non-zero ϵ). Thus, while due to this exploration the RL module still discovers actions with higher rewards, because the Q-values are already stable, these discoveries do not significantly affect the value function. Instead, they lead to a higher reward accumulation.

APPENDIX D
ADDITIONAL DETAILS FOR THE EXPERIMENTS WITH THE FLC-BASED PERSONALIZATION MODULE

This appendix includes the rule base and the fuzzy membership functions that were used in the experiments of the paper for the FLC-based personalization module.

TABLE 7. The rules that determine the absolute component of the reward based on the new state-of-mind of the person the SAR interacts with.

New state-of-mind	Absolute reward component
Very good	2.5
Good	1.5
Neutral	0
Bad	-1.5
Very bad	-2.5

In particular, Table 5 gives the fuzzy rule base that has been designed for personalizing the 7 behavioral elements of the SAR according to the scores for the three personality traits, extroversion, agreeableness, and neuroticism. Figures 23 and 24 illustrate the fuzzy membership functions that represent the linguistic terms that describe the, respectively, antecedent and consequent of the fuzzy rules in this table.

Finally, Figures 25 and 26 show screenshots of parts of the first online survey that was designed and conducted for the experiments of this paper.

TABLE 8. Fuzzy rule base for determining the relative component of the reward.

	Antecedent 1	and/or	Antecedent 2		Consequent
If	*old state-of-mind is Very bad*	and	*new state-of-mind is Very bad*	Then	*relative reward is Negative*
If	*old state-of-mind is Very bad*	and	*new state-of-mind is Bad*	Then	*relative reward is Neutral*
If	*old state-of-mind is Very bad*	and	*new state-of-mind is Neutral*	Then	*relative reward is Positive*
If	*old state-of-mind is Very bad*	and	*new state-of-mind is Good*	Then	*relative reward is Very positive*
If	*old state-of-mind is Very bad*	and	*new state-of-mind is Very good*	Then	*relative reward is Very positive*
If	*old state-of-mind is Bad*	and	*new state-of-mind is Very bad*	Then	*relative reward is Negative*
If	*old state-of-mind is Bad*	and	*new state-of-mind is Bad*	Then	*relative reward is Negative*
If	*old state-of-mind is Bad*	and	*new state-of-mind is Neutral*	Then	*relative reward is Neutral*
If	*old state-of-mind is Bad*	and	*new state-of-mind is Good*	Then	*relative reward is Positive*
If	*old state-of-mind is Bad*	and	*new state-of-mind is Very good*	Then	*relative reward is Very positive*
If	*old state-of-mind is Neutral*	and	*new state-of-mind is Very bad*	Then	*relative reward is Very negative*
If	*old state-of-mind is Neutral*	and	*new state-of-mind is Bad*	Then	*relative reward is Negative*
If	*old state-of-mind is Neutral*	and	*new state-of-mind is Neutral*	Then	*relative reward is Neutral*
If	*old state-of-mind is Neutral*	and	*new state-of-mind is Good*	Then	*relative reward is Positive*
If	*old state-of-mind is Neutral*	and	*new state-of-mind is Very good*	Then	*relative reward is Very positive*
If	*old state-of-mind is Good*	and	*new state-of-mind is Very bad*	Then	*relative reward is Very negative*
If	*old state-of-mind is Good*	and	*new state-of-mind is Bad*	Then	*relative reward is Negative*
If	*old state-of-mind is Good*	and	*new state-of-mind is Neutral*	Then	*relative reward is Neutral*
If	*old state-of-mind is Good*	and	*new state-of-mind is Good*	Then	*relative reward is Neutral*
If	*old state-of-mind is Good*	and	*new state-of-mind is Very good*	Then	*relative reward is Positive*
If	*old state-of-mind is Very good*	and	*new state-of-mind is Very bad*	Then	*relative reward is Very negative*
If	*old state-of-mind is Very good*	and	*new state-of-mind is Bad*	Then	*relative reward is Very negative*
If	*old state-of-mind is Very good*	and	*new state-of-mind is Neutral*	Then	*relative reward is Negative*
If	*old state-of-mind is Very good*	and	*new state-of-mind is Good*	Then	*relative reward is Neutral*
If	*old state-of-mind is Very good*	and	*new state-of-mind is Very good*	Then	*relative reward is Positive*

TABLE 9. Stochastic model for the emotional response of one simulated participant to the changes in the volume of the speech by the SAR.

Current state-of-mind	*The existing options for the parameter corresponding to the 'volume of speech' of the SAR*															
	Increase a lot				**Increase**				**Decrease**				**Decrease a lot**			
	The ranges of the score for the emotion surprise for the participant and the likelihood that the emotion score lies in this range after experiencing the behavioral change of the SAR															
	[0,25]	**(25,50]**	**(50,75]**	**(75,100]**	**[0-25]**	**(25,50]**	**(50,75]**	**(75,100]**	**[0,25]**	**(25,50]**	**(50,75]**	**(75,100]**	**[0,25]**	**(25,50]**	**(50,75]**	**(75,100]**
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%	40%	35%	10%	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%

APPENDIX E
ADDITIONAL DETAILS FOR THE EXPERIMENTS WITH THE RL-BASED ADAPTATION MODULE

This appendix includes the rule base and the fuzzy membership functions that were used in the experiments of the paper for estimating the state-of-mind and the rewards for the RL-based adaptation module.

In particular, Table 6 gives the filtered fuzzy rule base that has been designed for estimating the state-of-mind based on the feedback about the emotions of the person. Figures 27 and 28 illustrate the fuzzy membership functions that represent the linguistic terms that describe the, respectively, antecedent and consequent of the fuzzy rules in this table.

Table 7 shows the rules that are used to determine the absolute component of the reward, based on the new state-of-mind of the person that the SAR interacts with. Table 8 includes the fuzzy rule base that is used for determining the relative component of the reward, with Figure 29 showing the fuzzy membership functions that represent the linguistic terms that describe the consequent of the fuzzy rules in this table.

Finally, Table 9 shows an example stochastic model that has been developed for the emotional response of a simulated participant to the changes in the volume of the speech by the SAR.

APPENDIX F
ABBREVIATIONS

The following table includes the definition for the abbreviations that are used throughout the paper:

- **FLC:** Fuzzy logic control
- **FIS:** Fuzzy inference system
- **HRCI:** Human-robot cognitive interaction
- **RL:** Reinforcement learning
- **SAR:** Socially assistive robot
- **RMSE:** Root mean squared error

ACKNOWLEDGMENT

The authors would like to thank Dr. Davide Dell'Anna for his very helpful and valuable insights in designing the experiments.

REFERENCES

[1] D. Feil-Seifer and M. J. Mataric, "Socially assistive robotics," in *Proc. 9th Int. Conf. Rehabil. Robot.*, Jun. 2005, pp. 465–468.

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

[3] I. Papadopoulos, R. Lazzarino, S. Miah, T. Weaver, B. Thomas, and C. Koulouglioti, "A systematic review of the literature regarding socially assistive robots in pre-tertiary education," *Comput. Educ.*, vol. 155, Oct. 2020, Art. no. 103924.

[4] T. Fong, I. Nourbakhsh, and K. Dautenhahn, "A survey of socially interactive robots," *Robot. Auto. Syst.*, vol. 42, nos. 3–4, pp. 143–166, Mar. 2003.

[5] 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 *Proc. 1st IEEE/RAS-EMBS Int. Conf. Biomed. Robot. Biomechatronics*, Feb. 2006, pp. 483–488.

[6] T. Ascensão and A. Jamshidnejad, "Autonomous socially assistive drones performing personalized dance movement therapy: An adaptive fuzzy-logic-based control approach for interaction with humans," *IEEE Access*, vol. 10, pp. 15746–15770, 2022.

[7] D. Dell'Anna and A. Jamshidnejad, "Evolving fuzzy logic systems for creative personalized socially assistive robots," *Eng. Appl. Artif. Intell.*, vol. 114, Sep. 2022, Art. no. 105064.

[8] I. Leite, C. Martinho, and A. Paiva, "Social robots for long-term interaction: A survey," *Int. J. Social Robot.*, vol. 5, no. 2, pp. 291–308, Apr. 2013.

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

[10] C. Clabaugh and M. Matarić, "Robots for the people, by the people: Personalizing human-machine interaction," *Sci. Robot.*, vol. 3, no. 21, p. 7451, Aug. 2018.

[11] A. Tapus, M. Mataric, and B. Scassellati, "The grand challenges in socially assistive robotics," *IEEE Robot. Autom. Mag.*, vol. 14, no. 1, pp. 20–29, Jan. 2007.

[12] A. Tapus and M. J. Mataric, "Socially assistive robots: The link between personality, empathy, physiological signals, and task performance," in *Proc. AAAI Spring Symp., Emotion*, Mar. 2008, pp. 133–140.

[13] J. Heredia, E. Lopes-Silva, Y. Cardinale, J. Diaz-Amado, I. Dongo, W. Graterol, and A. Aguilera, "Adaptive multimodal emotion detection architecture for social robots," *IEEE Access*, vol. 10, pp. 20727–20744, 2022.

[14] C. D. Kidd and C. Breazeal, "Robots at home: Understanding long-term human–robot interaction," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Sep. 2008, pp. 3230–3235.

[15] D. Dell'Anna and A. Jamshidnejad, "SONAR: An adaptive control architecture for social norm aware robots," *Int. J. Social Robot.*, vol. 16, nos. 9–10, pp. 1969–2000, Oct. 2024.

[16] 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, Mar. 2019.

[17] K. Tsiakas, M. Abujelala, and F. Makedon, "Task engagement as personalization feedback for socially-assistive robots and cognitive training," *Technologies*, vol. 6, no. 2, p. 49, May 2018.

[18] M. L. M. Patrício and A. Jamshidnejad, "Dynamic mathematical models of theory of mind for socially assistive robots," *IEEE Access*, vol. 11, pp. 103956–103975, 2023.

[19] H.-L. Cao, P. G. Esteban, A. De Beir, R. Simut, G. Van de Perre, D. Lefeber, and B. Vanderborght, "A survey on behavior control architectures for social robots in healthcare interventions," *Int. J. Humanoid Robot.*, vol. 14, no. 4, Dec. 2017, Art. no. 1750021.

[20] M. Maroto-Gómez, F. Alonso-Martín, M. Malfaz, Á. Castro-González, J. C. Castillo, and M. Á. Salichs, "A systematic literature review of decision-making and control systems for autonomous and social robots," *Int. J. Social Robot.*, vol. 15, no. 5, pp. 745–789, May 2023.

[21] N. Akalin and A. Loutfi, "Reinforcement learning approaches in social robotics," *Sensors*, vol. 21, no. 4, p. 1292, Feb. 2021.

[22] A. H. Qureshi, Y. Nakamura, Y. Yoshikawa, and H. Ishiguro, "Robot gains social intelligence through multimodal deep reinforcement learning," in *Proc. IEEE-RAS 16th Int. Conf. Humanoid Robots (Humanoids)*, Nov. 2016, pp. 745–751.

[23] H. W. Park, I. Grover, S. Spaulding, L. Gomez, and C. Breazeal, "A model-free affective reinforcement learning approach to personalization of an autonomous social robot companion for early literacy education," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, Jul. 2019, pp. 687–694.

[24] P. Cunningham, M. Cord, and S. J. Delany, "Supervised learning," in *Machine Learning Techniques for Multimedia*. Springer, 2008.

[25] Z. Ghahramani, "Unsupervised learning," in *Advanced Lectures on Machine Learning*, vol. 3176. Springer, 2003.

[26] A. Tapus, C. Tapus, and M. J. Mataric, "The use of socially assistive robots in the design of intelligent cognitive therapies for people with dementia," in *Proc. IEEE Int. Conf. Rehabil. Robot.*, Jun. 2009, pp. 924–929.

[27] K. Bartl-Pokorny, M. Pykała, P. Uluer, D. E. Barkana, A. Baird, H. Kose, T. Zorcec, B. Robins, B. W. Schuller, and A. Landowska, "Robot-based intervention for children with autism spectrum disorder: A systematic literature review," *IEEE Access*, vol. 9, pp. 165433–165450, 2021.

[28] 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 Robot. AI*, vol. 6, pp. 1–18, Nov. 2019.

[29] Q. Hou and J. Dong, "Finite-time membership function-dependent H_∞ control for T-S fuzzy systems via a dynamic memory event-triggered mechanism," *IEEE Trans. Fuzzy Syst.*, vol. 31, no. 11, pp. 4075–4084, Nov. 2023.

[30] W. Bleidorn, P. L. Hill, M. D. Back, J. J. A. Denissen, M. Hennecke, C. J. Hopwood, M. Jokela, C. Kandler, R. E. Lucas, M. Luhmann, U. Orth, J. Wagner, C. Wrzus, J. Zimmermann, and B. Roberts, "The policy relevance of personality traits," *Amer. Psychologist*, pp. 1056–1067, Jun. 2019.

[31] S. Rossi, G. Santangelo, M. Staffa, S. Varrasi, D. Conti, and A. D. Nuovo, "Psychometric evaluation supported by a social robot: Personality factors and technology acceptance," in *Proc. 27th IEEE Int. Symp. Robot Human Interact. Commun. (RO-MAN)*, Aug. 2018, pp. 802–807.

[32] T.-H.-H. Dang and A. Tapus, "Stress game: The role of motivational robotic assistance in reducing User's task stress," *Int. J. Social Robot.*, vol. 7, no. 2, pp. 227–240, Apr. 2015.

[33] A. Vinciarelli and G. Mohammadi, "A survey of personality computing," *IEEE Trans. Affect. Comput.*, vol. 5, no. 3, pp. 273–291, Jul. 2014.

[34] L. Robert, R. Alahmad, C. Esterwood, S.-M. Kim, S. You, and Q. Zhang, "A review of personality in human–robot interactions," *Found. Trends Inf. Syst.*, vol. 4, no. 2, pp. 107–212, Jan. 2020.

[35] C. Esterwood and L. P. Robert, "Personality in healthcare human robot interaction (H-HRI): A literature review and brief critique," in *Proc. 8th Int. Conf. Human-Agent Interact.*, Nov. 2020, pp. 87–95.

[36] A. Tapus, C. Ţăpuş, and M. J. Matarić, "User—Robot personality matching and assistive robot behavior adaptation for post-stroke rehabilitation therapy," *Intell. Service Robot.*, vol. 1, no. 2, pp. 169–183, Apr. 2008.

[37] S. Jung, H. T. Lim, S. Kwak, and F. Biocca, "Personality and facial expressions in human–robot interaction," in *Proc. 7th ACM/IEEE Int. Conf. Hum.-Robot Interact. (HRI)*, Mar. 2012, pp. 161–162.

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

[39] M. Islam, M. Mazumder, D. Schwabe-Warf, Y. Stephan, A. R. Sutin, and A. Terracciano, "Personality changes with dementia from the informant perspective: New data and meta-analysis," *J. Amer. Med. Directors Assoc.*, vol. 20, no. 2, pp. 131–137, Nov. 2018.

[40] J. C. Anestis, T. R. Rodriguez, O. C. Preston, T. M. Harrop, R. C. Arnau, and J. A. Finn, "Personality assessment and psychotherapy preferences: Congruence between client personality and therapist personality preferences," *J. Personality Assessment*, vol. 103, no. 3, pp. 416–426, May 2021.

[41] 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 Res.*, vol. 30, no. 7, pp. 857–870, Oct. 2020.

[42] E. van Thiel. *Personality Test*. Accessed: Apr. 23, 2024. [Online]. Available: https://www.123test.com/personality-test/

[43] L. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning—I," *Inf. Sci.*, vol. 8, no. 3, pp. 199–249, Jan. 1975.

[44] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning—II," *Inf. Sci.*, vol. 8, no. 4, pp. 301–357, Jan. 1975.

[45] E. H. Mamdani, "Application of fuzzy algorithms for control of simple dynamic plant," *Proc. Inst. Electr. Engineers*, vol. 121, no. 12, pp. 1585–1588, 1974.

[46] L. A. Zadeh, "From computing with numbers to computing with words—From manipulation of measurements to manipulation of perceptions," *IEEE Trans. Circuits Syst.*, vol. 573, pp. 36–58, Jan. 2001.

[47] Y. Bai and D. Wang, "Fundamentals of fuzzy logic control-fuzzy sets, fuzzy rules and defuzzifications," in *Advanced Fuzzy Logic Technologies in Industrial Applications*. Cham, Switzerland: Springer, 2006, pp. 17–36.

[48] L. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, no. 3, pp. 338–353, Jan. 2003.

[49] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *IEEE Trans. Syst., Man, Cybern.*, vols. SMC–15, no. 1, pp. 116–132, Jan. 1985.

[50] 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, 2017.

[51] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, U.K.: Cambridge Univ. Press, 2018.

[52] J. Querengässer and S. Schindler, "Sad but true?–How induced emotional states differentially bias self-rated big five personality traits," *BMC Psychol.*, vol. 2, no. 1, p. 14, Dec. 2014.

[53] S. S. Izquierdo and L. R. Izquierdo, "Mamdani fuzzy systems for modelling and simulation: A critical assessment," *J. Artif. Societies Social Simul.*, vol. 21, no. 3, p. 2, Jul. 2018.

[54] N. Simmons-Mackie and D. Kovarsky, "Engagement in clinical interaction: An introduction," *Seminars Speech Lang.*, vol. 30, no. 1, p. 5, Feb. 2009.

[55] A. Lopez-Rincon, "Emotion recognition using facial expressions in children using the NAO robot," in *Proc. Int. Conf. Electron., Commun. Comput. (CONIELECOMP)*, Feb. 2019, pp. 146–153.

[56] R. Plutchik, *Emotions in the Practice of Psychotherapy: Clinical Implications of Affect Theories*. Washington, DC, USA: American Psychological Association, 2009.

[57] 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," *Sci. Rep.*, vol. 10, no. 1, p. 4367, Mar. 2020.

[58] B. Kikhia, T. G. Stavropoulos, S. Andreadis, N. Karvonen, I. Kompatsiaris, S. Sävenstedt, M. Pijl, and C. Melander, "Utilizing a wristband sensor to measure the stress level for people with dementia," *Sensors*, vol. 16, no. 12, p. 1989, Nov. 2016.

[59] G. J. Landry, J. R. Best, and T. Liu-Ambrose, "Measuring sleep quality in older adults: A comparison using subjective and objective methods," *Frontiers Aging Neurosci.*, vol. 7, p. 166, Sep. 2015.

[60] G. Dulac-Arnold, N. Levine, D. J. Mankowitz, J. Li, C. Paduraru, S. Gowal, and T. Hester, "Challenges of real-world reinforcement learning: Definitions, benchmarks and analysis," *Mach. Learn.*, vol. 110, no. 9, pp. 2419–2468, Sep. 2021.

[61] A. Alanazi, "Using machine learning for healthcare challenges and opportunities," *Informat. Med. Unlocked*, vol. 30, Mar. 2022, Art. no. 100924.

[62] A. B. Hostetter and A. L. Potthoff, "Effects of personality and social situation on representational gesture production," *Gesture*, vol. 12, no. 1, pp. 62–83, Sep. 2012.

[63] M. Munster. *Fuzzy-logic-based Decision Making Module Survey*. Accessed: Jul. 26, 2025. [Online]. Available: https://tudelft.fra1.qualtrics.com/jfe/form/SV_ehUK09qhgzXHkyi

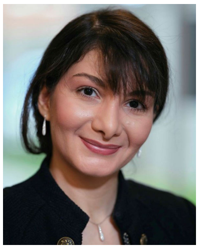
[64] M. Munster. *Reinforcement-learning-based Decision Making Module Survey*. Accessed: Jul. 6, 2025. [Online]. Available: https://tudelft.fra1.qualtrics.com/jfe/form/SV_09tvBCdiQBRiD1s

[65] L. D. Natale, B. Svetozarevic, P. Heer, and C. N. Jones, "Computationally efficient reinforcement learning: Targeted exploration leveraging simple rules," in *Proc. 62nd IEEE Conf. Decis. Control (CDC)*, Dec. 2023, pp. 2334–2339.

[66] M. Kawato and K. Samejima, "Efficient reinforcement learning: Computational theories, neuroscience and robotics," *Current Opinion Neurobiol.*, vol. 17, no. 2, pp. 205–212, Apr. 2007.

[67] W. Liu, Y. Li, and H. Tang, "LDQN: A lightweight deep reinforcement learning model," in *Proc. IEEE Smart World Congr. (SWC)*, Dec. 2024, pp. 1693–1698.

MARCEL MUNSTER received the B.Sc. degree in aerospace engineering and the M.Sc. degree in control and operations from Delft University of Technology, The Netherlands, in 2018 and 2023, respectively. Currently, he is a Data Science Trainee in the financial services industry, The Netherlands. His research interests include automation, social robotics, and artificial intelligence.



ANAHITA JAMSHIDNEJAD received the Ph.D. degree (cum laude) from Delft University of Technology (TU Delft), The Netherlands, in 2017. She is currently an Assistant Professor with TU Delft, leading the Mathematical Decision Making Group and directing the AI*MAN Laboratory. Her main research interests include systems theory for modeling human cognition, model-based predictive steering methods, fuzzy logic, integrated control paradigms, and applied to autonomous and social robots.