A Cognitive Architecture Incorporating Theory of Mind in Social Robots towards Their Personal Assistance at Home

O. Can Görür and Şahin Albayrak, *Member, IEEE*

***Abstract*—Recent studies show that robots are still far from being long-term companions in our daily lives. With an interdisciplinary approach, this position paper structures around coping with this problem and suggests guidelines on how to develop a cognitive architecture for social robots assuring their long-term personal assistance at home. Following the guidelines, we offer a conceptual cognitive architecture enabling assistant robots to autonomously create cognitive representations of cared-for individuals. Our proposed architecture places *Theory of Mind* approach in a *metacognitive process* first to empathize and learn with humans, then to guide robot's high-level decision-making accordingly. These decisions evaluate, regulate and control robot's *cognitive process* towards understanding, validating and caring for interacted humans and serving them in a personalized way. Hence, robots deploying this architecture will be trustworthy, flexible and generic to any human type and needs; in the end, they will establish a secure attachment with interacted humans. Finally, we present a use-case for our *novel cognitive architecture* to better visualize our conceptual work.**

I. Introduction

Recent engineering achievements on the robotic technologies have led to robots that are robust enough to be put in close interaction with humans in domestic environments. However, the fact that robots have only recently been deployed out of their lab environments leaves it controversial whether or not their capabilities will be satisfactory enough to be accepted in the long-term within proper human environment, i.e., the real world. Doubts are driven by the fact that the *Novelty Effect*, the first response to a new technology, also exists in HRI and the behaviors and the attitudes of humans towards robots change negatively as it wears off in long-term [1].

To support the longevity of robot usage in close interaction with humans, robots need to be social actors achieving social intelligence [2]. For this purpose, robots need to create cognitive representations of interacted humans by processing sensory inputs as well as learning from previous interactions [3]. In the case of social robots assisting at home, cognitive representations should comprise understanding personal needs and preferences of individuals living at home and adaptively responding to meet those needs [4]. However, recent studies showed that robots have deficient capacity to understand human's changing affective and motivational states (to empathize) and adapt to respond autonomously in long-term [5]. This suggests that robots are still far from creating such complex cognitive representations of interacted humans; as a result, they mostly depend on human commands or well-structured scenarios to initiate their interactions. The lack of adaptation to various changing human needs result in failure to keep users engaged over repeated and long-term interactions [1].

Adaptation of robots to dynamic human needs is an open-ended task that spans longer intervals of time. Implementation of autonomous robot architectures executing such open-ended tasks is non-trivial and suggested to be best achieved by memory-centered cognitive architectures [6]. The attention of these architectures has recently been drawn to social HRI, suggesting their applicability in the field along with further developments [7]. Moving from [7], it is our intuition that further strategies should be taken on cognitive architectures towards making them capable of adaptively creating cognitive representations of interacted humans, while making their structure simple enough to be applicable to robots deployed in real human environments.

In this position paper, we define a set of high-level guidelines in developing such a cognitive architecture specifically for social robots in their long-term personal assistance at home. The guidelines are extracted by linking the listed requirements for assistant social robots to the state-of-the-art cognitive architectures. Following these guidelines, we offer a conceptual cognitive architecture enabling assistant robots to autonomously create cognitive representations of cared-for individuals. These representations comprise human's personal needs and preferences. Our proposed architecture, in the end, places *Theory of Mind* approach [8] in a *metacognitive process* first to empathize and learn with humans, then to guide robot's high-level decision-making accordingly. These decisions evaluate, regulate and control robot's *cognitive process* based on twofold: *i)* exerting responsiveness behaviors, i.e., *understanding, validating* and *caring* for interacted humans as examined in [9]; *ii)* serving humans in a personalized way. Hence, robots deploying this architecture will be trustworthy, flexible and generic to any human type and needs; in the end, they will establish a secure attachment with interacted humans.

In Section II we present the related works, a brief literature survey on cognitive architectures and their requirements to implement social planning in assistive robots. Section III.A moves from this survey and extracts guidelines through developing a cognitive architecture for social robots towards their long-term personal assistance at home. Section III.B contains our novel cognitive architecture, a prototype following the guidelines as a proof of concept. In Section III.C we define a use-case scenario, as an exemplary application created using our cognitive architecture. Finally, the conclusion summarizes the article and our future plans.

O. C. Görür is with German Turkish Advanced Research Center for ICT, Berlin, Germany (e-mail: orhan-can.goeruer@gt-arc.com).

S. Albayrak is with DAI-Labor, Technische Universität Berlin, Germany (e-mail: sahin.albayrak@dai-labor.de).

II. Related Works

Many challenges are stated for social robots towards their assistance in daily life home environment. Tapus et al. [4] suggests that robots need to realize natural interaction with humans by means of empathizing with its user to *understand* personal needs and preferences and learning from the user to *adapt* its capabilities to the user's personality in providing customized interaction. Moreover, they added that the adaptation needs to cover both the short-term changes for individual differences and the long-term changes for allowing engagement over repeated and long-term interactions. That is, assistive robots towards their social interaction need to create adaptive cognitive representations of their cared-for humans. This level of cognitive process, achieving such open-ended and dynamic tasks in human interaction scenarios, is argued to be best achieved with human brain inspired cognitive architectures that emphasizes the role of memory in constantly learning the world and human [6], [10]. In these models, lifelong learning is achieved by separating the content from the architecture (e.g., control mechanisms, estimation models, learning processes etc.), where the architecture learns the content (e.g., needs of humans and how they are met) over the course of problem solving [6].

There are a few robotic cognitive architectures that have been developed as embodied approaches, integrating sensory-motor learning for recognizing and responding to outside stimuli. They use hybrid approaches, where subsymbolic level is used to recognize the outside stimuli, translate them into symbols and forward them to symbolic level (condition-action rules) to match with appropriate outputs. Being one of them, ACT-R/E [10] applies some HRI tasks including gaze following and Theory of Mind. The major drawback of ACT-R/E is that the learning is mostly limited to subsymbolic layer making it deficient in adaptively generating new symbolic rules (high-level knowledge) out of learnt stimuli. In other words, it is deficient in higher-level reasoning and problem solving at an abstract level. ADAPT [11] and SS-RICS [12] are other examples of embodied architectures, which are also limited to low-level tasks (e.g., navigation, SLAM).

It is stated that self-regulated learning is crucial in open-ended tasks, which is the case for assistive social robots in their long-term interactions, and it requires both *cognitive process* for constructing knowledge and *metacognitive process* for monitoring, controlling and regulating the learnt knowledge, i.e., high-level reasoning [13], [14]. CLARION cognitive architecture successfully integrates a metacognitive layer evaluating and creating new rules (knowledge) by combining reinforcement learning (Q-learning) with symbolic planning [14]. Each selected action decision is evaluated and regulated through Q-learning. MIDCA is another example of integrating metacognition process [15]. However, these two architectures are not embodied and have no application of HRI leaving their computational complexity towards their application to mobile robots and their abilities to exert social intelligence controversial.

Creating adaptive cognitive representations of humans requires a thorough understanding of human cognitive information. For this purpose, *Theory of Mind (ToM)* approach, the ability to understand human mental states like intention, receive a significant attention [8]. ACT-R/E successfully applies ToM approach in testing their embodied architecture [10]. Although the estimation results are remarkable, the application is limited to the given scenario where human and robot patrol an area. That is, robot runs cognitive simulations on hand-coded human beliefs only related to patrolling task. This is actually reflected in the aforementioned deficiency of ACT-R/E in problem solving at an abstract level, which causes its applications to be *ad hoc* (scenario specific). In a more recent study [16], a robot estimates human's belief on the joint actions of a shared plan to decipher work division between human and robot in a collaborative task. The authors implement a ToM manager in estimating human agent's mental state that is defined as human's belief on the state of his/her actions, plans and goals. Through this definition, human beliefs to be estimated are generalized towards their usage in broad range of assistant tasks of robots in human collaboration. Although this approach highly inspires our study, offered architecture is not focusing on personalized assistance, thus it is not adapting to human agents. Moreover, their belief estimation process does not take into account human emotional states and human reactions to robot moves; howbeit, in reality they have a significant impact on.

As Baxter recently states [7], cognitive architectures needs to integrate further strategies from social robotics towards their conformation to social human interactions. They need to enable planning of robot behaviors to be predictable, consistent and reliable for the interacted human. However, he adds that highly complex structure of available architectures practically limits this conformation when they are performed on real human interactions within real environments. On the other hand, studies on social robotics also highlight the missing integration of cognitive information about the human (intentions, beliefs, needs etc.) into the highest-level decision-making of robotic architectures towards their natural human interactions [17], [18]. As indicated in [9], these decisions of robots in general need to strive for exerting *responsiveness behaviors* listed as understanding, validating and caring for interacted humans in order to achieve a secure attachment between cared-for individual and the robot.

Moving from the outlined points, it is our belief that a *metacognitive process* integrating ToM approach is what is missing in conventional cognitive architecture towards their application to assistive social robots. *Metacognitive process* needs to utilize human mental states (to empathize) in controlling robot's cognitive tasks while always striving for exerting responsiveness behaviors and assisting cared-for humans with their goals. Moreover, the architectures should reduce their system complexity and should be compatible and adaptable to various assistant tasks and human types. To our knowledge, there is no such an embodied cognitive architecture integrating the listed requirements above.

III. General Framework

In this section, we present our guidelines in developing a cognitive architecture for social robots towards their personal assistance. Then, we introduce our conceptual cognitive architecture in detail. Finally, we give a use-case for the application of such an architecture on a robot in a real home environment assigned to assist an older person.

A. *Guidelines for Cognitive Process in Assistive Social Robots*

The motivation in listing the guidelines is to introduce new challenges when the fields of cognitive architectures and assistive social robots are engineered together. We use these guidelines in defining the components and their connection (input-output) strategies to compose our architecture. We do not claim with certainty that following the guidelines will output a robot supporting all aspects of real social assistance. Rather, these high-level guidelines are the minimum requirements that are extracted from the studies outlined in the related works. The guidelines are listed next.

1) Implement the cognitive architecture in a way that it will not be ad hoc, but be flexible and generic.

Social HRI scenarios put human in the center of the planning, which results in very dynamic contexts. Hence, robot architectures shall be built independent of the tasks and scenarios, then shall learn over the course of problem-solving.

2) Ensure adaptability of robot behaviors to various changing human behaviors.

Human behaviors differ from one day to another. This requires adapting to both short-term changes in user's mental state (e.g., tough day at work) and long-term personal habits and preferences. Therefore, architecture shall acquire and regulate knowledge lifelong, learning with human agents.

3) Recognize human mental states (Theory of Mind) to infer the true need and preferences of the user.

To empathize and adaptively respond to changing human needs, robot shall constantly recognize the mental states of its user. Mental states in assistant tasks shall include emotional state, intention and belief of the user.

4) Use human mental states in meta-level robot decision-making, regulating low-level cognitive process.

Metacognitive process targets for assessing the overall success of the cognitive system and regulating its process accordingly. In the case of assistive social robots, the assessment shall be based on changes in the recognized human mental states (see *Guideline 3*). Robot shall understand these changes and make new decisions, such as re-planning, which assures its adaptability (see *Guideline 2*).

5) Start with stereotyped plans, model preferences of each user and use this model in planning for personalization.

Every person has different social preferences and reacts differently to similar responses. Following Kirsch et al. [19] the planning shall: *i)* move from stereotyped behaviors suggested by social psychology studies; *ii)* adaptively learn user's social preferences and robot actions that creates appropriate responses on the user and update the *user model* with the learnt knowledge; *iii)* use the memory in planning behaviors through assisting the user in a personal way.

6) Make decisions that the user can anticipate and also approves towards establishing secure attachment.

Planning shall be goal-driven for user to anticipate. Particularly, the robot shall have internal goals where validating and caring for the user shall always be two of them throughout robot lifetime (*Responsiveness behaviors* of robots examined by Hoffman et al. [9] for secure attachment). Therefore, in assistive social robots, success criterion for the metacognitive evaluation of cognitive process (see *Guideline 4*) shall always be towards getting the approval of the user to robot actions. As a result, the user always trusts robot that it plans through assisting him in user's way, which in the end establishes a trustworthy relation and secure attachment between robot and the user.

7) Empathize and respond autonomously: decisional and functional autonomy.

Understanding, adapting and planning shall be connected in a closed-loop manner (input-output relation) constructing full decisional and functional autonomy. This supports robot to engage in intuitive and long-term interactions [5].

B. *Proposed Approach – CASOR Cognitive Architecture*

We contemplate the guidelines in designing a cognitive architecture that comprises of understanding, adapting and planning in a closed loop manner (*Targeting Guideline 7*). Our approach is based on our conceptual cognitive architecture given in Fig.1, named as CASOR acronym of "Cognitive Architecture for Assistive SOcial Robots". It describes the core processes of understanding human behavior and mental state from a robotic perspective (empathizing), learning from experiences and modelling the user and the world through social interactions. The goal is to create a task-invariant infrastructure that allows implementation of various skills (e.g., sensory/motor) and learns new contents, in this case user's personal preferences and needs. Therefore, our architecture offers a more generic and adaptive solution for assistive social robotics towards implementing various use-cases. (*Targets Guideline 1*)

CASOR is divided into two levels as *Cognitive Level* and *Meta-Cognitive Level*. The former comprises *Sensing*, *Actuating* and *Memory* components, where the latter has *Theory Of Mind (ToM)* and *Update Module* as illustrated in Fig.1. *Cognitive Level* is the low-level of the architecture and follows the plans set by meta-level. It holds sensory skills in recognizing outside stimuli and actuator skills for executing plans in forms of robot actions. Whereas *Meta-Cognitive Level* infers human mental states, evaluates success of the current plan through achieving human goals and then interrupts cognitive process and generates new plans, if necessary. The functionalities of each component and how we plan to achieve them are briefly explained below.

1) Sensing Component

This component consists of two levels of sensing as illustrated in Fig.1. Low- level sensing involves environment mapping, object recognition and recognition of human's presence, face, gaze, body joints (skeletal tracking), pose, basic facial expressions and basic speech commands. High-level sensing, on the other hand, utilizes the low-level cues in recognizing emotional states and actions of the user. In assistant tasks, emotional state is used as human feedback (see Fig.1) and is in the form of approval/disapproval detected using cues like smiling face or speech like "thank you" (*targeting Guideline 6)*. Before depicting the action

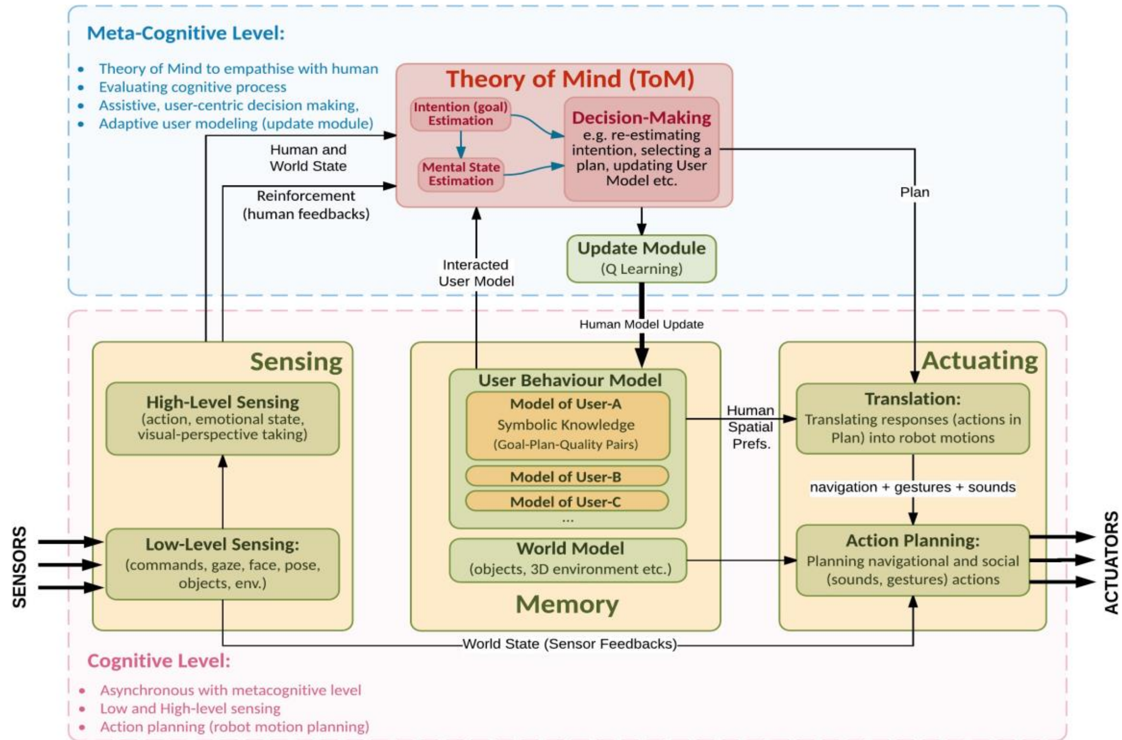

Figure 1 CASOR – Conceptual Robot Cognitive Architecture

recognition, we give an exemplary definition of an action to better describe the context in which it functions:

action = ⟨ *agent, type, object, location* ⟩

where for instance, act1 = ⟨User-A, grasping, book, onTable⟩. More contents can be added to the *action* definition, such as precondition and effect, yet the *type* is the most descriptive. *Sensing* recognizes action *type* by conventional action recognition methods using Hidden Markov Models. Estimation models utilize body joints as features for basic human actions like sitting, walking, grasping; whereas human gaze, head pose and objects (from low-level) are used to estimate what and where human is looking at, i.e., *visual-perspective taking*.

By defining actions in such a symbolic way, it is possible to introduce new *actions* using trained (known) action *types* in combination with other descriptors like trained *objects*. This is towards making the system flexible to various use-cases (*targeting Guideline 1*). *Sensing* component, then, is able to detect these *actions* in humans, which are to be used in goal (intention) estimation and plan descriptions.

2) Memory Component

CASOR is able to adapt to its own user and his/her changing preferences by modeling user's needs (goals) and preferences (plans) to meet these needs. *Memory* component is only for storing user models consist of symbolic knowledge in the form of productions (rules), i.e., *goal-plan-quality* pairs. *Quality* term is the quality of selecting a rule and it is calculated by user's approval (reward) in response to its previous selection (see *Update Module*). *Goals,* we also call as intentions, are personal goals of cared-for human that are also the goal of the robot to assist the human (for details, see *ToM Component*). *Plans* are action sequences in realizing these goals. An exemplary plan description is given as:

plan = ⟨ *agent list, action sequence, goal* ⟩

where *action sequence* consists of ordered *actions*, *agent list* holds the actor agent of each *action* in the sequence and *goal* stands for the goal the *plan* allows to achieve.

Memory is accessed by *Decision-Making* component of *ToM* to retrieve the specific user model for *rule* selection, and by *Update Module* to update *rule* qualities or to generate new ones (see Fig.1). *Memory* component starts with manually entered *plan(s)* for each *goal* of a user. In time, by being constantly updated through changing user preferences or states (see *Update Module*), *Memory* holds the personalized model of the user constructing a long-term memory. (*Targets Guideline 2, 5*)

3) Actuating Component

After *Decision-Making* component selects a *rule*, the set of *actions* of the rule's *plan* are translated here to navigational, sound and gestural actions for robot. For example, a symbolic representation of an *action* = ⟨ *robot, grasping, the glasses, onTable* ⟩ is translated into the actions of navigating to the location of the glasses and grasping it. *Action Planning* component does motor planning and actuates the motors to interact with the physical world.

4) Theory of Mind (ToM) Component

CASOR incorporates ToM approach to have robot take perspective of the user in estimating person's goal and preferred plan to reach the goal in the context. In *ToM* component, first step is to estimate intention and mental state of the user, then to incorporate them into meta-level decision-making. (*Targets Guideline 4*)

- *Intention Estimation:* Assistance is a joint action and its first step is to share a goal. Intentions in our architecture are called user goals, e.g., in daily life scenarios user habits like watching TV, reading book, etc. Robot shares these goals to assist human in achieving them. Goals can also be symbolically defined. An exemplary description is given as:

goal = 〈*agent, model, world state*〉

where *model* holds the trained model for estimation of the *goal*, *world state* defines the desired states goal wants to reach, such as desired human actions.

Features for training intention models are set of observed *actions* (recognized by *Sensing*) leading to an intention [18]. For example, looking at a book and grasping it may indicate the intention of reading. One HMM is used to train for each intention *model* with recorded *action* sequences (i.e., features) while a person is realizing this *goal*. Goal estimation then functions by checking the relevance of an *action* recognized (emission probability of the *action* in the *model*) and the transition of the *actions* (transition probabilities of the *model*). Estimated goals are used in mental state estimation and decision-making, which makes it a crucial part of CASOR. (*Targets Guideline 3*)

- *Mental State Estimation:* Inspired from [16], mental state in CASOR is the belief of human agent on the state of his/her current estimated *goal* (e.g., in progress, succeeded, aborted), *plan* (e.g., in progress) and *action* (e.g., aborted, help needed, done) in robot's view. For example, the state of *action-A* of the user is estimated as *help needed*, whereas *plan-A* and *goal-A* are *in progress*. Mental state estimation is a state transition model to be developed, for example, using Markov decision process, where states are the states of *goals/plans/actions* and transitions between states are based on: *i)* the current user *action*; *ii)* estimated intention (*goal*); *iii)* user reactions to robot moves in forms of approval /disapproval (*i* and *iii* from *Sensing*). We note that robot estimates *succeeded* as the belief of the user on user's goal only if robot constantly detects user's approval to its actions. Therefore, robot always strives for user's approval keeping his/her autonomy. Human mental state is the feedback of the system that helps robot evaluate the current success of its decisions. By constantly estimating it, robot realizes its mistakes and finally understands person's true preferences on the context to assist successfully. This gives robot the ability to adapt to user's momentary needs and preferences (short-term adaptation). (*Targets Guidelines 2,3,6*)

- *Decision-Making:* Meta decisions for CASOR are given as, but not limited to: re-estimating a *goal*, selecting a *plan*, triggering *Update Module*. Firstly, re-estimation of user *goal* (intention) is decided when robot estimates "aborted" for user's belief in the status of his/her estimated *goal*. Any unexpected action detection, which is an *action* that is not in the defined *plan* for a given *goal*, leads to the estimation of *goal aborted*. As a result, robot is able to detect its mistake in intention estimation of the human, which is already non-trivial even for us as humans. Secondly, plan selection is through symbolic planning. Algorithms are to be developed to match human *goal* with *rules* stored in the model of the user and to select one *rule* based on the *quality* values (see *Memory Component*). A new *rule* is selected each time the user's belief for a *plan* or a *goal* is estimated to be *aborted* or *succeeded*. Finally, *Update Module* is triggered for two reasons: to update the *quality* values of *rules* after each interaction, to generate new *plans*. (*Targets Guideline 4*)

5) Update Module

CASOR integrates associative reinforcement learning method to its symbolic planner in updating learnt knowledge with respect to constantly estimated human mental and emotional states. *Quality* value of a *rule* reflects the dynamic quality of selecting the *rule*. Rewards are the received user approvals (positive emotional states such as smiling or saying "thank you") in response to executed robot *actions* of a *plan*. The more approvals the robot receives (e.g., for each *action* under the given *plan*), the higher the *quality* value for selected *rule* will be. Thereby, selected *plans* always reflect user's approved way to achieve a *goal* for a trustworthy relationship. (*Targets Guideline 6*)

Moreover, *Update Module* is able to create new *plans* and so new *rules*. Any change in *action sequence* or *agent list* (see *Memory Component*) of *plan* descriptors constitutes a new *plan*. For example, *Update* is triggered to take over mopping upon the estimation of *help needed*. This changes the work-division (*agent list*) in a *plan* for house cleaning, thus creating a new *plan*. Associative learning approach matches the newly created *plans* with corresponding *goals*. By constantly fine tuning the user model, robot learns the most favored, secured and personalized *plans* to serve its user (long-term adaptation). (*Targets Guideline 2, 5*)

C. A Use-Case for the Proposed Cognitive Architecture

Our strategy is to keep the architecture task-invariant as much as possible. Developers have the ability to introduce new sensing (e.g., action *types*, *objects* and so *actions*) and actuation capabilities through implementing new applications as assistant tasks.

One possible use-case could be where a robot is assisting User-A, a conscious older man lives alone, in his daily activities. Robot's aim is to autonomously detect in which action User-A needs help and take over this action to assist in achieving his goal. User-A's family defines his habit of "reading" as a *goal* to the system with related *actions* given as (following the notation introduced in Section III.B):

- *AC1 (Action-1) =* 〈*User-A/Robot, Look for, Glasses*〉*,*
- *AC2 =* 〈*User-A/Robot, Navigate, toGlasses, PRE: AC1*〉*,*
- *AC3 =* 〈*User-A/Robot, Fetch, Glasses, PRE: AC2*〉*,*
- *AC4 =* 〈*User-A/Robot, Look for, Book*〉*,*
- *AC5 =* 〈*User-A/Robot, Navigate, toBook, PRE: AC4*〉*,*
- *AC6 =* 〈*User-A/Robot, Fetch, Book, PRE: AC5*〉*,*
- *AC7 =* 〈*User-A, Sit, onYellowCouch*〉

where the actor of an action, *agents*, may be the robot or User-A, and *PRE* is the precondition of action. Simple *goal* definition and manually entered initial *plan* are given as:

- *GO1 (Goal 1: Reading) =* 〈*User-A, model_GO1*〉*,*
- *PL0 (Initial plan) =* 〈*All Agents: User-A, AC4+AC5+ AC6+AC1+AC2+AC3+AC7, GO1*〉

where *PL0* is the initial plan for realizing the goal *GO1* and User-A is initially set as actor *agent* of all *ACs* under *PL0* with the given order. Ultimately, robot's goal is to detect when exactly (which *action* in the *plan*) User-A needs help to assist him with the *action* upon the user's consent.

Robot starts by estimating User-A's intention from the given *models*. In this iteration, User-A's goal is estimated as "reading" since *Sensing* recognized *AC4* in User-A (he is looking at the table where the book stands). According to the plan *PL0*, the only plan in *Memory* (*Model of User-A*), robot thinks that User-A must be planning to take the book (*AC5*). However, User-A's inaction for a while leads to mental state (MS) estimation of "User-A *needs help* in his action". It may be that User-A has *aborted GO1* but the ongoing look towards the table and MS transition probabilities (the belief that a person doesn't easily give up on his/her goal when it is *in progress*) makes robot reason that goal *GO1* is still *in progress*. Then, robot quickly asks User-A if it should bring the book. User-A smiles, which is a reward for the robot. Next action being "pick the glasses", the robot knows where User-A left his reading glasses and offers to take them. This time User-A says "don't!". He didn't want robot to pick the glasses because he was afraid that it could break them. Robot detects his disapproval, estimating "the *action* is *aborted"*. User-A picks the glasses leading robot reason that User-A doesn't *need help* in taking his glasses. After User-A sits back on his couch (*AC7* is recognized), robot concludes that initially it made the correct intention estimation of reading. Now that the work-division (active *agent list*) of *PL0* has changed, *Update Module* saves it as a new plan to *User-A Model* with the collected rewards. The plan is (*PL1*): Robot is responsible for fetching the book but not the glasses, i.e., actor *agent* of *AC4,5,6* is robot. This new plan is actually User-A's preferred plan for the goal of reading.

A week after, User-A has a terrible pain on his legs. Upon the estimation of "reading" goal, robot will not fetch the glasses as it has learnt that way. However, detected inaction of User-A will make MS estimation eventually result in "User-A *needs help*". As a result, robot asks and fetches the glasses creating another *plan* (*PL2*) in the user model. It could also be that User-A is not able to see without the glasses. This way, robot can adapt to such changing situations by empathizing with the user and reasoning when he needs help. If this new plan, where robot also fetches the glasses, is selected and succeeded more frequently on the upcoming days, robot will more likely select it assuming that it is User-A's favorite (having more reward).

IV. CONCLUSION

In this paper, we propose an approach for developing assistive social robots that bear a cognitive architecture to provide a long-term personalized assistance at home environment. We presented existing work in the field and pointed out the gaps to achieve our ultimate aim. The guidelines stated in the paper are attainable for a successful cognitive architecture required for personal long-term assistance of robots at home.

As future work, we are planning to implement our proposed architecture, CASOR. All the components except *Theory of Mind* are to be developed using state-of-the-art solutions, where *ToM* component and its integration to the cognitive process are to be our main contributions. Finally, CASOR is to be evaluated through a use-case scenario, similar to the one given, in successfully deciding when and how to assist older people in achieving their goals at home.

REFERENCES

[1] C. D. Kidd and C. Breazeal, "Robots at home: Understanding long-term human-robot interaction," in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS*, 2008, pp. 3230–3235.

[2] K. Dautenhahn, "Socially intelligent robots: dimensions of human-robot interaction.," *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, vol. 362, no. 1480, pp. 679–704, 2007.

[3] G. Bodenhausen and K. Hugenberg, "Attention, perception, and social cognition," in *Social cognition: The basis of human interaction*, 2009, pp. 1–22.

[4] A. Tapus, M. J. Mataric, and B. Scassellati, "The Grand Challenges in Socially Assistive Robotics," *IEEE Robotics and Automation Magazine*, vol. 14, no. 1, pp. 35–42, 2007.

[5] I. Leite, C. Martinho, and A. Paiva, "Social Robots for Long-Term Interaction: A Survey," *International Journal of Social Robotics*, vol. 5, no. 2, pp. 291–308, 2013.

[6] U. Kurup and C. Lebiere, "What can cognitive architectures do for robotics?," *Biologically Inspired Cognitive Architectures*, vol. 2, pp. 88–99, 2012.

[7] P. Baxter, "Memory-Centred Cognitive Architectures for Robots Interacting Socially with Humans," in *2nd Workshop on Cognitive Arhitectures for Social Human-Robot Interaction at HRI'16*, 2016.

[8] L. J. Byom and B. Mutlu, "Theory of mind: mechanisms, methods, and new directions.," *Frontiers in human neuroscience*, vol. 7, no. August, p. 413, 2013.

[9] G. Hoffman, G. E. Birnbaum, K. Vanunu, O. Sass, and H. T. Reis, "Robot responsiveness to human disclosure affects social impression and appeal," *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction - HRI '14*, pp. 1–8, 2014.

[10] G. Trafton, L. Hiatt, A. Harrison, F. Tamborello, S. Khemlani, and A. Schultz, "ACT-R/E: An Embodied Cognitive Architecture for Human-Robot Interaction," *Journal of Human-Robot Interaction*, vol. 2, no. 1, pp. 30–55, 2013.

[11] D. P. Benjamiin, D. Lyons, and D. Lonsdale, "ADAPT: A cognitive architeture for robotics," in *ICCM*, 2004, pp. 337–338.

[12] E. Avery, T. Kelley, and D. Davani, "Using cognitive architectures to improve robot control: Integrating production systems, semantic networks, and sub-symbolic processing," *15th Annual Conference on Behavioral Representation in Modeling and Simulation (BRIMS)*, no. Minsky 1974, 2006.

[13] K. W. Thiede, M. C. M. Anderson, and D. Therriault, "Accuracy of metacognitive monitoring affects learning of texts.," *Journal of Educational Psychology*, vol. 95, no. 1, pp. 66–73, 2003.

[14] R. Sun, X. Zhang, and R. Mathews, "Modeling meta-cognition in a cognitive architecture," *Cognitive Systems Research*, vol. 7, no. 4, pp. 327–338, 2006.

[15] M. Cox, T. Oates, and D. Perlis, "Toward an Integrated Metacognitive Infrastructure," *2011 AAAI Fall Symposium*, pp. 74–81, 2011.

[16] S. Devin and R. Alami, "An Implemented Theory of Mind to Improve Human-Robot Shared Plans Execution," in *11th ACM/IEEE International Conference on Human-Robot Interaction (HRI'16)*, 2016, pp. 319–326.

[17] R. Prada and A. Paiva, "Human-agent interaction: Challenges for bringing humans and agents together," in *Proc. 3rd International Workshop on Human-Agent Interaction Design and Models (HAIDM 2014)*, 2014, pp. 1–10.

[18] O. C. Görür and A. M. Erkmen, "Intention and Body-Mood Engineering via Proactive Robot Moves in HRI," in *Handbook of Research on Synthesizing Human Emotion in Intelligent Systems and Robotics*, J. Vallverdú, Ed. IGI Global, 2015, pp. 256–284.

[19] A. Kirsch, T. Kruse, E. A. Sisbot, R. Alami, M. Lawitzky, D. Brščić, S. Hirche, P. Basili, and S. Glasauer, "Plan-Based Control of Joint Human-Robot Activities," *KI - Künstliche Intelligenz*, pp. 1–9, 2010.