DR 2.2: Context and engagement-based strategy selection

Joost Broekens, Rifca Peters, Frank Kaptein, Bert Bierman

TU Delft, Delft, The Netherlands

‹joost.broekens@gmail.com›

Project, project Id:	EU H2020 PAL / PHC-643783
Project start date:	March 2015 (48 months)
Due date of deliverable:	28-feb-2017
Actual submission date:	28-feb-2017
Lead partner:	TU Delft
Revision:	final
Dissemination level:	PU

In this document we report on the work in work package 2 that relates to context, engagement, and explanation of robot actions. Context and engagement directly relate to this deliverable, explanation of robot action relates to work needed for task 2.1 to be able to implement multi-user explanation of the actions proposed by the system. We have focused on several fundamental issues in this year. First, can children perceive differences in robot educational style? This is important for engagement because human teachers use different styles and the ability to adapt educational style to an individual increases educational efficiency and effectiveness. Second, we have evaluated (with children and healthcare professionals) the PAL Control tool developed last year. PAL Control is our health care professional authoring tool. Again this is important for engagement and also for contextualized goal setting. Being able to set learning goals appropriately for children contributes to focusing on an achievable set of goals and contributes to structuring the learning process. Goal setting has been integrated with the quiz module of the PAL system. This enables the system to propose educational quizzes dependent on the context as defined by the learning goals. Third, we have shown in an experiment at a diabetes autumn camp that children and adults prefer robot action explanations differently. This is important fundamental knowledge enabling us (in the following years) to define how the agent should explain its strategy to the children, parents, and healthcare professionals.

Executive Summary

This report is the second deliverable report for WP2. In this document we report on the work in work package 2 that relates to context, engagement, and explanation of robot actions. Context and engagement directly relate to this deliverable, explanation of robot action relates to work needed for task 2.1 to be able to implement multi-user explanation of the actions proposed by the system. We have focused on several fundamental issues in this year.

First, we investigated the question if children can perceive differences in robot educational style. This is important for engagement because human teachers and caregivers use different styles to personalize their behavior towards children. This ability to adapt educational style to an individual increases educational efficiency and effectiveness. If we want the PAL system to be able to maximize learning efficiency and effectiveness, personalization of educational style is therefore important. However, for style to have an impact, we first need to know if style is perceived at all. In two experiments, one at two different primary schools, and one at the autumn camp in the Netherlands, we tested if the perception of style could be manipulated. These experiments showed that this is indeed possible, effects of some style elements were small but significant. We plan to extend these studies in the next year to first maximize the effect and then measure the effect of style on learning. A paper is under construction.

Second, we have evaluated (with children and healthcare professionals) the PAL Control tool developed last year, our health care professional authoring tool. Again this is important for engagement and also contextualized goal setting. Being able to set learning goals appropriately for children contributes to focusing on an achievable set of goals and contributes to structuring the learning process. The evaluation took place as a qualitative study to find out if the structure and interface was suitable for health care professionals to set goals for children. Important findings include that there is the need to also use the interface for assessment of the child’s knowledge and skill level and that this must be facilitated by the interface. A paper will be published on this topic (see Annex 1). The goals set by the child and care professional have been integrated as guiding force for the content of the quiz module of the PAL system. This enables the system to propose educational quizzes dependent on the context as defined by the learning goals.

Third, we have shown in an experiment at a diabetes autumn camp that children and adults prefer robot actions explained differently. Although both children and adults prefer goal-based explanations over belief-based explanations (e.g., I

propose to play a quiz with you because I want to play a game with you, vs., I propose to play a quiz with you because I know you like to play games). This is important fundamental knowledge enabling us (in the following years) to define how the agent should explain its strategy to the children, parents, and healthcare professionals. A paper is submitted on this topic (see Annex 2).

Role of the Robot Style, Goal setting and Explanation in the PAL system and prototype

The overall idea of the PAL project is to provide the child with a long-term and engaging experience. Robot style, strategic setting of goals, and suitable explanations of the system of why it proposes particular activities to the child are essential elements of providing an engaging experience over the long run. We explain these shortly.

Robot style is important because it facilitates engagement. Human educators change their educational style dependent on the child and learning task. For example, dominant communication is needed at a different moment than submissive communication. If a robot cannot change its communication style, it will be perceived – in the long run – as boring and non-personal and at the very least not adaptive towards the learning needs of the child. This could result in lower engagement or dropping out.

Setting appropriate learning goals is extremely important to give focus and a sense of progress to the child. This needs to be done collaboratively, so that the child feels he/she is problem owner. As such, a goal setting interface, and a method to then adapt the PAL's system behavior based on these goals, facilitate the learning process.

Explaining why something is done is important for understanding why something is (or is not) important, and for attributing priority. Humans do this naturally. Children and their parents could benefit if smart agents, such as the PAL agent, would be able to explain its actions. For example, if the PAL agent proposes to play a quiz about insulin, then it helps the child to know that this is proposed because the agent thinks the child likes to play quizzes. However, parents might want to know that the reason for playing this game is to learn the child more about insulin. In other words, different users might need different explanations, generated by the same PAL agent, i.e., the PAL agent needs to be able to generate multi-user explanations.

Tasks, objectives, results

1.1 Planned work

This report was planned to report on the work done in task 2.2. This work is about multimodal role selection (e.g., PAL selects to be a peer while friends are present but a tutor when the child is alone at night), and, about social norms and values about PAL behavior in different roles and contexts that could help select appropriate roles in an ethical way. Further, work was planned on the evaluation of the PAL control goal setting. Finally work was planned in task 2.3 on engagement detection and integration of engagement and context in the behavior of the PAL agent.

1.2 Actual work performed

Actual work this year is shifted slightly, though all work is directly relevant to the overall goal of the PAL project and specified in WP2 in the proposal. We report on the evaluation of PAL control (goal setting) and the use of the goals as context for the PAL strategy (as planned). We report on robot behavioural style perception by children, as this is important for long-term engagement (and strongly related to work in Task 5.2). Finally, we report on multi-user explanation (planned in task 2.1 for year 3). We realise that this is a change of plans (mainly in what happened when). However, the achievements this year have shown to be directly relevant for the PAL project (either as knowledge for future work or as activities during camps), and, as mentioned, fit the overall agenda of work package to, i.e., managing the strategic elements of the PAL system, and, taking engagement context and multi-user aspects into account.

PAL Control: evaluation of the HPC authoring tool and use of goals in quizzes

The authoring tool, including the tree-based interface for goal setting, progress monitoring and attainment registration, has been evaluated in the Netherlands and Italy. In total 7 HCPs (6 Dutch nurses, 1 Italian doctor), and 35 children (aged 7-12 M/F) and their parents participated in the study. In face-to-face consultation the HCP, together with a child and his/her parent(s), set personal learning goals to be pursued by the child using the MyPAL application (see D5.1 section 3.1.1) at home for the next three weeks. Eleven of these consultations were observed and audio-recorded. Additionally, in the Netherlands, training sessions were carried out with each nurse prior to the first consultation using a think-out-loud protocol. Semi-structured interviews were carried out posterior with all professionals.

All consultations included assessment of the child's current abilities, goal suggestions by the professional, conformation by the child and/or parent, and registration of goal state (active, inactive or attained) in the authoring tool. These steps were repeated for individual goals by a top-down walk through of the goal-tree. Although the selection mechanism was easily understood and goal setting was done effectively, the underlying goal structure was not clear. Nurses showed diffficulties understanding the topic clusters and achievement concept. Several visual cues (e.g., icons, shapes) have been discussed to point out the structure. These suggestions need to be implemented and evaluated, see also the work of described in D1.3 section 3.2.5. Further, rather than selecting a goal of special relevance to the child, after assessment of current abilities goals are set for each topic on the level not yet acquired, resulting in a broad selection of goals on various topics. This (unexpected) assessment functionality provided the nurses with valuable insight of the child's abilities. Further investigation on the feasibility of broad versus focused goal setting is required.

Nurses were enthusiastic about the systems capabilities to inform, both HCPs and children, about progress calculated based on quiz performance. However, the specification of tasks contributing to goal progression is limited and needs expansion. Further, lack of feedback on whether a goal was registered attained manually or by the system hindered reflection on progress. Aside, clear visualization of personal objectives and progress in the child interface (i.e., MyPAL) may increase the effect of goal setting on motivation and learning gain.

On the content level, goal descriptions were not specific enough about the precise ability that will be acquired when the goal is reached. For example, knowledge can be obtained either on the level of remembering a fact, understanding the implication, or being able to apply knowledge in new situations [4]. This is interesting because in diabetes self-management training a gap often arises between having factual knowledge and application in daily life. Moreover, attitudinal objectives are important to the domain and should be included. The diabetes education ontology structure is updated accordingly, content revision (objective instances/individuals) is planned this year.

For more details on the evaluation method and results, we refer to the paper in Annex 1.

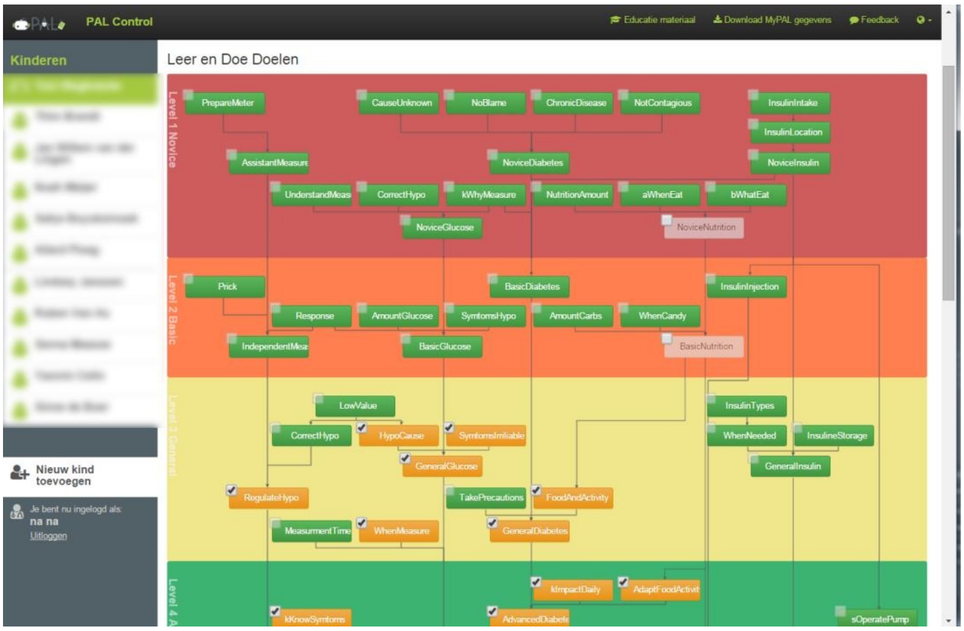


Figure 1. The authoring tool (PAL Control), displaying the diabetes learning goals (i.e., knowledge and skills to attain to progress towards self-management) and achievements with current progress for a 10 year old boy. Attained objectives are green, yellow ones are active. Coloured horizontal bars depict difficulty levels. Topics are arranged vertically.

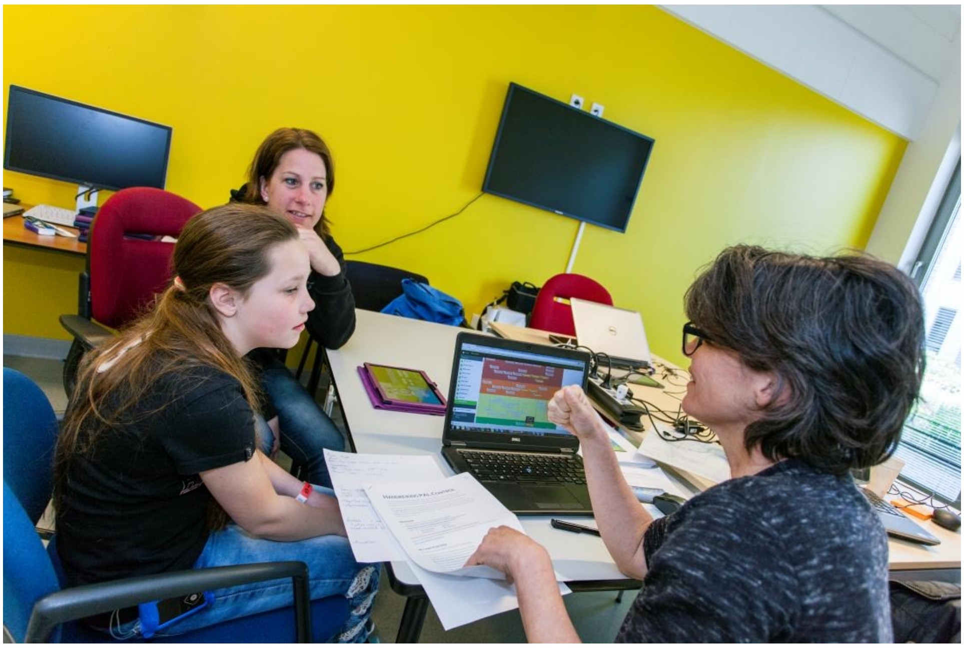


Figure 2. Healthcare professional assesses, together with a child and parent, the child's current knowledge and skills, registers goal attainment for already acquired abilities, and sets personal learning goals to be pursued in the next weeks.

The diabetes self-management learning goals have specific tasks to practice the ability associated with that goal. For example, to learn locations for insulin injection the child can play a quiz on this matter. These objectives (both goals and tasks) and relations between them are formalized in the diabetes education (goal) ontology. This enables personalization of content and automatic calculation and registration of progress. An early example is provided in the paper in Annex 5.

Goals are used in the functioning of quizzes done with children in the following way. First, the quiz is initialized with a child *id*. Then the quiz retrieves from the database which questions have already been posed (these will be excluded). The DialogueManager (see WP4) will propose the option to "ask question with goal X", with X being part of the set of active goals. When this action is selected, the quiz module will select a random question that belongs to goal X. And the child can answer the quiz. In this way, only questions that are appropriate for the current learning goals are asked.

Style-based robot behaviour

We are not only concerned with *what* is appropriate robot behaviour, but also *how* this should be communicated. In human-human interaction, educators heavily relay on their ability to identify and respond to social signals. Deliberate or based on implicit competencies, human educators adapt their style of interaction to the learner. For example a teacher can take up different teaching styles e.g. expert or facilitator [2]. These styles are related to the expression of warmth: the expert style is related to low warmth, opposed to high warmth for facilitator style. More open gestures are perceived as high warmth, forward gestures are perceived as low warmth [3].

To further our study on *how* actions should be communicated in child-robot interactions we conducted a series of perception studies, validating our model of warmth and competence expression by display of non-verbal robot behaviour. Following the line of work from Nguyen et al. [3] we developed a model of non-verbal behaviour cues (i.e., posture, gesture and paralinguistic parameters) to express high/low warmth or competence, and created a library of NAO behaviours fitting this model. High warmth is related to soft voice, low pitch, chin up, bodily orientation towards the audience, and open, spreading gestures both semantic and syntactic. High competence is related to stable posture, and frequent gestures. The model is applied to a robot giving a short informal, interactive lecture, resulting in four configurations (warmth x competence) of robot behaviours (Figure 3).

The model was evaluated with 101 children at 2 primary schools and 21 children at diabetes camp in the Netherlands. Children's perceptions of the robot's level of warmth and competence were measured by 20 adjectives rated on a 3-point Likert scale. The warmth and competence scores are calculated from a weighted loading for each adjective on that factor based on expert ratings. We developed an instrument requiring physical activity (placing stickers) to increase trustworthiness of responses and avoid the primacy effect (i.e., consistently selecting either the low or high extreme) [5].

A preliminary analysis indicated that subtle changes in robot behaviour or context (location, user group, content) influenced how children perceive the robot. Children did perceive the robot displaying competence-related behaviours (stable posture and frequent gestures) as more competent than a low-competence robot. A robot displaying warmth-related behaviours (low voice and pitch, head tilt, and spreading gestures) was perceived as warmer than a low-warmth robot only if the robot was high-competent. Additionally, independent of the robot behaviour configuration, warmth scores were higher at schools than diabetes camp. This work is planned for submission early 2017.

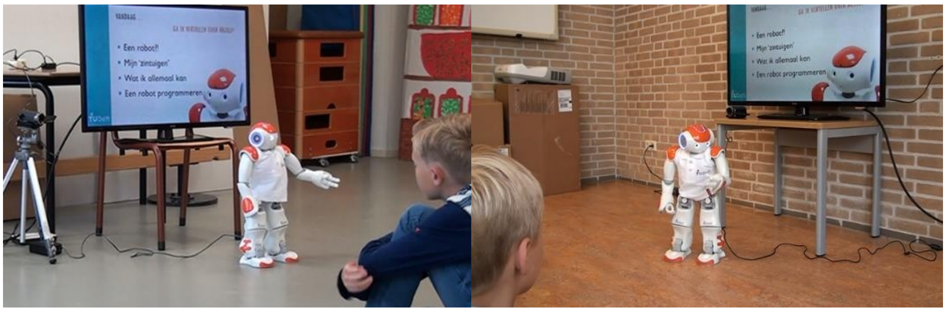


Figure 3. Stills of the behaviour accompanying a statement in two of the four robot style configurations. Left: high warmth and high competence. Right: low warmth and low competence

Multi-user explanations

The PAL-system needs to facilitate human-agent collaboration by explaining its users why the Pal-agent behaved in particular ways. This is part of task 2.1, and part of WP2. The PAL-agent autonomously interacts with the children for prolonged periods of time. It helps them to cope with medical health issues. Explainable AI (XAI) facilitates increases a user's trust and understanding in the system they are working with [12, 13]. Furthermore, it facilitates shared patient-caregiver responsibility for the child's diabetes regimen at preadolescence.

When humans amongst each other explain their behavior, then we do this by means of folk psychology, i.e., we communicate the beliefs and goals that caused us to perform certain actions [9]. Previous work on XAI provides guidelines on generating folk psychology based explanations for behavior of intelligent systems [7]. However, this work did not yet take multi-user explanations into account (i.e., different users may prefer different explanations for the same actions).

We belief that XAI (especially when based on folk psychology) may be received differently by different users. Folk psychology based explanations evolve as humans mature. For example, very young children (4 years old) are not good at understanding someone might belief something that is false [8]. Furthermore, children and adults alike are inclined to belief that others have similar beliefs and knowledge as we do [6]; however, adults have more (and different) knowledge. Another reason to assume differences between child and adult explanations can be found in adult learning psychology. Adults strongly desire to know the goals they are pursuing when provided with instructional material, so that they can link it to their already existing knowledge [10, 11].

The PAL-agent can use beliefs and goals when explaining its actions by implementing existing XAI technology [7]; however, we need to determine what (subset of these) beliefs and goals are most relevant to specific types of users, (e.g., children might prefer different explanations than adults). For example, when the NAO-robot tells the child that there is a hypo when blood glucose measurement is below 4.0 mmol/L, then we may explain this by saying it wants to teach the child how to detect and treat having a hypo. Other reasoning that drove the PAL-agent may be that it thinks the child does not know when one has a hypo, that it thinks the child is in a good mood to learn new information, and that it wants to be a good diabetes assistant. Providing all this information to the user may be too much, earlier work has confirmed that explanations should not be too long [6, 7]. We need to determine what information the PAL-agent should communicate to its users.

We started our research on multi-user explanations in a diabetes autumn camp in the Netherlands. Here we implemented XAI technology for a Nao-robot. The Nao-robot provided children and their parents with several examples where it acted in particular ways to help a fictional child ‘Jimmy’ to cope with diabetes. For all these examples the Nao-robot proposed two explanations for its behavior. One explanation that communicated the robot’s goal that it was pursuing, and one explanation that communicated the triggering condition (belief) that caused the robot to perform this particular action. For example, the Nao-robot told Jimmy to take a dextrose when experiencing a hypo. It can explain this by (goal) telling it want to teach the child to cope with having hypos; or, (belief) it thinks the child does not know how to correct its blood sugar when experiencing a hypo. Figure 4 shows the set-up of the experiment. The Nao-robot is presenting the examples and explanations. On the laptop screen the participant can read-back the example and explanations presented by the Nao-robot.

From this experiment we learned that adults (the parents) have a significantly higher tendency to prefer goal-based action explanations over belief-based action explanations. This is fundamental knowledge for implementing an XAI module in the PAL-system. Annex 4 discusses the performed experiment in more detail.

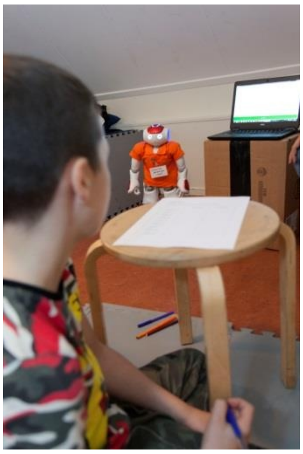


Figure 4: Experiment on multi-user explanations

Conclusion, relation to milestones and feedback

In this year we have reached several major outcomes. First, we have evaluated the goal setting interface with health care professionals, parents and children. This gave us important guidelines to further development the goal setting interface for profesisonals. As mentioned in the review feedback, this is important for the success of the PAL project, as proper goal setting is crucial for the children’s adherence and motivation as well as acceptance by health care professionals (as they need to work with the goal setting tool).

Second, we have gathered, to our knowledge for the first time, evidence that robots can use communication style in their communication towards children. This is a major breakthrough as this opens up the road towards personalization of robot behaviors towards children and learning goals.

Third, we have shown that there are individual differences in preferences towards how a robot explains its actions to users. This shows the need for personalized explanations of actions of the PAL robot and avatar.

With regards to the review feedback about system evaluation, we believe the PAL project has a good apraochs towards evaluation of the system. The PAL prototype system is tested several times per year with all stakeholders. For the coming year an evaluation is planned in which several functionalities will be introdcuced sequentially so that the impact on the children's system usage can be evaluated each time a new piece of functionality is introduced. This allows us to investigate the impact of particular funcitonality on adherence, system usage and motivation. The use of controls in this project is possible, but should be arranged by the hospitals if needed.

Also, reviewers stated that they expect the robot to allow for more compex interaction during the next evaluation. This is the case, we will allow for example simple small-talk, goal-based quizes, and other activities. However this is a result of all partners together, and not specifically tied to this work package.

With regards to milestone 2.1, the prototype context and engagement-based strategy selection system integrated in PAL, we believe that we have successfully achieved this. First, goals are now integerated in the activity selection of the PAL system (e.g., in quizes, games and other activities). A prototype sentiment analysis system has been develeoped that can be used as input for the PAL agent. The system has two different operational modes depending on the context: hospital or home. What is currently mising is that the PAL system on the tablet is not selecting different behavior based on the context (e.g., at home or at school). However, given that the focus has been on integrating the learning- goal related context and user sentiment, we feel this is not a huge amiss.

References

1. Schell, J., *The art of game design: A book of lenses.* . 2008., Amsterdam: Morgan Kaufman Publishers.
2. Grasha, A.F., *A Matter of Style: The Teacher as Expert, Formal Authority, Personal Model, Facilitator, and Delegator.* College Teaching, 1994. **42**(4): p. 142-149.
3. Nguyen, T.-H.D., et al., *Modeling Warmth and Competence in Virtual Characters*, in *Intelligent Virtual Agents: 15th International Conference, IVA 2015, Delft, The*

Netherlands, August 26-28, 2015, Proceedings, W.-P. Brinkman, J. Broekens, and D. Heylen, Editors. 2015, Springer International Publishing: Cham. p. 167-180.

4. Bloom, B. (1956). Taxonomy of educational objectives. Vol. 1: Cognitive domain. New York: McKay.
5. Espinoza, R. R. et al., (2011). Child-Robot Interaction in The Wild: Advice to the Aspiring Experimenter. ICMI'11.
6. Keil, F. C. (2006). Explanation and understanding. Annual review of psychology, 57, 227.
7. Harbers, M., Broekens, J., Van Den Bosch, K., & Meyer, J. J. (2010, January). Guidelines for developing explainable cognitive models. In Proceedings of ICCM (pp. 85-90).
8. Mayringer, H. W. H. (1998). False belief understanding in young children: Explanations do not develop before predictions. International Journal of Behavioral Development, 22(2), 403-422.
9. Malle, B. F. (2004). How the mind explains behavior: Folk explanations, meaning, and social interaction. MIT Press.
10. Lieb, S., & Goodlad, J. (2005). Principles of adult learning.
11. Knowles, M. S. (1970). The modern practice of adult education (Vol. 41). New York: New York Association Press.
12. Lam, D. N., & Barber, K. S. (2005, July). Comprehending agent software. In Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems (pp. 586-593). ACM.
13. Shortliffe, E. H., Davis, R., Axline, S. G., Buchanan, B. G., Green, C. C., & Cohen, S. N. (1975). Computer-based consultations in clinical therapeutics: explanation and rule acquisition capabilities of the MYCIN system. Computers and biomedical research, 8(4), 303-320.

Annex 1

Guidelines for Tree-based Learning Goal Structuring

Bibliography: Rifca Peters, Joost Broekens, Mark Neerincx, *Guidelines for Tree-based Learning Goal Structuring*, IUI 2017, March 13-16, 2017, Limassol, Cyprus.

Abstract Educational technology needs a model of learning goals to support motivation, learning gain, tailoring of the learning process, and sharing of the personal goals between different types of users (i.e., learner and educator) and the system. This paper proposes a tree-based learning goal structuring to facilitate personal goal setting to shape and monitor the learning process. We developed a goal ontology and created a user interface representing this knowledge-base for the self-management education for children with Type 1 Diabetes Mellitus. Subsequently, a co-operative evaluation was conducted with healthcare professionals to refine and validate the ontology and its representation. Presentation of a concrete prototype proved to support professionals' contribution to the design process. The resulting tree-based goal structure enables three important tasks: ability assessment, goal setting and progress monitoring. Visualization should be clarified by icon placement and clustering of goals with the same difficulty and topic. Bloom's taxonomy for learning objectives should be applied to improve completeness and clarity of goal content.

Relation to WP. Setting appropriate learning goals is extremely important to give focus and a sense of progress to the child. This needs to be collaboratively, so that the child feels he/she is problem owner. As such a goal setting interface, and a method to then adapt the PAL's system behavior based on these goals, facilitate the learning process.

Availablity Unrestricted. Available for download (https://doi.org/10.1145/3025171.3025188)

Annex 2

Robot Style Perception by Children

Bibliography Rifca Peters, Joost Broekens, Mark Neerincx, submitted to ROMAN.

Abstract to be finalized

Relation to WP. Robot style is important because it facilitates engagements. Human educators change their educational style dependent on the child and learning task. For example, dominant communication is needed at a different moment than submissive communication. If a robot cannot change its communication style, it will be perceived – in the long run – as boring and non-personal and at the very least not adaptive towards the learning needs of the child. This may result in lower engagement or dropping out.

Availablity Not yet published

Annex 3

Kaptein et al (2016), "CAAF: A Cognitive Affective Agent Programming Framework"

Bibliography Kaptein, F., Broekens, J., Hindriks, K. V., & Neerincx, M. (2016, September). CAAF: A Cognitive Affective Agent Programming Framework. In International Conference on Intelligent Virtual Agents (pp. 317-330). Springer International Publishing.

Abstract Cognitive agent programming frameworks facilitate the development of intelligent virtual agents. By adding a computational model of emotion to such a framework, one can program agents capable of using and reasoning over emotions. Computational models of emotion are generally based on cognitive appraisal theory; however, these theories introduce a large set of appraisal processes, which are not specified in enough detail for unambiguous implementation in cognitive agent programming frameworks. We present CAAF (Cognitive Affective Agent programming Framework), a framework based on the belief-desire theory of emotions (BDTE), that enables the computation of emotions for cognitive agents (i.e., making them cognitive affective agents). In this paper we bridge the remaining gap between BDTE and cognitive agent programming frameworks. We conclude that CAAF models consistent, domain independent emotions for cognitive agent programming.

Relation to WP. Cognitive agent programming frameworks facilitate the development of agents like the PAL-agent. This publication shows how a computational model of emotion is added to such a framework. This modelling of emotions provides high-level current- and desired state information that is used in the action selection module (WP3). The work presented here is further implemented in the PAL-system, where emotions are simulated concerning the child's performance while playing the quiz game.

Availablity Unrestricted. Available for download
(http://dx.doi.org/10.1007/978-3-319-47665-0_28)

Annex 4

Kaptein et al (2016) “Adults prefer goal-based agent-action explanations over belief-based explanations more so than children”

Bibliography Kaptein, F., Broekens, J., Hindriks, K. V., & Neerincx, M. (2017). Adults prefer goal-based action explanations over belief-based explanations more so than children *to appear*

Abstract Explainable Artificial Intelligence (AI) enables BDI-based agents to explain their actions to their users. This is achieved by narrating the agent's beliefs and goals that caused it to perform the action. However, many beliefs and goals generally precede a BDI-based agent's decision to perform an action. Explanations fail if they contain too much information. The key challenge is to choose what information should be communicated to a user. In addition, different users may need different explanations (i.e., we need *multi-user explanations*). In this paper, we define algorithms for goal-based and belief-based action explanations, based on previous work on explaining agents. We let a Nao-robot provide two user groups (children and adults) with example actions explained by these two explanation algorithms. For every action, we asked them what explanation they prefer. From this, we learned that adults have a significantly higher tendency to prefer of goal-based agent-action explanations over belief-based explanations. This is a first, important step in addressing the challenge of multi-user explanations.

Relation to WP. This publication is part of task 2.1, where an Explainable AI module is developed to facilitate shared patient-caregiver responsibility. The intelligence here is directly relevant to the authoring tool, parental monitoring interface, and PAL strategic reasoning engine, where it can explain the caregiver, parent, and patient *why* the PAL-agent acted in a particular way.

Availablity Not yet published

Annex 5

Ontologies for social, cognitive and affective agent-based support of child's diabetes self-management

Bibliography Neerincx, M., Kaptein, F., van Bekkum, M., Krieger, H. U., Kiefer, B., Peters, R., Broekens, J., Demiris, Y., & Sapelli, M. (2016). Ontologies for social, cognitive and **affective** agent-based support of child's diabetes self-management. *Artificial Intelligence for Diabetes*, 35.

Abstract The PAL project is developing: (1) an embodied conversational agent (robot and its avatar); (2) applications for child-agent activities that help children from 8 to 14 years old to acquire the required knowledge, skills and attitude for adequate diabetes self-management; and (3) dashboards for caregivers to enhance their supportive role for this self-management learning process. A common ontology is constructed to support normative behavior in a flexible way, to establish mutual understanding in the human-agent system, to integrate and utilize knowledge from the application and scientific domains, and to produce sensible human-agent dialogues. This paper presents the general vision, approach, and state of the art

Relation to WP This paper shows the current progress and vision of the PAL-system. The development of ontologies facilitates the development of strategic goal selection, and normative behavior. Furthermore, the ontologies facilitate mutual understanding between the **different** users and the PAL-agent. This paper shows the high level design of the PAL-system, and how the ontologies support the development and intelligence of this system.

Guidelines for Tree-based Learning Goal Structuring

Rifca Peters
Delft University of Technology
Delft, The Netherlands
r.m.peters@tudelft.nl

Joost Broekens
Delft University of Technology
Delft, The Netherlands
d.j.broekens@tudelft.nl

Mark A. Neerincx
Delft University of Technology
Delft, The Netherlands
m.a.neerincx@tudelft.nl

ABSTRACT

Educational technology needs a model of learning goals to support motivation, learning gain, tailoring of the learning process, and sharing of the personal goals between different types of users (i.e., learner and educator) and the system. This paper proposes a tree-based learning goal structuring to facilitate personal goal setting to shape and monitor the learning process. We developed a goal ontology and created a user interface representing this knowledge-base for the self-management education for children with Type 1 Diabetes Mellitus. Subsequently, a co-operative evaluation was conducted with healthcare professionals to refine and validate the ontology and its representation. Presentation of a concrete prototype proved to support professionals' contribution to the design process. The resulting tree-based goal structure enables three important tasks: ability assessment, goal setting and progress monitoring. Visualization should be clarified by icon placement and clustering of goals with the same difficulty and topic. Bloom's taxonomy for learning objectives should be applied to improve completeness and clarity of goal content.

ACM Classification Keywords

H.5.2 Information Interfaces and Presentation (e.g. HCI): User Interfaces; H.5.3 Information Interfaces and Presentation (e.g. HCI): Group and Organization Interfaces

Author Keywords

Diabetes; Healthcare; Education; Learning goal -setting -attainment; Personalization; Collaboration; Visualization; Knowledge-base.

INTRODUCTION

Advancements in media technologies provide new opportunities for education. For example, Intelligent Tutoring Systems (ITSs) provide immediate tailored instructions or feedback to a learner to facilitate effective learning while lessening the students dependency on a teacher. Also, consider eHealth applications that have been designed to increase a person's knowledge and control over health and well-being. Especially in self-regulated learning motivation is highly important to optimize adherence to the education program [18]. Research suggests that goal setting and feedback on goal attainment enhance motivation (e.g., [7, 10, 16]) and learning gain (e.g., [5, 9]). Moreover, personal goal setting allows for tailoring of the learning process, this is applicable to personalization of educational technology (e.g., [14]).

Incorporating personal learning goals in educational technologies requires knowledge of learning goals relevant to the domain, a mechanism to set personal learning goals and to share this information between different types of users (e.g., doctor and patient, teacher and student) and with the system, and means to monitor learning progress. Ability-trees are used in games to structure and visualize skills that allow the player to tailor character development and game-play. Gamification has been applied to educational technologies. For example, task completion is rewarded with points or achievements. Using an ability-tree for learning goals is an interesting approach to provide a solution for goal structuring, setting, and monitoring.

In this paper we propose guidelines for a tree-based learning goal model and user interface to support collaborative goal setting. In a case study, on self-management education for children with Type 1 Diabetes Mellitus, we explore requirements for a tool to set personal learning goals. We developed a knowledge-base (ontology) formalizing learning goals and tasks based on medical protocols. We created a user interface presenting these learning goals in a tree-based graph, and enabling personal goal setting. Based on a co-operative evaluation we formulated guidelines to improve the design.

BACKGROUND

Learning Goals

Effective learning requires commitment, adherence and motivation, which can be increased by learning goals [7, 9]. Goals enhance motivation independent of their source (i.e., assigned, self- or collaborative set), if relevance is provided [11]. However, performance is lower for unexplained, assigned goals than self- or collaboratively set goals [12]. Contrary, Kleinrahm et al. [10] found that cooperative goal setting and reflection increased motivation.

Black and Wiliam [5] concluded that awareness of goals and goal attainment improves learning gain. Similarly, goal-setting theorist believe that feedback results in setting higher goals [12]. Which in turn, leads to better performance [11]. This fits Vygotsky's [21] theory on the zone of proximal development (ZPD), predicting that experiences slightly advancing current abilities encourage and advance learning [17].

IUI 2017, March 13-16, 2017, Limassol, Cyprus
© 2017 ACM. ISBN 978-1-4503-4348-0/17/03...$15.00
DOI: http://dx.doi.org/10.1145/3025171.3025188

In instructional classroom learning goals are often strategically chosen to align with institutional or national standards. However, in self-paced learning the education process can and should be tailored by the learner. Self-regulated learning (SRL) increases learning gain, but is more demanding in terms of effort and thus motivation [18]. SRL theorists believe that strategies such as goal setting, self-monitoring and self-evaluation are vital for effective learning [23]. Motivation comes from goal orientation, self-efficacy believes, task value, and outcome expectations [18, 15]. SRL benefits from a mastery goal orientation (i.e., focus on learning of the ability), while a performance orientation (i.e., focus on demonstration of abilities) declines performance [2, 22, 15].

Ability Trees

In graph theory a tree is defined as an undirected graph in which two nodes are connected by exactly one edge. In a directed graph edges have an associated direction, and in a rooted tree one node is designated as the root. In computer science a tree is a non-linear, hierarchical data structure represented by a root node and linked children or sub-trees [19]. The direction can be from (out-tree) or to (in-tree) the root [13]. In the remainder of this paper we use *tree* as data structure.

Graph theory based models are used in e-learning environments guiding the self-paced learning process, and enabling personalization thereof (e.g., [4, 20]). Learning object graphs formalize the structure of a course, representing mandatory and recommended learning objects—including objectives but mainly content—and relations between them. Through assessment and authoring, objectives are selected, constructing individual learning paths that align with the learner's profile.

Tree-based data structures are used in games allowing players to customize their game experience. For example, in the strategy game Civilization players can choose to develop skills in alphabet or mathematics, but only after having achieved the writing skill. In role-playing games such as Diablo (I and II), players develop their character using points to gain magical powers. These so called ability-trees are visual, hierarchical representation of possible sequences of developments. Abilities are displayed in branching paths and open up after completing required prerequisites. These structures, based on abilities opposed to levelling, allow players to excel in some areas while progressing more slowly, or not at all, in others. Expected is that ability-trees are familiar, and therefore understandable, to children.

CASE: DIABETES SELF-MANAGEMENT EDUCATION

Current Practises in Diabetes Education

Type 1 Diabetes Mellitus (T1DM), diagnosed by a growing number of children, is a high impact digestion disease which requires daily self-management. Thus, to improve well-being and avoid complications, long-term behaviour change is necessary [8]. Learning objectives are personal and change while ageing. Therefore, self-management education is highly personalized. It is directed by challenges faced in daily life and aimed at gradual development of attitudes, knowledge and skills needed for autonomous self-management.

In the Netherlands, formalization of learning goals is limited to (annual) check lists arranged by topic and age such as the *weet & doe-doelen* (knowledge & skill goals) composed by the Dutch organization for diabetes nurses EADV (http://www.eadv.nl). Hierarchical relations between goals are implied; a goal can be prerequisite for one or more others (e.g., injecting insulin requires knowledge of appropriate body parts). Further, goals become increasingly complex for older children. As a result, goals may cover multiple topics and thus precede or succeed goals on different topics. Moreover, specific goals are irrelevant to some children (e.g., pump users do not necessarily need to learn injecting themselves).

Active involvement of patients in the disease management and education process is essential [1]. Objectives should be defined collaboratively between patient and caregivers. This is in line with the Motivational Interviewing (MI) guiding style adopted in healthcare counselling. The principle of MI [16] is to explore and resolve a patient's ambivalent feelings towards change, opposed to coercing or persuading. MI is believed to increase the patients commitment to developing self-management abilities.

Diabetes Education Framework

The PAL project[1] develops mHealth technology providing educational support to children with T1DM. The aim is to gradually increase children's self-management abilities and responsibilities. The envisioned system includes an embodied conversational agent (robot and avatar), extra-curricular educational child-agent activities, and an authoring tool.

Authoring Tool

The authoring tool is a web-based application for healthcare professionals (HCPs) designed to support goal setting, progress monitoring, and attainment registration. The current state is a functional, but minimal, prototype presenting diabetes self-management learning goals, and providing an interface to set personal goals or register attainment together with a child.

Diabetes Learning Goal Structure

We formalized the learning goals, as proposed by the EAVD, in an ontology (Figure 1). Learning goals are classified by type (i.e., knowledge or skill) and values are given for difficulty and topic. Additionally, restrictions are added for prerequisite goals. (The progress and state properties are specific to a child and values are given at a later time.) Further, achievements are added for each topic and difficulty combination, and tasks (e.g., 'win a quiz on insulin') are linked to learning goals (e.g., 'know locations for insulin injection'). The restrictions and relations allow the system to provide personalized content, and calculate and update goal progress automatically.

We created a tree-based visualization of the goals and achievements (Figure 2) because the merging structure of an in-tree fits the diabetes learning goals; from leafs with a single focus topics (e.g., 'Nutrition'or 'Insulin') to multilevel topics (e.g., 'Nutrition in social context') and ultimately the root node 'Self-management'. Further, tree presentations have been applied successfully to structure abilities in games.

[1] Personal Assistant for a healthy Lifestyle: http://www.pal4u.eu/

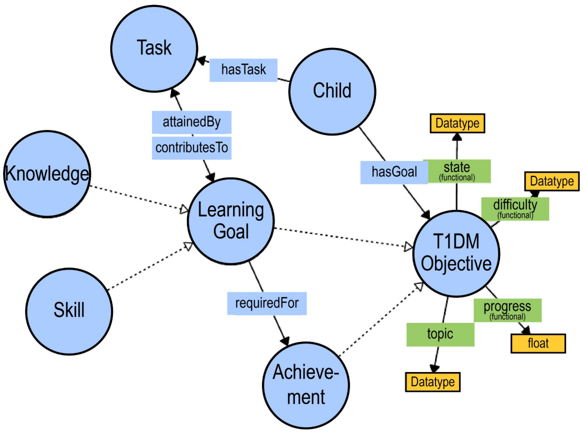


Figure 1. The Diabetes Education ontology. Nodes depict the objective types (classes). Arrows depict object or data properties, dotted arrows depict subtype relations. Self-management learning goals are instantiated in the knowledge or skill class.

Nodes represent a learning goal (i.e., knowledge or skill) or achievement. Edges depict connections to prerequisite nodes. Attaining all learning goals on one topic at one difficulty grants the achievement and unlocks the possibility to advance on this topic. For example, a child who attained all difficulty 1 goals on glucose (i.e., knows why measurement is needed, how to correct a hypo, and understands the measurement value) is granted the achievement 'Novice Glucose', and may advance to the next level. Edges connecting a node to multiple, represent nodes prerequisite to more than one others. For example, knowledge of the correct response to a glycemic value is required for both 'Basic Glucose' and 'Independent Measurer'. The proposed model facilitates tailoring to a child's situation and development by selection of personal learning goals, while enforcing to have obtained prerequisite abilities.

Collaborative Goal Setting Interface

A minimal graphical interface was created to support goal setting and progress monitoring (Figure 2). It displays the goal-tree and provides mechanisms to switch between goal states (active, inactive and attained). Active goals, pursued in the near future, are yellow. Attained goals are green. Inactive goals are greyed-out but visible to raise awareness, allowing children to ask about them. The goal state is changed by clicking the node (register attainment) or the selection box at the top-left corner (activate). Upon activation, all prerequisite goals activate automatically, the user can inactivate (irrelevant) goals clicking the selection box. The goal-tree is presented top-down, first displaying goals for new learners, to avoid frequent scrolling.

METHODOLOGY

To gain insight about diabetes education protocols and elicit user requirements, interviews were conducted with HCPs at an early stage. Implicit knowledge and experience appeared fundamental to forming of the, highly personalized, educational process. Besides, HCPs are not used to thinking from a technological design perspective. Hence, development of a goal ontology and user interface were no trivial tasks. Therefore, we selected a co-operative, formative evaluation method providing a minimal example and collaboratively composing guidelines for further development.

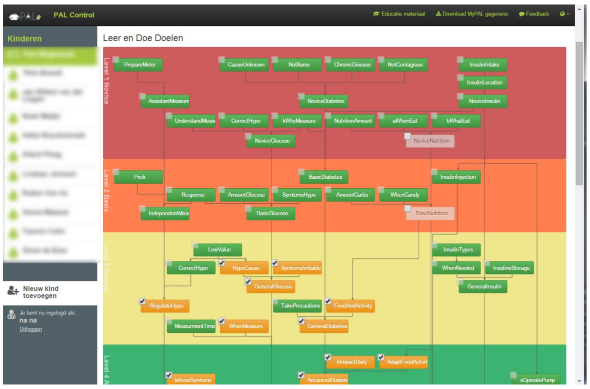


Figure 2. The authoring tool (PAL Control), displaying the diabetes learning goals (i.e., knowledge and skills to attain to progress towards self-management) and achievements with current progress for a 10 year old boy. Attained objectives are green, yellow ones are active. Coloured horizontal bars depict difficulty levels. Topics are arranged vertically.

Evaluations were conducted with 7 HCPs (6 Dutch nurses, 1 Italian doctor), and 35 children (aged 7-12 M/F) and their parents, in 3 hospital (2 Dutch, 1 Italian) in May-June 2016. Each child visited the hospital two times, once at the start and end of a three week period. In between children played educational activities at home. The first consultation covered personal goal setting using the authoring tool. In the second meeting progress was discussed. An observer was present and audio, keystrokes, and clicks were recorded during at least each nurse's earliest consultation. Additionally, in the Netherlands, training sessions of approximate half an hour were carried out prior to consultations using a think-out-loud protocol whilst preparing goals for the first child. Semi-structured interviews were conducted posterior with all professionals.

RESULTS AND DISCUSSION

A total of 11 goal setting consultations have been observed and analysed. A typical consultation was attended by a paediatric diabetes nurse, child, and one parent and lasted for about 10 minutes. All meetings included assessment of the child's current abilities, goal suggestions by the professional, conformation by the child and/or parent, and registration of goal state (active, inactive or attained) in the authoring tool. These steps were repeated for individual goals by a top-down walk through of the goal-tree.

Assessment of the child's current abilities was mostly straight forward: the nurse asked whether or not the child knows or can do x, where x is a specified goal. In six sessions answers were given by the child and parent in collaboration or turn, in three sessions the child responded alone, in two sessions only the parent was involved (both cases a 7-year old child). In four occasions assessment was done more implicitly by 'small-talk' (e.g., "Your horseback riding right? Do you have any difficulties with your diabetes then?"). If agreed on goal attainment, it was registered as such. A goal was set active if no agreement was reached. Assessment was suspended for a topic if goals were set active.

Although the selection mechanism was easily understood and goal setting was done effectively, the *topic clusters* were not clear. For example, Nurse 1 was looking at glucose goals in search for a goal on nutrition. Further, the *achievement concept* was hard to grasp. For example, instead of activating goals and achievement, Nurse 1 explained that she did not select the achievement because the related goals were not yet attained, and Nurse 3 marked achievements attained while related goals were still active. In addition, *handling irrelevant goals* was troublesome for two nurses. For example, Nurse 4 registered 'Insulin Injection' attained because the child, as a pump user, did not need to learn how to inject insulin by pen (opposed to leaving it inactive). Moreover, questions were raised about the *intent or meaning for specific goals*. For example, Nurse 3 doubted whether to activate 'Insulin-type Needed', because the child did know the facts but was not yet able to apply this knowledge in daily situations. Further, nurses were unable to select goals they had in mind for a child because they were not present in the goal-tree. For all but one, this were attitude goals such as feeling more secure. In posterior interviews with HCPs, the following five issues on the goal ontology and user interface were discussed.

First, the *complexity of the goal-tree* was too high. For the nurses, the meaning of 'achievement' (i.e., *not* concerning new knowledge or skill) was not clear. Furthermore, the user interface was not clear: Icons to clarify the topic were proposed, and all HCPs suggested or favoured visualization of achievements by displaying related goals in nested nodes. One nurse suggested showing only the active difficulty level, hiding others. To support tailored learning goals, this must be done per topic or a collapse-on-select mechanism could be considered: when selecting an achievement it unfolds and enables (de)activating containing goals.

Second, the *clarity and completeness of the goal content* showed shortcomings. The ontology did not include attitude goals and lacked distinction between factual knowledge and the ability to apply this knowledge. Application of Bloom's Taxonomy [6] for educational objectives and gradual development of more complex skills might solve these shortcomings, distinguishing cognitive (knowledge), psycho-motor (skill) and affective (attitude) processes. The cognitive dimension includes multiple levels such as remembering, understanding and applying [3]. Explicitly formulating goals on these levels bridges knowledge development, practice and assessment.

Third, *feedback on goal attainment for progress monitoring* was missing (i.e., whether a goal was registered attained manually or by the system). When the start and end meeting were attended by different nurses, they were unsure about newly attained goals. So, information about goal-state changes should be provided for nurse shiftes and time periods (i.e., simple visualization of goals attained since the last meeting by placement of an icon).

Fourth, *goal setting and assessment of current abilities* were partly supported. Although goals can be selected at any difficulty, while the recursive mechanism ensures that prerequisites are activated, nurses started at the top of the tree and worked their way down assessing each goal. As a result, an overview of current abilities was created, and for each topic unattained goals were set active. This is different from the current practise where a single focus is chosen based on the child's experiences. Nurses had different preferences, either favouring several goals allowing the child varied experiences in the learning framework or favouring a mechanism selecting a focus topic from active goals. Nurses agreed that the top-down goal assessment provided a valuable overview of the current state of abilities of a child. It was time consuming, but nonetheless considered more usable than current check-lists. Three nurses suggested to let the child play a game (e.g., with the robot) to assess abilities and serve as input for goal setting.

Fifth, *collaboration with the child during goal setting* was partly supported. According to the nurses, the authoring tool eased interaction between them and the child. However, goal setting was less collaborative than desired. Children's involvement was limited to (dis)agreement; they did not proactively discuss specific goals, while active involvement is key to motivation for behaviour change [16]. Children's involvement can be improved by allowing them to select their personal focus from active goals. Moreover, HCPs may benefit from training in collaborating with a child using the authoring tool. Nonetheless, goal setting was believed helpful making the education process more interesting and transparent to the children.

The present study has some limitations such as a lack of quantitative data proposed by the number of participants and method. We do not report on usability statistics because they do not provide novel information for research and development of intelligent user interfaces. We plan to expand our user base and methods to evaluate alternative interfaces and investigate the effect of goal setting on learning outcomes. Although, other structures might be feasible as well, we have chosen a tree-structure because this fits our domain. Further research, presenting alternative structures, is needed to make any conclusions on structure preferences. Suggestions provided in this paper are applicable to tree-based structures.

CONCLUDING GUIDELINES

The main challenges addressed in the present work are the development of an ontology for diabetes self-management education and an interface to support collaborative personal goal setting and monitoring. An authoring tool was created for this purpose and co-evaluated with healthcare professionals.

Guideline 1: The authoring tool should provide *clear, visual feedback* on goal *structure*, and *active state and progress*. For example, by usage of icons depicting topic and state changes.

Guideline 2: The *different concepts* of the model (e.g., goal and achievement) should consistently have a *different representation* (e.g., shape) in the user interface .

Guideline 3: The domain should be fully covered in goal content. In our case *differentiation* between *affective and cognitive*, and *factual and application* objectives should be embedded (e.g., Bloom's taxonomy).

Guideline 4: The authoring tool should support, next to goal setting, progress monitoring and attainment registration, *assessment of current abilities*.

ACKNOWLEDGMENTS

This work is funded by the EU Horizon 2020 PAL project (grant number 643783). We would like to thank the healthcare professionals from Gelderse Vallei, MeanderMC and Ospedale San Raffaele, without their efforts this work would not have been possible.

REFERENCES

1. American Diabetes Association. 2013. Diagnosis and classification of diabetes mellitus. *Diabetes Care* 36, SUPPL.1 (2013), 67–74.
2. C. Ames. 1992. Classrooms: Goals, Structures, and Student Motivation. *Educational Psychology* 84, 3 (1992), 261–271.
3. L.W. Anderson, D. R. Krathwohl, and B.S. Bloom. 2001. *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Allyn & Bacon.
4. Y. Atif, R. Benlamri, and J. Berri. 2003. Learning Objects Based Framework for Self-Adaptive Learning. *Education and Information Technologies* 8, 4 (2003), 345–368.
5. P. Black and D. Wiliam. 1998. Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice* 5, 1 (1998), 7–74.
6. B.S. Bloom. 1956. Taxonomy of educational objectives. Vol. 1: Cognitive domain. (1956).
7. J.F. Bryan and E. A. Locke. 1967. Goal setting as a means of increasing motivation. *Applied Psychology* 51, 3 (1967), 274–277.
8. D. Freeborn, T. T. Dyches, S. O. Roper, and B. Mandleco. 2013. Identifying challenges of living with type 1 diabetes: Child and youth perspectives. *Clinical Nursing* 22 (2013), 1890–1898.
9. A. M. Grant. 2012. An integrated model of goal-focused coaching : An evidence-based framework for teaching and practice. *International Coaching Psychology Review* 7, 2 (2012), 146–165.
10. R. Kleinrahm, F. Keller, K. Lutz, M. Kölch, and J. Fegert. 2013. Assessing change in the behavior of children and adolescents in youth welfare institutions using goal attainment scaling. *Child and Adolescent Psychiatry and Mental Health* 7, 33 (2013).
11. E.A. Locke and G.P. Latham. 2006. New Directions in Goal-Setting Theory. *Current Directions in Psychological Science* 15, 5 (2006), 265–268.
12. E.A. Locke and G. P. Latham. 2002. Building a Practically Useful Theory of Goal setting and Task Motivation: A 35-Year Odyssey. *American Psychologist* 57, 9 (2002), 705–717.
13. K. Mehlhorn and P. Sanders. 2008. *Algorithms and Data Structures: The Basic Toolbox*. Springer Science & Business Media.
14. M.A. Neerincx, F. Kaptein, M. A. Van Bekkum, H. Krieger, B. Kiefer, R. Peters, J. Broekens, Y. Demiris, and M. Sapelli. 2016. Ontologies for social, cognitive and affective agent-based support of child's diabetes self-management. In *Proc. ECAI'16*. 35–38.
15. P. R. Pintrich. 1999. The role of motivation in promoting and sustaining self-regulated learning. *Educational Research* 31, 6 (1999), 459–470.
16. S. Rollnick, W. R. Miller, and C. C. Butler. 2008. *Motivational Interviewing in Health Care: Helping Patients Change Behavior*. The Guilford Press, New York, London.
17. B. R. Schadenberg, M. A. Neerincx, F. Cnossen, and R. Looije. In Press. Personalising game difficulty to keep children motivated to play with a social robot: A Bayesian approach. *Cognitive Systems Research* (In Press).
18. D. H. Schunk and B. J. Zimmerman. 2008. *Motivation and Self-Regulated Learning: Theory, Research, and Applications*. Routledge.
19. T. A. Sudkamp. 2005. Mathematical Preliminairies. In *Languages and machines: an introduction to the theory of computer science*. Addison Wesley.
20. A. N. Viet and D. H. Si. 2006. ACGs: Adaptive Course Generation System - An Efficient Approach to Build E-learning Course. In *Proc. CIT'06*. 259–265.
21. L. S. Vygotsky. 1980. *Mind in society: The development of higher psychological processes*. Harvard university press.
22. C. A. Wolters, L. Y. Shirley, and P. R. Pintrich. 1996. The relation between goal orientation and students' motivational beliefs and self-regulated learning. *Learning and Individual Differences* 8, 3 (1996), 211–238.
23. B. J. Zimmerman. 1986. Becoming a self-regulated learner: Which are the key subprocesses? *Contemporary Educational Psychology* 11, 4 (1986), 307–313.

Robots Educate in Style: The Effect of Context and Non-verbal Behaviour on Children's Perceptions of Warmth and Competence

Rifca Peters[1], Joost Broekens[1] and Mark A. Neerincx[1,2]

Abstract— **Social robots are entering the private and public domain where they engage in social interactions with non-technical users. This requires robots to be socially interactive and intelligent, including the ability to display appropriate social behaviour. Yet, it remains unclear what behaviour style is appropriate and *how* to express this; no comprehensive, validated model exists of non-verbal behaviour as perceived by the user. Based on a literature survey, we created a model of non-verbal behaviour to express high/low warmth and competence —two dimensions fundamental to social perception. We applied this model to a NAO robot, and evaluated this in a field-perception study at primary schools and a diabetes camp in the Netherlands. For this, we developed, based on expert ratings, an instrument measuring perceived warmth, competence, dominance and affiliation. The competence model appeared successful, and an interaction effect was found for perceived warmth. Moreover, the context significantly influenced children's perceptions of the robot. Our study shows that even subtle manipulations influence how children perceive an educational robot.**

I. Introduction

While robots emerge as collaboration partner in training and education (e.g., [20], [34], [19]), the urgency towards social intelligence is stressed [14]. Educational robots take the role (i.e., a consistent behaviour pattern that evokes certain percepts) of tutor, tool or peer; learning from, about or with robots [23]. Different roles have been linked to different learning activities. For example, for basic learning tasks a peer robot was preferred above a tutor robot [25], but for language learning a tutor was preferred [30]. The quality of human teacher-student interactions is partly determined by the teacher's ability to adapt their style (i.e., patterns in behaviour variations and associated attitudes) to the student and activity [16]. Thus, roles can be performed with various styles, and effective educational robots should, like human educators, be able express appropriate styles.

Roles and style are being used in experimental human-robot interactions (e.g., [22], [35]). However, these studies seldom report a validated model of behaviours to express this. For example, in [22] a *motivator-robot* was intended to signal *empathy* and *trustworthiness* by listener behaviour (gaze, nod and 'listening expression'), but validation of this model has not been reported. Attempting to bridge the social intelligence gap between humans and machines, social signal processing (SSP) focusses on modelling, analysis and synthesis of non-verbal behaviour. Social signals are, often unconscious, expressions of someone's attitude towards interactions such as agreement, dominance and hostility, displayed by non-verbal behaviour cues (e.g., prosody, gestures, posture) [32], [27]. Research includes modelling of agent non-verbal behaviour based on analysis of human behaviour (e.g., [8], [1], [24]). In the multidisciplinary fields work is being done understanding human behaviour, and investigating *what* robot role is appropriate. However, the question of *how* to stylize and express this receives less attention. To the best of our knowledge, no comprehensive, validated model exists of non-verbal behaviour cues to express style, nor of intermediate steps such as social attitudes contributing to style or behaviour cues to communicate attitudes.

In this paper we evaluate a model of non-verbal behaviours to express *warmth* and *competence* for an educational robot. Based on previous literature we created a model of prosodic, posture, and gesture parameters (e.g., fluency, spread) intended to evoke perceived high/low competence and warmth. We evaluated our model, applied to a NAO robot giving an interactive lecture, in a 2×2 (warmth×competence) between-subject perception study at primary schools in the Netherlands and did a follow-up study at a camp for children with type 1 diabetes mellitus (T1DM). Children did perceive the robot displaying competence-related behaviours (stable posture and frequent gestures) as more competent than a low-competence robot. A robot displaying warmth-related behaviours (low voice and pitch, head tilt, and open gestures) was perceived warmer than a low-warmth robot only if the robot was high-competent. Additionally, independent of the robot style, warmth ratings showed higher at schools than at diabetes camp.

II. Background

In human educational interactions, educators adapt both content and style of communication to the learner and task at hand. This ability is believed to be crucial for effective and motivating educational interactions. Style describes, opposed to personality traits and role, the current observable behaviour—the way a role is performed. This can be trained and used strategically.

A. Teaching Style

Grasha [16] defined teaching style as a pattern of needs, beliefs and behaviours that affect information presentation, student interaction, class management, and supervision. Based on literature, observation and interviews, the author constructed a conceptual model of five teaching styles:

* This work is funded by the EU Horizon2020 PAL project (grant 643783)

[1] Interactive Intelligence Group, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands r.m.peters@tudelft.nl

[2] TNO, Postbus 23, 3769 ZG Soesterberg, The Netherlands mark.neerincx@tno.nl

Expert (transmitting information), *Formal authority* (provide feedback and establish boundaries), *Personal model* (show example), *Facilitator* (encourage critical thinking), and *Delegator* (available in the background during project work). Although all teachers are believed to posses all styles to some degree, certain clusters of styles appeared most frequent and have been linked to specific situations and strategies. Teaching style is selected based on student capabilities (e.g., knowledge, responsibility and motivation), teacher's need for control, and teacher's willingness to build and maintain relationships. For example, teacher-centred styles (i.e., expert and formal authority) are linked to teacher control, and limited student capabilities, while student-centred styles (i.e., facilitator and delegator) are linked to teachers willing to loosen control, and more capable students.

B. Interpersonal Circumplex

To create awareness of own behaviour and learn predicting and directing behaviour of others, skill training, such as at the Dutch police academy and teacher training, includes learning and practice the interpersonal circumplex. This model—also know as Leary's Rose—defines interaction stance by two axes: *dominance* and *affiliation* [21], [33]. The horizontal affective-axis depicts willingness to cooperate. The vertical dominance-axis depicts the degree of power. Commonly, the circumplex is partitioned into eight octants, each reflecting a gradual blend to axes (Fig. 1a). Further, Leary's theory states two interaction rules: dominance is complementary and affiliation symmetric, meaning that, opposed stance evokes opposed stance and dominance evokes submissiveness. [31].

A cooperative style has been linked to maintaining contact, a competitive style to aversion. For example, head nods [15], gaze, and open posture [4] are associated with high affiliation. A dominant style is commonly stereotyped as loud and obtrusive, while submissive style is believed to show discrete, unnoted behaviour. Research suggest that loudness, vocal control, and gaze are associated with perceived dominance [12], [6]. Moreover, dominant people are believed to lean forward, use more gestures, have open and up-strait posture, and orient towards others [7]. However, a meta-analysis of the relation between dominance and non-verbal behaviours reported that previous work has been inconclusive or based on limited data. The authors concluded that the relation between dominance and non-verbal behaviour exists to different degrees and even directions, depending on the person and situation [18].

C. Stereotype Content Model

Willingness to build and maintain relationships is related to warmth expressions. The stereotype content model (SCM) defines *warmth* and *competence* as two fundamental dimensions of social perception [13], [11], suggested to account for 82% of variance in perception [9]. Perceived warmth (kindness, empathy, friendliness and trustworthiness) evaluates valance of intent. Perceived competence (intelligence, power, efficacy and skill) assesses the ability to act on these intentions. Perceived warmth and competence generate emotions of admiration, envy, pity, and disgust towards someone, and predict active/passive and facilitative/harmful behaviour patterns (Fig. 1b).

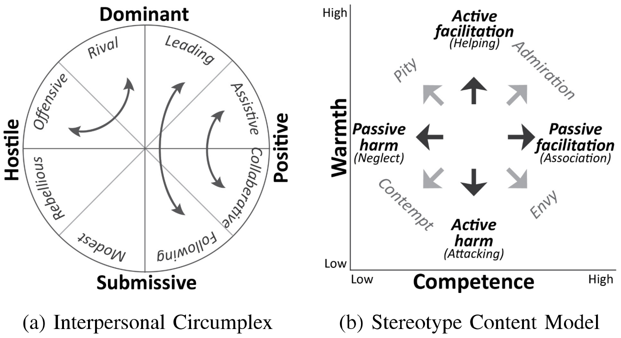
Perceived warmth is believed to be evoked by sincere smiles, head tilt, nodding, leaning forward, and open gestures [6], [10]. While coldness is expressed by closed hands, cutting motion, chin down, and the body pivoting away [10]. Further, vocal warmth (low pitch and volume) has been associated with affiliated, warm behaviours [10]. Upright posture and open gestures, are predictors of perceived power, and associated with competence [6], [10]. Fiddling is suggested to signal low control and confidence, therefore resulting in low-competence judgement [10]. Moreover, warmth judgements are made before, and are likely to influence, competence evaluations—persons evaluated as high-warmth, are likely to be judged more competent [17], [6].

In summary, style can be defined as a behaviour pattern signalling our attitude towards a person or situation, affecting how others evaluate us and respond subsequently. We described four constructs (i.e., *warmth*, *competence*, *dominance*, and *affiliation*) used to evaluate others and their associated non-verbal behaviours.

III. Related Work

In human-robot interaction (HRI) *interaction style* is often defined as a combination of behaviours that evoke a perceivable robot role (e.g., [26], [35]). The roles that educational robots take are tool, tutor, or peer; leering about, from or with robots [23]. For example, in a self-help support program, a (virtual and physical) iCat instructor fulfilled the roles of *buddy* (asking about well-being and showing empathy) or *educator* (informing and asking about health), but also *motivator* (asking about and providing feedback on desired changes) [22]. And, in collaborative play, a NAO robot took the role of *peer* (collaborative behaviour) or *tutor* (scaffolding support) [35]. Recent studies explore the preference for or effect of robot roles. However, a clear report lacks on *how* these roles are performed—the behaviours displayed to evoke role perception. Further, it remains unclear if participants did perceive the robot as intended.

Although, up until now, studies validating behaviour models have been focusing on emotion expression (e.g., [34]), recently interest sparked in modelling of social perceptions



(a) Interpersonal Circumplex (b) Stereotype Content Model

Fig. 1: Models of social interaction evaluation and prediction

such as warmth and dominance. In virtual humans, perceived dominance was influenced positively by head tilt and eyebrow raising, and negatively by head nodding and gaze aversion [2]. Weak support was found that vocal control displayed by a virtual suspect influenced perceived dominance and affiliation [28]. Perceived dominance was higher when started speaking in overlap than bridged or gaped turn-taking. Similarly, in agent-agent interactions continuing at overlap contributed to perceived dominance, and perceived affiliation was higher when speaking started in silence compared to started speaking in overlap [5].

Presence of gestures positively influenced perceived competence compared to absence of gestures [3]. A more elaborate model [1] of virtual agent behaviours expressing warmth and competence appeared successful [24]. The high-warmth characters were perceived as warmer than low-warmth characters, and high-competence characters were perceived as more competent than low-competence characters. The level of competence did not seem to affect the perception of warmth, but high-warmth agents were perceived as more competent than low-warmth agents.

IV. Robot Style Model

Based on the non-verbal behaviours associated with warmth and competence described above, we created a model of non-verbal behaviours for a NAO robot expressing four different style configurations: *high-warmth and high-competence* (HwHc); *high-warmth and low-competence* (HwLc); *low-warmth and high-competence* (LwHc); and *low-warmth and low-competence* (LwLc). The focus on warmth and competence was chosen because of the extensive, applicable work done in [24]—expression of dominance and affiliation are subject to subsequent studies.

First, we selected cues applicable to our set-up. For example, the 'sync' cue was rejected because of limitations with respect to timing of behaviours proposed by the software used (Section V-A), and the NAO robot is incapable of facial expressions. The resulting model is presented in Table I.

Next, we applied our model to the robot by annotating available behaviours, creating a library of animations fitting the style configurations. Lastly, we added fitting behaviours to the text sequence (script). For example, when robot says 'Dus let goed op' (Pay attention), for HwHc the open *StateLeft* behaviour and *head-up* are selected, and for LwLc the closed *CapisceLeft* behaviour and *head-down* (Fig. 2).

V. Study 1: Perceived Warmth and Competence

A. Method

Evaluating the effect of non-verbal behaviour on how children perceive an educational robot's style of interaction, we conducted a 2 × 2 (*warmth* × *comptence*) between-subject perception study at primary schools. For this, a PowerPoint presentation with a text and animation script for four robot styles has been created using the RoboTutor framework (`https://github.com/RoboTutor`). The 10-minute interactive lecture on robotics, given by a NAO robot (`www.alderbaran.com`), included three multiple choice questions, which could be answered using the Turningpoint polling system (`https://www.turningtechnologies.com`). The robot speech was accompanied by non-verbal behaviours, which varied between groups (*HwHc, HwLc, LwHc, LwLc*). Participant's perceptions of the robot's level of *warmth*, *competence*, *dominance*, and *affiliation* were collected after the lecture by rating 20 adjectives on a three-point Likert scale.

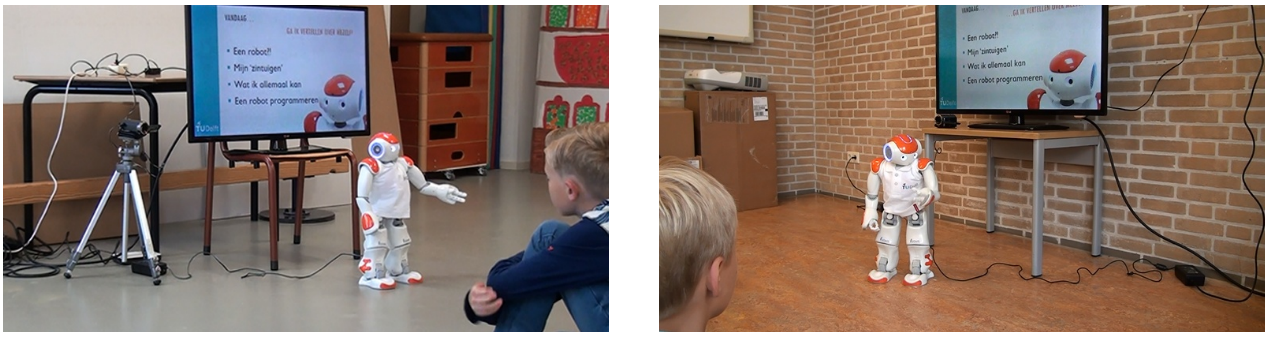
Measurement: Perceived *competence*, *warmth*, *dominance*, and *affiliation* were measured by a non-validated instrument developed for the present study. Children provided their perception of the robot by rating 20 adjectives on a tree-point Likert scale. Perception score for each dependent variable are calculated by multiplication of word-rating with a loading based on expert ratings.

The adjectives (translated from Dutch: *bossy, nagging, clumsy, friend, popular, playful, follower, loner, angry, honest, fights, knowledgeable, boring, nice, listener, confident, educational, helpless, helpful, dumb*) are chosen from words

TABLE I
Overview of the behaviour cues for the four robot style configurations (warmth × competence). ⋆ = competence cues, • = warmth cues, – = behaviour cue for warmth × competence.
*conditions used in second experiment

High Warmth, **High Competence*** *Paralinguistic* • low pitch • low volume *Body Posture* ⋆ stable • directed at audience • head tilt – gaze fixed at audience *Hand Gestures* • open • semantic & syntactic ⋆ frequent (every sentence) – mid-peripheral – centre-centre	**Low Warmth,** **High Competence** *Paralinguistic* • high pitch • high volume *Body Posture* ⋆ stable • pivot away • chin down – gaze fixed at audience *Hand Gestures* • closed • semantic only ⋆ frequent (every sentence) – low-centre – mid-centre-centre
High Warmth, **Low Competence** *Paralinguistic* • low pitch • low volume *Body Posture* ⋆ wobbling • directed at audience • head tilt – gaze fixed at audience *Hand Gestures* • open • semantic & syntactic ⋆ low frequency – low/mid-peripheral	**Low Warmth,** **Low Competence*** *Paralinguistic* • high pitch • high volume *Body Posture* ⋆ non-stable • pivot away • chin down – gaze diversion *Hand Gestures* • closed • semantic only ⋆ low frequency – low/high-centre/peripheral – fiddling

[1] To conserve space, we refer the interested reader to the authors original work for details on the model.



(a) High Warmth, High Competence

(b) Low Warmth, Low Competence

Fig. 2: Stills of behaviour accompanying a statement in two of the four robot style configurations

commonly used to describe various positions in Leary's rose or the SCM, and likely to be present in young children's vocabulary. Children rated whether each word would describe the robot (yes, sometimes/maybe, no), and placed the sticker in the corresponding column of a response leaflet. The words were provided as stickers, because physical activity is suggested to reduce the primacy effect (selecting extreme high or low for all items), thus increasing reliability [29]. Ratings were coded on a numeric interval scale [2–0], where a higher value is associated with better fitting description of the robot. A missing value was coded 99, and replaced by the population mean for that word.

The loadings for individual words on the dependent variables are calculated from expert ratings. Eleven experts in human-computer interaction provided for each adjective a rating [-2–2] on four bipolar scales. First, we assessed the reliability of ratings, words with $SD \geq 1$ are excluded for that construct. Further, words with $Mdn = 0$ are excluded for that construct because the association was weak. From the remaining words (10 for competence, 13 for warmth, 5 for dominance, and 11 for affiliation), we took the mean as loading value. Resulting is a table of 20 adjectives and their loading, if any, on each dependent variable (Table II).

TABLE II
LOADINGS FOR 20 ADJECTIVES (TRANSLATED FROM DUTCH) ON THE DEPENDENT VARIABLES. THE VALUES PRESENT THE MEAN OF THE EXPERT RATINGS, COMPLIANT WITH OUR VARIANCE (≥ 1) AND MEDIAN ($\neq 0$) CRITERIA, FOR THE ADJECTIVES ON FOUR BIPOLAR SCALES.

Adjective	Competence	Warmth	Dominance	Affiliation
Bossy	–	-1.18	2.00	-0.55
Nagging	-0.73	-0.73	–	-1.09
Clumsy	–	–	–	-1.73
Friend	–	1.91	–	1.36
Popular	0.82	0.91	0.55	–
Playful	–	1.36	–	1.82
Follower	–	–	–	–
Loner	–	–	–	–
Angry	–	-1.18	1.09	-1.09
Honest	0.82	1.00	–	1.46
Fight	–	-1.36	1.36	-1.64
Knowledgeable	1.64	–	–	–
Boring	–	-0.64	–	–
Nice	–	1.82	–	1.09
Listener	0.91	1.09	–	1.46
Confident	–	-0.55	1.00	–
Educational	1.64	–	–	0.91
Helpless	-1.45	–	–	–
Helpful	0.64	1.27	–	1.64
Dumb	-1.91	–	–	–

Participants: A total of 101 children, from two primary schools (S1, n=40; S2, n=61) in the Netherlands, participated in our study. Children at S1 are 10-13 years of age ($M = 11.43, SD = 0.64$), and enrolled in 7th (n=9) or 8th (n=31) grade. Children at S2 are aged 5-8 ($M = 6.52, SD = 0.65$), all enrolled in third grade. Gender is evenly distributed (S1 male=20, female=20; S2 male=30, female=28), three children did not report their gender. All participants were naive to the research aim and had little to no previous experience with a NAO robot. Children within each class were assigned to one of four robot style configurations. Children from the same school, in the same condition were merged into one group. This way, at both schools, all groups contained a minimum of 10 children, controlled for age, class and gender.

Procedure: Before the experiment, in the classroom, the researcher was briefly introduced to the children and children were assigned a group. Afterwards, when all children participated, children were given the opportunity to asks questions about the robot and experiment, and a demonstration was given. The following steps were repeated for each group:

- children entered, were seated and given a 'clicker';
- researcher introduced the robot and instructed the children to remain seated after the lecture;
- researcher started the selected script;
- robot gave a lecture displaying stylized behaviours;
- researcher explained the questionnaire, accompanied by a brief example rating the popular Disney figure Simba;
- children spread across the room;
- researcher handed the children stickers with 20 adjectives, followed by a response leaflet and pencil; and
- children provided their individual ratings.

B. Results

We explored differences in children's perceptions of robots displaying high/low warmth and competence related behaviours. Although, K-S tests indicated non-normal distributions, we decided to do parametric tests because we are interested in the interaction effect and the sample size is fair. We performed a MANOVA.

Using Pillai's trace there was a near significant interaction effect of intended warmth × competence on how children perceived the robot, $V = 0.08, F(4, 95) = 2.09, p = 0.088$. Separate univariate ANOVAs on the outcome variables revealed a significant interaction effect for intended warmth × competence on perceived affiliation, $F(1, 97) = 4.42, p = 0.038$, and perceived warmth, $F(1, 97) = 4.09, p = 0.046$. Further, a near significant main effects were found for intended competence on perceived competence, $F(1, 97) = 3.30, p = 0.072$, and intended warmth on perceived dominance, $F(1, 97) = 3.81, p = 0.054$.

In other words, children perceived the robot displaying high-competence behaviours (stable posture, frequent gestures) as more competent ($M = 10.00$) than low-competence behaviours (unstable posture, low frequency of gestures) ($M = 9.20$), regardless of the intended level of warmth (Fig. 3a). And the robot displaying high-warmth behaviours was indeed perceived as slightly more warm ($M = 16.18$) than low-warmth robots ($M = 15.83$) (Fig. 3b). However, there is an interaction effect, the competence model influenced the effects of the warmth model on how children perceived the robot; robots displaying high-warmth behaviours where perceived as more warm than low-warmth robots, only if the robot expressed high-competence. High-competence robots displaying high-warmth behaviours were perceived more competent ($M = 16.58$) than low-warmth robots($M = 15.42$). Low-competence robots displaying high-warmth behaviours were perceived as considerably less warm ($M = 15.76$) than low warmth robots ($M = 16.21$) (Fig. 3c).

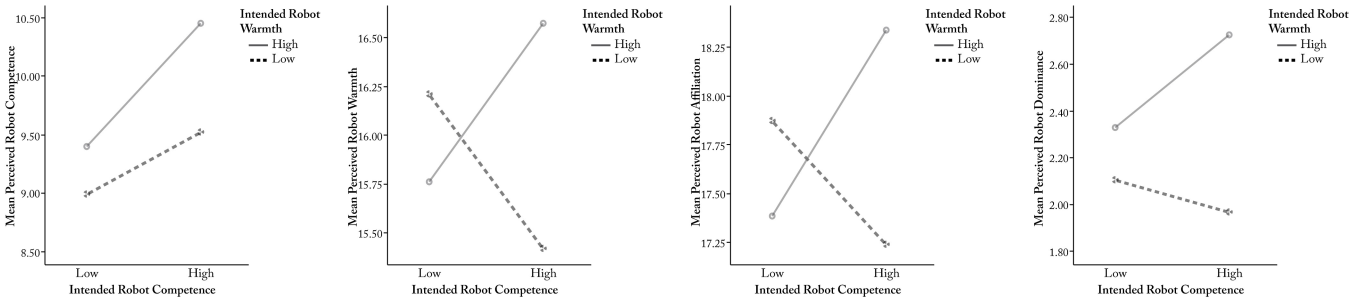
To investigate the bias of individual words on perception scores, we explored significant differences in children's word-ratings. We performed non-parametric Mann-Whitney tests, once for each construct, because normality could not be assumed. Comparison of high and low competence samples, showed a significant difference in ratings for 'Follower' ($U = 913.00, z = -2.49, p = 0.013, r = -0.25$) and 'Helpless' ($U = 798.50, z = -2.69, p = 0.007, r = -0.28$), near significant differences were present for 'Helpful' ($U = 933.00, z = -1.83, p = 0.068, r = -0.19$). No difference in word-ratings have been found for 'Dumb', 'Nice', 'Knowledgeable', and 'Nagging'. Comparison of high and low warmth samples, showed near significant differences for 'Popular' ($U = 1290.00, z = 1.83, p = 0.067, r = 0.19$), 'Loner' ($U = 991.50, z = -1.76, p = 0.078, r = -0.18$), 'Angry' ($U = 1323.00, z = 1.72, p = 0.086, r = 0.17$), 'Confident' ($U = 1305.00, z = 1.65, p = 0.099, r = 0.17$), and 'Helpful' ($U = 1307.00, z = 1.83, p = 0.068, r = 0.19$). Since all children reported exactly the same rating for 'Dumb', again there was no difference between samples.

C. Discussion

The results indicate that children did perceive robots differently based on their non-verbal behaviours—the set-up, PowerPoint, textscript and questions were similar for all groups. However, the differences are small. This can be due to subtlety of the manipulation or measurement.

Variations in non-verbal behaviour were kept subtle, attempting to avoid creating a caricature and keep behaviour believable. However, observers of all four conditions were not able to note any difference, indicating that we have been to careful selecting and adapting animations. Future studies exploring the effect of enlarged differences in behaviour between styles is needed.

The measurement instrument, created based on literature and in corresponds with experts, was partly validated. Word-loadings are based on expert ratings, however, children not necessarily have the same interpretation of words. Analysis of individual word-ratings revealed that children did provide significantly different ratings of words that were not included in calculation of the outcome measures; children ratings of 'Loner' differed significantly between high/low intended warmth, but this word did not contribute to calculation of perceived warmth. Similarly, 'Follower' was significantly different for, but was not included in calculation of, competence. Moreover, 'Dumb' and 'Knowledgeable' are used to calculate perceived competence with notable loadings of resp. -1.91 and 1.64. However, no differences in word-ratings were present. This may indicate a sealing effect; children never think of the robot as being dumb. Or it may result from a priming effect; the robot stated to know a lot about robotics and teach the children about this subject. Discarding 'Knowledgeable' from calculation of competence scores yields a significant difference in perceived competence between



(a) Mean perceived competence.

(b) Mean perceived warmth (range - 11.27 to 18.727).

(c) Mean perceived affiliation.

(d) Mean perceived dominance.

Fig. 3: Study 1 results showing mean perception scores. Competence model displayed on horizontal axis, warmth model by seperate lines.

high/low intended competence, $F(1,97) = 4.41, p = 0.038$. Indicating that validation and improvement of the reliability of our measurement instrument are worthwhile.

Our findings on competence are in line with the suggestions from the literature: stable posture and frequent gestures evoke higher perceptions of competence. Our findings on warmth seemed in line with the literature: low pitch and volume, body directed at audience, head tilt, open (semantic and syntactic) gestures, overall, evoked higher perceptions of warmth. However, perceived warmth declined in the HwLc condition compared to LwLc. Indicating that low-competence behaviours (unstable posture, infrequent gestures) reversed the effect of warmth behaviours. Alternatively, although opposed to the literature, the fiddling and gaze diversion only present in the LwLc condition could be accountable for the effect. It could be that additional head and hand movements make the robot more lively and therefore perceived as more warm. Our findings partially contradict the theory that warmth is evaluated before competence and therefore characters perceived as warm are likely to be found competent. If the robot displayed high-competence behaviours, then the robot displaying high-warmth behaviours was indeed perceived as more competent than low-warmth robots. However, in the low-competence condition, the robot displaying high-warmth behaviours was perceived as less competent than the low-warmth robot. Indicating that warmth expression enhances the competence effect rather than bias towards perceived high competence.

VI. Study 2: Context Dependency

A. Method

To explore the effect of context (i.e., location, content, and usergroup) on how children perceive the robot, we conducted a 2 × 2 (*robot style* × *context*) follow-up study at a camp for children with T1DM. Only two groups could be formed due to practical limitation, limiting us to two robot style configurations (HwHc and LwLc). The procedure and materials were similar to Study 1, except for the lecture topic being usage of the MyPAL app—an app developed to support children coping with and learning about T1DM.

Participants: A total of 72 children participated, 52 from the first study at schools (HwHc=26, LwLc=26), and 20 from camp. We recruited 21 children with T1DM, of which 6 had experienced interacting with a NAO robot before during previous studies. One participant was excluded from further analysis because the questionnaire was not understood and completed. Leaving 20 participants from camp (male=12, female=8), aged 8-11 ($M = 9.20, SD = 1.10$), divided in two groups (HwHc=9, LwLc=11).

B. Results

We explored differences in children's perceptions of the robot between the two contexts (school and camp) and between robot styles (HwHc and LwLc). We performed MANOVA analysis to explore main and interaction effects between robot styles and contexts of use.

Using Pillai's trace, there was a significant main effect for context on perception scores, $V = 0.16, F(4,65) = 3.18, p = 0.019$. Separate univariate ANOVAs on the outcome variables revealed significant effects of context on perceived competence, $F(1,68) = 6.61, p = 0.012$, and affiliation, $F(1,68) = 6.99, p = 0.010$, and near a significant effect on perceived warmth, $F(1,68) = 3.52, p = 0.065$. No significant main effect for robot style on perception scores have been found. However, univariate analysis revealed a near significant effect for robot style on perceived competence, $F(1,68) = 3.03, p = 0.086$. No interaction effects have been found for context × robot style, $V = 0.07, F(4,65) = 1.14, p = 0.345$.

The result shows that the context (location, users, and content) of the activity (interactive presentation by a NAO robot) influenced perceived warmth, affiliation and competence. Overall children with T1DM at camp perceived the robot as less warm ($M = 15.28$), affiliated ($M = 16.61$) and competent ($M = 8.14$) than children at school (warmth $M = 16.39$, affiliation $M = 18.11$, competence $M = 9.72$) (Fig 4a, 4b). Further, we were able to replicate the effect of robot style on perception scores with 20 new participants. Robots displaying high-competence and high-warmth behaviours were perceived more competent ($M = 9.94$) than low-competence and low-warmth robots ($M = 8.65$), independent of the context (Fig. 4c).

C. Discussion

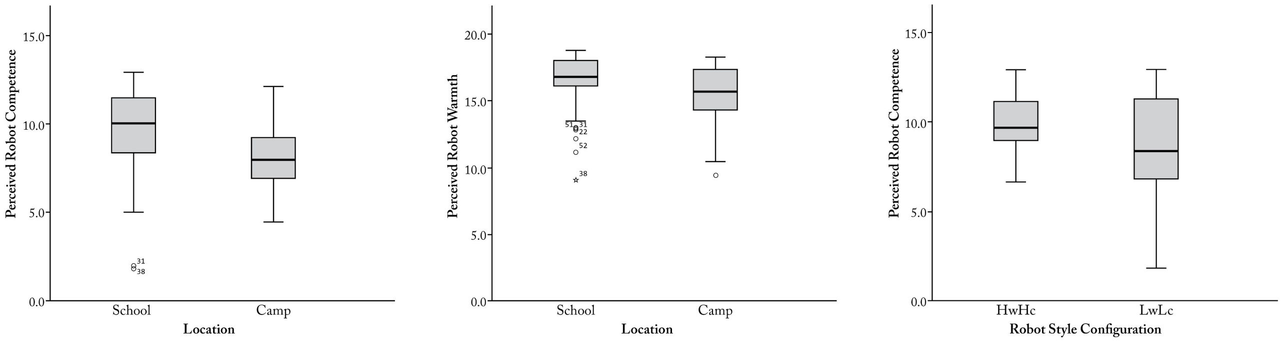
The results indicate that children at camp perceived the robot less warm, affiliated and competent than children at school, independent of the robot style. Changing the users, content or location of an activity can influence how children perceive an educational robot, meaning that results obtained in one context not necessarily translate to another context.

Lower perceived competence may indicate that the setting of Study 1 primed the children in thinking of the robot as competent because it would teach them. Although we expected the children at camp to see the robot as a friendly helper, and therefore perceive it as more warm and affiliated, this was not the case. Possibly, when novelty wears of, children are more rigorous evaluating the robot. Or children who are familiar with the robot compare its behaviour to previous experiences and find the robot less warm in the current activity compared to for example playing a game together. Further research is needed to provide solid conclusions on the relation between context and perception scores.

The lack of difference between perception of robot styles at camp may result from the small samples or too subtle differences in behaviours. Although this was not the main purpose of the second study, it would be interesting to repeat the study and explore the effect of robot style using a larger sample size or more exaggerated behaviour manipulations

VII. Conclusion

We evaluated an educational robot displaying non-verbal behaviours to express high/low warmth and competence with children at primary schools and camp, and showed that even



(a) Perceived competence split by context (range -11.64 to 12.909).

(b) Perceived warmth split by context (range -11.27 to 18.727).

(c) Perceived competence scores split by robot style (range -11.64 to 12.909).

Fig. 4: Study 2 results for perceived warmth and competence.

subtle manipulations in robot behaviour influence children's perceptions of the robot's level of warmth and competence. Although this was a fist attempt, and further research is needed replicating the study with enlarged behaviour manipulations and other social perception factors, to our knowledge this is the first evidence of style in robot teaching.

ACKNOWLEDGMENT

Selection and Participation of Children: We recruited 101 children (aged 5-13, M/F) from primary schools in Delft and Voorschoten, the Netherlands urban area. An invitation to participate in the study was offered to teachers who requested a robot demonstration. 21 children diagnosed with Type 1 Diabetes Mellitus (aged 8-11, 8 female, 13 male) were recruited by the Dutch Diabetes Association DVN. The research was approved by the Human Research Ethics Committee of Delft University of Technology.

REFERENCES

[1] R. Akker, M. Bruijnes, R. Peters, and T. Krikke. Interpersonal stance in police interviews: content analysis. *Computational Linguistics in the Netherlands Journal (CLIN Journal)*, 3:193–216, 2013.

[2] A. Arya, L. N. Jefferies, J. T. Enns, and S. DiPaola. Facial actions as visual cues for personality. *Computer Animation and Virtual Worlds*, 17(3-4):371–382, 2006.

[3] K. Bergmann, F. Eyssel, and S. Kopp. A second chance to make a first impression? how appearance and nonverbal behavior affect perceived warmth and competence of virtual agents over time. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7502 LNAI:126–138, 2012.

[4] D. S. Berry and J. S. Hansen. Personality, Nonverbal Behaviour, and Interaction Quality in Female Dyads. *PSPB*, 26(3):278–292, 2000.

[5] A. Cafaro, N. Glas, and C. Pelachaud. The Effects of Interrupting Behavior on Interpersonal Attitude and Engagement in Dyadic Interactions. In *AAMAS 2016*, pages 911–920. International Foundation for Autonomous Agents and Multiagent Systems, 2016.

[6] L. L. Carli, S. LaFleur, and C. Loeber. Nonverbal behavior, gender, and influence. *Personality and Social Psychology*, 68(6):1030–1041, 1995.

[7] D. R. Carney, J. A. Hall, and L. S. LeBeau. Beliefs about the nonverbal expression of social power. *Journal of Nonverbal Behavior*, 29(2):105–123, 2005.

[8] M. Chollet, M. Ochs, and C. Pelachaud. A multimodal corpus for the study of non-verbal behavior expressing interpersonal stances. In *IVA 2013 Workshop Multimodal Corpora: Beyond Audio and Video*, 2013.

[9] A. J. Cuddy, S. T. Fiske, and P. Glick. Warmth and Competence as Universal Dimensions of Social Perception: The Stereotype Content Model and the BIAS Map. 40(07):61–149, 2008.

[10] A. J. Cuddy, M. Kohut, and J. Neffinger. Connect, Then Lead. *Harvard Business Review*, (July-August), 2013.

[11] A. J. C. Cuddy, P. Glick, and A. Beninger. The dynamics of warmth and competence judgments, and their outcomes in organizations. *Research in Organizational Behavior*, 31:73–98, 2011.

[12] N. E. Dunbar and J. K. Burgoon. Perceptions of power and interactional dominance in interpersonal relationships. *Journal of Social and Personal Relationships*, 22(2):207–233, 2005.

[13] S. T. Fiske, A. J. C. Cuddy, and P. Glick. Universal dimensions of social cognition: warmth and competence. *Trends in Cognitive Sciences*, 11(2):77–83, 2007.

[14] T. Fong, I. Nourbakhsh, and K. Dautenhahn. A survey of socially interactive robots. *Robotics and Autonomous Systems*, 42(3-4):143—-166, 2003.

[15] R. Gifford. The Role of Nonverbal Communication in Interpersonal Relations. In L. Horowitz and S. Strack, editors, *Handbook of interpersonal psychology: Theory, research, assessment, and therapeutic interventions*, chapter 11, pages 171—-190. Wiley, NY, 2011.

[16] A. F. Grasha. A Matter of Style: The Teacher as Expert, Formal Authority, Personal Model, Facilitator, and Delegator. *College Teaching*, 42(4):142—-149, 1994.

[17] L. Guerrero and T. Miller. Associations between Nonverbal Behaviors and Initial Impressions of Instructor Competence and Course Content in Videotaped Distance Education Courses. *Communication Education*, 47(1):30—-42, 1998.

[18] J. a. Hall, E. J. Coats, and L. S. LeBeau. Nonverbal behavior and the vertical dimension of social relations: a meta-analysis. *Psychological bulletin*, 131(6):898–924, 2005.

[19] L. Hall, C. Hume, S. Tazzyman, A. Deshmukh, S. Janarthanam, and H. Hastie. Map Reading with an Empathic Robot Tutor. In *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, number 11th, pages 567—-567, 2016.

[20] J. B. Janssen, C. C. Van Der Wal, M. a. Neerincx, and R. Looije. Motivating children to learn arithmetic with an adaptive robot game. In *Social Robotics*, volume 7072 LNAI, pages 153—-162. Springer, 2011.

[21] T. Leary. Interpersonal diagnosis of personality. *American Journal of Physical Medicine & Rehabilitation*, 37(6):331, 1958.

[22] R. Looije, M. a. Neerincx, and V. D. Lange. Children s responses and opinion on three bots that motivate , educate and play. *Journal of Physical Agents*, 2(2):13–20, 2008.

[23] O. Mubin, C. J. Stevens, S. Shahid, A. A. Mahmud, and J.-j. Dong. A REVIEW OF THE APPLICABILITY OF ROBOTS IN EDUCATION. *Technology for Education and Learing*, 1:209—-0015, 2013.

[24] T.-h. D. Nguyen, E. Carstensdottir, N. Ngo, M. S. El-nasr, M. Gray, D. Isaacowitz, and D. Desteno. Modeling Warmth and Competence in Virtual Characters. In *Intelligent Virtual Agents*, volume 9238, pages 167—-180, Delft, 2015. Springer.

[25] S. Y. Okita, V. Ng-Thow-Hing, and R. Sarvadevabhatla. Learning together: Asimo developing an interactive learning partnership with children. In *RO-MAN 2009-The 18th IEEE International Symposium on Robot and Human Interactive Communication*, pages 1125–1130. IEEE, 2009.

[26] S. Y. Okita, V. Ng-Thow-Hing, and R. K. Sarvadevabhatla. Multimodal

Approach to Affective Human-Robot Interaction Design with Children. *ACM Transactions on Interactive Intelligent Systems*, 1(1):1—-29, 2011.

[27] M. Pantic, R. Cowie, F. D'Errico, D. Heylen, M. Mehu, C. Pelachaud, I. Poggi, M. Schroeder, and A. Vinciarelli. Social Signal Processing: The Research Agenda. In *Visual Analysis of Humans*, chapter 26, pages 511—-538. Springer London, 2011.

[28] R. M. Peters. How turn-taking influences the perception of a suspect in police interviews, 2014.

[29] R. Ros, M. Nalin, R. Wood, P. Baxter, R. Looije, Y. Demiris, T. Belpaeme, A. Giusti, and C. Pozzi. Child-robot interaction in the wild: advice to the aspiring experimenter. In *Proceedings of the 13th international conference on multimodal interfaces*, pages 335–342. ACM, 2011.

[30] M. Saerbeck, T. Schut, C. Bartneck, and M. D. Janse. Expressive robots in education: varying the degree of social supportive behavior of a robotic tutor. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1613–1622. ACM, 2010.

[31] R. Verstegen and H. P. B. Lodewijks. *Interactiewijzer - Analyse en Aanpak van interactieproblemen in professionele opvoedingssituaties.* Uitgeverij van Gorcum, 1999.

[32] A. Vinciarelli, M. Pantic, D. Heylen, C. Pelachaud, I. Poggi, F. D'Errico, and M. Schroeder. Bridging the Gap Between Social Animal and Unsocial Machine: A Survey of Social Signal Processing. *IEEE Transactions on Affective Computing*, 3(1):69—-87, 2011.

[33] J. S. Wiggins. An informal history of the interpersonal circumplex tradition. *Journal of personality assessment*, 66(2):217–233, 1996.

[34] J. Xu, J. Broekens, K. Hindriks, and M. Neerincx. Effects of Bodily Mood Expression of a Robotic Teacher on Students. In *International Conference on Intelligent Robots and Systems (IROS)*, pages 2614–2620, 2014.

[35] C. Zaga, K. Truong, M. Lohse, and V. Evers. Exploring child-robot engagement in a collaborative task. In *Proceedings of the Child-Robot Interaction Workshop: Social Bonding, Learning and Ethics*, page 3, Lisbon, Portugal, 2014. Instituto de Engenharia de Sistemas e Computadores, Investigação e Desenvolvimento em Lisboa (INESC-ID).

CAAF: A Cognitive Affective Agent Programming Framework

F. Kaptein, J. Broekens, K. V. Hindriks, and M. Neerincx

Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands,
F.C.A.Kaptein@tudelft.nl

Abstract. Cognitive agent programming frameworks facilitate the development of intelligent virtual agents. By adding a computational model of emotion to such a framework, one can program agents capable of using and reasoning over emotions. Computational models of emotion are generally based on cognitive appraisal theory; however, these theories introduce a large set of appraisal processes, which are not specified in enough detail for unambiguous implementation in cognitive agent programming frameworks. We present CAAF (Cognitive Affective Agent programming Framework), a framework based on the belief-desire theory of emotions (BDTE), that enables the computation of emotions for cognitive agents (i.e., making them cognitive *affective* agents). In this paper we bridge the remaining gap between BDTE and cognitive agent programming frameworks. We conclude that CAAF models consistent, domain independent emotions for cognitive agent programming.

Keywords: models of emotionally communicative behavior · theoretical foundations and formal models · dimensons of intelligence, cognition and behavior

1 Introduction

Interaction with intelligent virtual agents is facilitated by providing such agents with affective abilities. For example, affective abilities in intelligent agents have been applied to facilitate *entertainment* [17, 23], to make an agent more *likable* for the user [3], to get *empathic* reactions from the user [7], and to create the so-called *the illusion of life* [2, 18], where characters are modelled to appear more life-like.

Cognitive agents can be programmed in frameworks like, e.g., GOAL [11], Jadex [16], or Jason [4]. A cognitive agent is an autonomous agent that perceives its environment through sensors and acts upon that environment with actuators [24]. It does so based on its *beliefs*, *desires* and *intentions*. Cognitive agents have a *mental state* and a *reasoning cycle* (see Figure 1). The mental state consists of *beliefs* and *desires*. Beliefs are the agent's representation of its environment. The agent can believe it is walking down the street, or that it is raining outside. Desires are things the agent *wants* to be true. For example, the agent can want to have an umbrella. The *intention* to get an umbrella reflects the agent's commitment to achieve that desire. After sensing *percepts* from the environment, the agent updates its mental state. Based on its beliefs, desires, and intentions, the agent reasons about its next action. The environment can change by itself, in response to an action of the agent, or actions from other agents that are situated in the same environment; thus, the agent may not always be *certain* of the exact *state of affairs* in its environment.

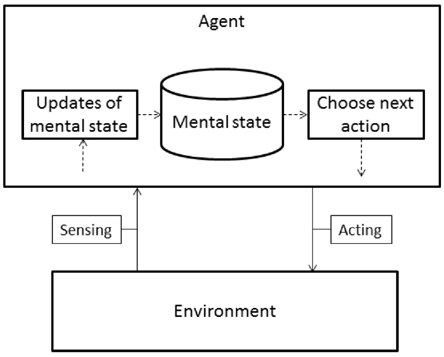


Fig. 1. The reasoning cycle of a cognitive agent.

By adding a computational model of emotion to cognitive agent programming frameworks, one can program intelligent agents capable of using and reasoning over emotions. Computational models of emotion are usually based on cognitive appraisal theories [13]. Cognitive appraisal theory proposes that emotions are consequences of cognitive evaluations (*appraisals*), relating the event to an individual's desires. For example, one is happy because one believes something to be true, and desires this to be true.

However, cognitive appraisal theories [12, 15, 25] typically introduce a large set of appraisal processes, which are not specified in enough detail for unambiguous implementation in cognitive agent programming frameworks. Psychological theories are developed to explain emotions for humans. These theories are thus not obligated to provide worked out computational specifications for the appraisals.

Here we address this problem by integrating a computational model of the belief-desire theory of emotions (BDTE) [19, 20] with a BDI (belief-desire-intention)-based, cognitive agent programming framework. We present CAAF, a Cognitive Affective Agent programming Framework. Emotions are computed based on BDTE for two reasons: 1) because it is conceptually close to the BDI agent framework; and 2) it does not introduce a large set of appraisals that are difficult to describe in a computational manner.

The two main contributions of this work are: 1) We define semantics for the programming constructs of cognitive agents, formalizing how an agent updates its *mental state*, and how emotions are computed. 2) We show when the agent should minimally (re)appraise, by proving that, under some circumstances, the computation of emotions stays consistent when reducing the frequency with which the agent's emotions are recomputed, thereby increasing the efficiency of the computation.

2 Motivation & Related Work

In this article we focus on computational models of emotion based on cognitive appraisal theory. A computational model of emotion describes the eliciting conditions for emotions, often including corresponding intensity. A popular appraisal theory among computer scientists, is the OCC-model [1, 15, 27]. The appraisal theory by Lazarus [12], and the sequential check theory (SCT) by Scherer [25, 26] have also found some attention among computer scientists. For example, the computational model EMA [10, 14] is mainly based on the appraisal theory by Lazarus [12], where the link between appraisal and coping is emphasized. EMA models how emotions develop and influence each other. For example, sadness can turn into anger at the responsible source. In [5] a formal notation for the declarative semantics of the structure of appraisal is proposed. Using this, a computational model of emotion is developed based on SCT.

The OCC model is the most implemented cognitive appraisal theory. Computational models based on the OCC model include AR [9], EM [18], FLAME[8], FearNot! [7], FAtiMA [6], and GAMYGDALA [17]. In AR [9] agents judge events based on their pleasantness, and whether they are confirmed, unconfirmed, or disconfirmed. For example, sadness is achieved when an agent confirms an unpleasant event. In EM [18] the aim is to build 'believable agents', agents that appear emotional and engage in social interactions. The EM architecture facilitates artists to model emotional agents in their applications. In FLAME the desirability of an event is modelled with fuzzy sets. For example, they define a fuzzy set 'undesirable event'. Individual events are then partly a member of this set, the amount of membership is adaptively learned over time. FearNot! is an application that helps children to cope with bullying. The agents use planning and expected utility to derive proper emotional responses. Currently the emotional responses in FearNot are triggered with a more enhanced model FAtiMA. FAtiMA divides the appraisal into different modules, all responsible for a separate part of the computation. This enables implementing such modules independently. GAMYGDALA is an emotion engine that can be added to games by annotating events with their influence on the beliefs and desires of different characters.

An underlying problem with many appraisal theories is that cognitive agent programming frameworks lack the required knowledge representations to compute most appraisal processes. For example, a computational model of emotion that aims to describe the OCC-model in total [15], including emotion intensities, needs to model 12 different appraisals. For many of these appraisals it is unclear *how* they should be implemented, e.g., *deservingness*, *sense of reality*, or *proximity*. Other appraisals, e.g., praiseworthiness, require complex constructs

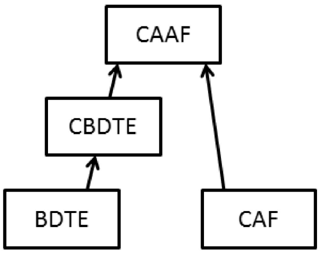


Fig. 2. CAAF is build upon CBDTE [20] and CAFs (Cognitive Agent programming Frameworks). With CAAF, we close the gap between CBDTE and CAFs, and provide a fully worked out, computational account of BDTE.

like norms and values to be represented by the agent. SCT [25, 26] additionally introduces multiple layers in the appraisal process. An event is first analysed in a reactive, bodily responsive, type of way, and later analysed with increasingly nuanced cognitive processes. The computational model of emotion, EMA [14], is mainly based on the appraisal theory by Lazarus [12]. EMA [14] aims to simplify the appraisal processes, introduced by the underlying appraisal theories, and models them from a knowledge representation consisting of beliefs, desires, intentions, and (decision-theoretic) plans. This is conceptually closer to cognitive agent programming frameworks; however, though these frameworks are suited for programming decision-theoretic plans, they do not always do so. This would thus put constraints on the agent programming frameworks for which we want to compute emotions.

The appraisals and knowledge representation proposed by the belief-desire theory of emotion (BDTE) [19, 20] are more compatible with cognitive agent programming frameworks. In BDTE, emotions are derived only from beliefs and desires. In its minimal form BDTE requires only two appraisals. This makes BDTE more suitable as a basis for simulated emotions for such frameworks.

In this paper we integrate a computational model of BDTE with a cognitive agent programming framework (CAF), hence developing CAAF. In [20], Reisenzein extended BDTE to a computational form (CBDTE). CBDTE has been referred to as a computational model of emotion [13]; however, Reisenzein acknowledges that the motivation behind developing CBDTE was not to develop a worked-out computational model, but rather to clarify aspects of BDTE [20]. Here, we build upon CBDTE, and close the gap between CAFs and CBDTE (see figure 2). Thus, this paper presents a *full* computational account of BDTE, and formalizes how a cognitive agent should (efficiently) compute emotions.

3 A Model of Emotion for Cognitive Agent Programming Frameworks

In this Section we present CAAF. We present the formal semantics needed to integrate BDTE with cognitive agent programming. Further, based on this formal system we show in Section 4 that emotions can be computed in an efficient way using the model presented here.

3.1 Semantics for a Basic Knowledge Representation & BDTE

The mental state of an agent requires a *knowledge representation.* The agent needs to *represent* states of affairs, to *store* these representations, and to *change* the stored representations.

Representing the states of affairs is achieved with a *language*. This language needs to define a syntax of *well-formed formulae.* We write $\varphi \in \mathcal{L}$ to denote that φ is a formula of language $\mathcal{L}$. Here, a formula is a single proposition that contains information about a *state of affairs*, i.e., it is a sentence that *expresses whether a state of affairs is true (or not).* We do not define how *logical connectives* work in this language, i.e., symbols that connect propositions such that the sense of the compound proposition depends only on the original sentences (for example, φ_1 *and* φ_2). The contribution of this paper is to define semantics for the programming constructs of cognitive agents, formalizing how an agent updates its *mental state*, and how emotions are computed.

Storing states of affairs is done with a *set.* The belief, desire and emotion base are represented in the semantics as a set of formulae, mapped to a value $[0,1]$. These bases are a subset of some language $\mathcal{L}$, but contain further information as well. A belief base has the form: $\Sigma : \langle C : \mathcal{L} \rightarrow [0,1] \rangle$, where C is mapping of a formula φ to (exactly one) certainty value between $[0,1]$. We denote $b\{\varphi \rightarrow c\} \in \Sigma$ for 'the agent believes φ with certainty c'. Furthermore, we add the constraint that if C contains the mappings $b\{\varphi \rightarrow c\}$ and $b\{\neg\varphi \rightarrow c'\}$, then $c = 1 - c'$. A desire base has the form $\Gamma : \langle U : \mathcal{L} \rightarrow [0,1] \rangle$, where U is mapping that maps formula φ to a utility value between $[0,1]$. We denote $d\{\varphi \rightarrow c\} \in \Gamma$ for 'the agent desires φ with utility (strength of desire) u'. Finally an emotion base has the form $\Upsilon : \langle I : \mathcal{L} \times \Theta \rightarrow [0,1] \rangle$, where $\theta \in \Theta$ is an emotion label (happy, unhappy, hope, fear, surprise, relieve, or disappointment), and I maps formula $\varphi \in \mathcal{L}$ and label $\theta \in \Theta$ to an intensity value between $[0,1]$. We denote $e\{\varphi \times \theta \rightarrow i\} \in \Upsilon$ for 'the agent has emotion θ (concerning formula φ) with intensity i'. Note that traditional boolean propositional logic (where formulae are either true or false, rather than mapped to a value between $[0,1]$) would be sufficient for programming cognitive (BDI-based) agents [11]. However, for the computation of many emotions in BDTE we need values between $[0,1]$. For example, an agent that applies for a new job cannot feel hope (according to BDTE) when it only knows if it got the job afterwards. It should reason over the certainty of getting this job. For example, after having a good job interview. Also note that the emotions in Υ contain a formula, rather than just a label and intensity. With this we model the apparent directedness of emotions, in line with BDTE [20]. One is happy *about* some formula, e.g., $\varphi =$ 'I will get a new job'.

Changing the knowledge representation is denoted with a combine operator $\oplus$. Given some set S and some set T containing a number of formulae, $S \oplus T$ denotes an update of S with T. $\oplus$ is a simple set join, with elements in set T taking priority over elements in set S, to allow updating of c, u and i in S. For all formulae $\varphi \in S$ and $\varphi \in T$, the mapping $\varphi \to n$ in the resulting set is taken from the set T. Thus, $\oplus$ is not symmetric, i.e., $S \oplus T \neq T \oplus S$.

Definition 1. (Combine $\oplus$)
Given some sets S, and T, which contain a number of elements $e = \{\varphi \to n\}$, where φ is a formula $\varphi \in \mathcal{L}$, and n a value $n \in [0,1]$. $S \oplus T$ is defined as follows:

$$e \in S \oplus T \qquad \textit{iff} \quad e \in T, \textit{ or } (e \in S \textit{ and } e \notin T)$$

A knowledge representation is a pair $\langle \mathcal{L}, \oplus \rangle$, where $\mathcal{L}$ is a language to represent states of affairs, and $\oplus$ defines how a set of formulae is updated with another set of formula. Using our definition of a knowledge representation, we can now formally define what a *mental state* of an agent is. We call this initial definition a 'Simple Mental State' because we will expand it later in the paper.

Definition 2. (Simple Mental State)
A mental state is a pair $\langle \Sigma, \Gamma \rangle$ where Σ is called a belief base, and Γ is a desire base.

The aim of the work presented here is to add *emotional reasoning* to these agent programming frameworks. The belief-desire theory of emotion (BDTE) [19, 20] provides a method for computing emotional responses based solely on ones beliefs and desires. For BDTE we need only the beliefs and desires, before and after an agent's update of its mental state. We could imagine that a computation of an agent program is a sequence of mental states $m_0, m_1, m_2, \ldots$. BDTE then enables the computation of an agent's emotions in a mental state m_i by using the belief- and desire base corresponding to mental states m_{i-1} and m_i. Based on BDTE we can define the inner workings of this function [20].

Definition 3 describes BDTE in a computational manner. This is based on CBDTE [20]. In function $R(\Sigma, \Sigma', \Gamma, \Gamma') \to \Upsilon$ (R for Reisenzein's appraisal [20]), we denote Σ as the belief base of mental state m_{i-1}, Γ as the desire base of mental state m_{i-1}, Σ' as the belief base in mental state m_i, and Γ' as the desire base of mental state m_i. The function $R(\Sigma, \Sigma', \Gamma, \Gamma')$ computes all new emotions resulting from changes in the mental state.

Definition 3. (BDTE R)
Given function $R(\Sigma, \Sigma', \Gamma, \Gamma') \to \Upsilon$. Let S be the set containing all φ such that $b\{\varphi \to c\} \in \Sigma$, $b\{\varphi \to c'\} \in \Sigma'$, $d\{\varphi \to u\} \in \Gamma$, and $d\{\varphi \to u'\} \in \Gamma'$, with $c \neq c'$, or $u \neq u'$. $S = \{\varphi_1, \ldots, \varphi_n\}$. If we iterate through S with $i = 1..n$, add the following emotions as follows: $\Upsilon = E_1 \oplus E_2 \oplus \ldots \oplus E_n$, such that:

$$\begin{array}{lll}
e\{\varphi_i \times \textit{happy} \to u\} \in E_i & \textit{iff} & c' = 1 \;\&\; u > 0 \\
e\{\varphi_i \times \textit{unhappy} \to u\} \in E_i & \textit{iff} & c' = 0 \;\&\; u > 0 \\
e\{\varphi_i \times \textit{hope} \to c' \times u\} \in E_i & \textit{iff} & 0 < c' < 1 \;\&\; u > 0 \\
e\{\varphi_i \times \textit{fear} \to (1 - c') \times u\} \in E_i & \textit{iff} & 0 < c' < 1 \;\&\; u > 0 \\
e\{\varphi_i \times \textit{surprise} \to 1 - c\} \in E_i & \textit{iff} & c' = 1 \\
e\{\varphi_i \times \textit{surprise} \to c\} \in E_i & \textit{iff} & c' = 0 \\
e\{\varphi_i \times \textit{relief} \to 1 - c\} \in E_i & \textit{iff} & c' = 1 \;\&\; u > 0 \\
e\{\varphi_i \times \textit{disappointment} \to c\} \in E_i & \textit{iff} & c' = 0 \;\&\; u > 0
\end{array}$$

For example, let $\varphi_1 =$ '*I got a new job*', $b\{\varphi_1 \to 1\} \in \Sigma'$ (i.e., the agent beliefs to have gotten a new job), and $d\{\varphi_1 \to 0.9\} \in \Gamma$ (i.e., the agent strongly desires to have gotten a new job), then Definition 3 prescribes $e\{\varphi_1 \times \mathit{happy} \to 0.9\} \in \Upsilon$ (i.e., the agent is very happy that it got a new job).

With these definitions we already have a framework to implement emotions, which basically works as proposed in previous work [20]. We might imagine that the computation of an agent program results in a sequence of mental states $m_0, m_1, m_2, \ldots$. Computing emotions can then be done by computing Υ over two consecutive mental states. However, this approach does not take into account that emotion intensities decay over time, how to deal with multiple appraisals of the same emotion label (θ), or the fact that you might want to store emotions for reasoning purposes. Furthermore, computation based on BDTE gives a large set containing multiple emotions for every formula φ the agent has in its mental state, meaning we need a method to abstract useful information from it.

3.2 Closing the Semantic Gap between BDTE and BDI

In this Section we expand the model such that BDTE can be used for agent programming in an efficient way, including decay, repeated appraisals, and querying the emotions. We start with expanding the mental state of an agent with an emotion base. With this we can store the current emotional state of an agent, and query this when needed.

Definition 4. (Mental State)
A mental state is a triple $\langle \Sigma, \Gamma, \Upsilon \rangle$ where Σ is called a belief base, Γ is a desire base, and Υ is an emotion base.

With an emotion base storing the emotional responses we can now define a function that gradually decays the intensities of the stored emotions. Function $d(\Upsilon, \Delta t)$ is responsible for decaying the emotional state Υ over time Δt. For the consistency of our model (see Section 4) we define Δt to be zero within one reasoning cycle of an agent. Between reasoning cycles, Δt is a function over the actual system time passed between the start of the previous and current reasoning cycle. Function decay is a mapping $d : \Upsilon \to \Upsilon'$, that decreases the intensity $i \in [0, 1]$ for all elements $e\{\varphi \times \theta \to i\} \in \Upsilon$.

Definition 5. (Decay Function d)
Let $e\{\varphi \times \theta \to i\} \in \Upsilon$. d is a function $d(\Upsilon, \Delta t) \to \Upsilon'$ defined as:

$$e\{\varphi \times \theta \to f(\theta, i, \Delta t)\} \in d(\Upsilon, \Delta t) \quad \textit{iff} \quad e\{\varphi \times \theta \to i\} \in \Upsilon$$

Where $f(\theta, i, \Delta t)$ is a function that decreases the intensity i, and for all emotions $e \in \Upsilon$ the emotion also exists in Υ' with a decayed intensity. The function can be initialized differently for every emotion label $\theta \in \Theta$. An example of exponential decay for happy would be: $f(\mathit{happy}, i, \Delta t) = i - i \times \Delta t$.

We adopt the view in [18] that decay may need different instantiations for different emotions, depending on the corresponding emotion label $\theta \in \Theta$. For example, hope and fear may decay slower than surprise. In our model an agent programmer can adjust the default decay function, for every emotion label independently.

The above defined functions come together in (i.e., are sub-functions of) function **EM**. This function is a mapping: $\mathbf{EM}(\Sigma \times \Sigma \times \Gamma \times \Gamma \times \Upsilon) \rightarrow \Upsilon$.

Definition 6. (Emotion Base Transformer **EM**)
Let Σ, Γ, and Υ be a belief base, desire base, and emotion base in some mental state m. Further, let Σ', and $dbase'$ be the belief base and desire base after some update on this mental state. Function $\boldsymbol{EM}(\Sigma \times \Sigma' \times \Gamma \times \Gamma' \times \Upsilon) \rightarrow \Upsilon'$ computes the emotion base in this updated mental state as follows:

$$\Upsilon' = d(\Upsilon, \Delta t) \oplus R(\Sigma, \Sigma', \Gamma, \Gamma')$$

This function is called when the belief base or desire base of an agent change. This happens through *updates*. There is a set of build-in updates that act on the mental state bases of the agent. Updates change the belief and desire bases of the agent. Whilst performing these updates, the agent will automatically add emotions to its emotion base Υ.

Definition 7. (Mental State Transformer $\mathcal{M}$)
Let $\varphi \in \mathcal{L}$, and $n \in [0, 1]$. The mental state transformer function $\mathcal{M}(update, m) \rightarrow m'$ *is a mapping from built-in updates (update = [insert, adopt, drop]) and mental states $m = \langle \Sigma, \Gamma, \Upsilon \rangle$ to mental states as follows:*

$$\begin{array}{ll} \mathcal{M}(\boldsymbol{insert}(\varphi, n), m) & = \langle \Sigma \oplus \{\varphi \rightarrow n\}, \Gamma, \Upsilon' \rangle \\ \mathcal{M}(\boldsymbol{adopt}(\varphi, n), m) & = \langle \Sigma, \Gamma \oplus \{\varphi \rightarrow n\}, \Upsilon' \rangle \\ \mathcal{M}(\boldsymbol{drop}(\varphi), m) & = \langle \Sigma, \Gamma \oplus \{\varphi \rightarrow 0\}, \Upsilon' \rangle \end{array}$$

with $\Upsilon' = \boldsymbol{EM}(\Sigma, \Sigma', \Gamma, \Gamma', \Upsilon)$, where Σ' is the belief base, and Γ' is the desire base in the resulting mental state m'.

Mental state bases are defined as sets, thus, if a previous mapping $\{\varphi \rightarrow n\}$ exists in the mental state, then the updates defined above overwrite the previous mapping. In BDTE the claim is made that emotions are subconscious meta-representations of ones beliefs and desires [20]. In the definition above, we model this with function **EM**, which automatically updates the emotions when updating the beliefs, and desires in the mental state.

Definition 8. (Transition rule)
Let m be a mental state, and `u` *be an update ([insert, adopt, drop]) performed in mental state m. The transition relation $\xrightarrow{\mathtt{u}}$ is the smallest relation induced by the following transition rule.*

$$\frac{\mathcal{M}(\mathtt{u}, m) \textit{ is defined}}{m \xrightarrow{\mathtt{u}} \mathcal{M}(\mathtt{u}, m)}$$

The execution of an agent as explicated above, results in a *computation.* A computation in this context is a list of mental states and corresponding updates, performed by the agent. The new mental state is derived from the transition rule in Definition 8. The agent chooses its next update from the set of possible updates in the current state, this set is filled through the rules defined by the programmer. The computation starts in the initial mental state of the agent.

Definition 9. (Mental Computation)
A mental computation is a sequence of mental states $m_0, \mathtt{u}_0, m_1, \mathtt{u}_1, m_2, \mathtt{u}_2, \ldots$ *such that for each* i *we have that* $m_i \xrightarrow{\mathtt{u}_i} m_{i+1}$ *can be derived using the transition rule of Definition 8.*

The emotion update function **EM** is triggered as part of the Mental State Transformer (Definition 7). It is a part of the mapping from $m_i \xrightarrow{\mathtt{u}_i} m_{i+1}$. Emotions are thus computed after every mental state change of an agent.

Figure 1 showed the reasoning cycle of an agent. The mental computation, defined in Definition 9, operates solely in the 'updates of mental state' box. This means that in the model presented here, an agent senses its environment and starts updating its mental state based on these observations. With these mental state updates, we now defined how emotions are automatically changed accordingly. After updating its mental state, the agent can choose a new action to perform in the environment, which in turn changes the environment. The agent then again senses the changes in the environment, and the cycle starts anew.

3.3 Querying the Emotion Base

Querying the emotion base of an agent is useful. For example, if one wants to know if the agent is happy then one should inspect the emotion base for formulae about which the agent is happy. However, a computation based on BDTE gives a large set containing multiple emotions for every formula φ the agent has in its mental state. We therefore need a function that abstracts over these formulae.

To model this, we define an overall *affective state*, which summarizes the agent's emotions. We compute this affective state with function A. This function computes abstractions from the emotion base that enable a programmer to, for example, query the overall happiness of an agent. It summarizes the emotions in some emotion base Υ. It does so by taking all formulae in the emotion base Υ, for all emotion labels $\theta \in \Theta$, and computing a single intensity from these emotions in Υ concerning the emotion label θ.

Besides the computational argumentation there is also a psychological argumentation to define the affective state. In [21] Reisenzein argues that emotions have a hedonic tone, different than that of beliefs and desires. It *feels* a certain way to have an emotion, which is essentially different from how a belief or desire feels. In his own words: "To account for the hedonic tone of emotions in BDTE, one must assume that 'emotional' belief-desire configurations cause a separate mental state that carries the hedonic tone. [21]" By means of an affective state we model this hedonic tone of emotions.

Definition 10. (Affective State Ω)
Ω is a function, that computes a generalized affective state which summarizes the emotions $e\{\varphi \times \theta \rightarrow i\} \in \Upsilon$ for some emotion label $\theta \in \Theta$.

$$\Omega(\theta, \Upsilon) = \log_2(\sum_{e\{\varphi\times\theta\rightarrow i\}\in\Upsilon} 2^{i\times 10})/10$$

In our model we have implemented $\Omega(\theta, \Upsilon)$ with a logarithmic function (Log_2 $(\sum 2^{i\times 10})/10$), where we sum over all emotions $e\{\varphi \times \theta \rightarrow i\} \in \Upsilon$ corresponding to label θ. Other possible functions might be normal combine: $i' = I/(I + 1)$, with I the summation of all intensities concerning θ), or a simple MAX function (taking the highest intensity emotion corresponding to θ.

From these functions the logarithmic is computationally speaking slightly less efficient; however, the function forces the resulting intensity to be as least as large as the highest value, but takes other values into account. For example, happiness about three different propositions: φ_1 = 'Getting a new job', φ_2 = 'Buying a new car', and φ_3 = 'Going out for dinner', with corresponding intensities: $[0.7, 0.6, 0.3]$, will compute to an overall happiness of 0.76 with logarithmic combine, to 0.62 with normal combine, and to 0.7 with the MAX function.

We do not claim that this is the only correct way to compute the overall affective state, but rather that an agent programmer *requires* a summary to efficiently query the emotion base, and that the here proposed approach will thus help the programmer.

4 Proof of Consistency when Minimizing the (Re)Appraisal of Emotions

In Section 3 we defined the (re)computation of an agent's emotions to occur after every mental state update. However, this is not a computationally optimal approach. In this Section we show how one can optimize this by showing when an agent should minimally (re)compute its emotions (i.e., when the agent should *(re)appraise*).

There are three conditions that should trigger a reappraisal: 1, An agent should reappraise before querying its emotion base, if it has updated its mental state since the last reappraisal, since otherwise it would query an outdated emotional state. 2, An agent should reappraise before a mental state update if the last reappraisal was in a previous reasoning cycle, otherwise the emotions are not correctly decayed. 3, An agent should reappraise when it performs a mental state update on a formula that had already been updated after the last reappraisal, otherwise the previous update will be lost. Since 1 and 2 directly follow from the formal semantics, we need only to show that 3 is true. We do so by proving that if we assume that updates refer to different formulae, appraisal can be postponed to the last update. From this one can infer point 3.

Theorem 1. *Consistency For Delayed Appraisal*
Let $\mathtt{u}_1, \mathtt{u}_2, .., \mathtt{u}_n$ be different mental state updates, with $\varphi_1, \varphi_2, \ldots, \varphi_n$ the formulae these updates refer to respectively. Furthermore, let $\mathtt{u}'_1, \mathtt{u}'_2, .., \mathtt{u}'_n$ be the same

mental state updates; however, for these mental state updates we define the Mental State Transformer (Definition 7) to delay updating the emotion base until $\mathtt{u}'_n$*. Furthermore let* $\varphi_1 \neq \varphi_2 \neq \ldots \neq \varphi_n$*. Consider the following two possible reasoning cycles:*

$$\begin{aligned} rc_1 : \quad & m_0 \xrightarrow{\mathtt{u}_1} m_1 \xrightarrow{\mathtt{u}_2} \ldots \xrightarrow{\mathtt{u}_n} m_n \\ rc_2 : \quad & m_0 \xrightarrow{\mathtt{u}'_1} m'_1 \xrightarrow{\mathtt{u}'_2} \ldots \xrightarrow{\mathtt{u}'_n} m'_n \end{aligned}$$

where rc_2 *delays updating the emotion base until update* $\mathtt{u}'_n$*. Under the constraint that* $\varphi_1 \neq \varphi_2 \neq \ldots \neq \varphi_n$*, we can derive that* $m_n = m'_n$*.*

To show the truth of this claim, let the knowledge bases corresponding to mental state m_i be denoted with, $m_i = \langle \varSigma_i, \varGamma_i, \varUpsilon_i \rangle$. Since $\varSigma$ and $\varGamma$ are updated normally we need only to show that $\varUpsilon_n = \varUpsilon'_n$. To this end, we first need to define a property of the definitions. We defined $\varDelta t$ in function d (decay) to be zero within one reasoning cycle. Furthermore, $d(\varUpsilon, 0) = \varUpsilon$. Due to this, we can ignore decay when comparing reasoning cycles rc_1 and rc_2. If we denote E_i to be the set of emotions resulting from function R in transition $m_{i-1} \xrightarrow{\mathtt{u}_i} m_i$, then we can write:

$$\begin{aligned} \varUpsilon_1 &= d(\varUpsilon_0, 0) \oplus E_1 \\ &= \varUpsilon_0 \oplus E_1 \\ \varUpsilon_2 &= d(\varUpsilon_0 \oplus E_1, 0) \oplus E_2 \\ &= \varUpsilon_0 \oplus E_1 \oplus E_2 \\ \varUpsilon_n &= \varUpsilon_0 \oplus E_1 \oplus E_2 \oplus \ldots \oplus E_n. \end{aligned}$$

The emotion base resulting from reasoning cycle 2 can be found with the same definitions. Since the update of the emotion base is delayed, the emotion base $\varUpsilon'_{n-1} = \varUpsilon_0$. Furthermore, the computation of new emotions (Definition 3) will consider all updated formulae:

$$\begin{aligned} \varUpsilon'_n &= d(\varUpsilon_0, 0) \oplus \{E_1 \oplus E_2 \oplus \ldots \oplus E_n\} \\ &= \varUpsilon_0 \oplus E_1 \oplus E_2 \oplus \ldots \oplus E_n. \end{aligned}$$

If $\varphi_1 \neq \varphi_2 \neq \ldots \neq \varphi_n$, then the emotions in sets $E_1, \ldots, E_n$ do not overwrite each other when added to the emotion bases. Therefore, we can conclude that $\varUpsilon_n = \varUpsilon'_n$. Together we can now also conclude $m_n = m'_n$.

5 Discussion

In this section we discuss some drawbacks of using BDTE as psychological background. BDTE models a limited range of emotions compared to other theories (BDTE models 7 emotions, while, for example, OCC models over 20 different emotions). Should an agent programmer want to use the emotions in the agent's decision making, then a smaller set of emotions might be more conceivable; however, there can also be domains in which the set of emotions modelled by BDTE is too limited. For example, when a programmer needs the agent to properly reason over empathic emotions like gratitude and remorse, then BDTE is inadequate in its current form.

Future work could thus complement this framework by modelling social emotions. In [22], Reisenzein discusses possible extensions of BDTE to take social emotions into account. For example, he proposes introducing *altruistic desires*. For example, pity is then explained as a form of displeasure following from the frustration of an *altruistic desire* (desiring something good for someone else). However, this does not provide explanations for all social emotions (e.g., anger). When adding social emotions, one might need to complement the presented framework with additional concepts such as norms.

6 Conclusion

In this paper we presented CAAF (a Cognitive Affective Agent programming Framework), a framework where emotions are computed automatically when agents update their mental states. We presented semantics showing the programming constructs of these agents in a domain-independent manner. With these constructs, a programmer can build an agent program with cognitive agents that automatically compute emotions during runs. We chose BDTE to compute new emotions because it is conceptually close to the BDI architecture and therefore allowed us to embed emotions without introducing many additional concepts in the mental states of the agents.

Our semantics facilitate incremental work. For example, if it is desirable to change the affective state (Definition 10) with a global mood, then one could change the function that computes the affective state (function A), without being forced to adjust the entire framework. One might also want to enable programmers to adjust the emotion base without changing the belief base. Definition 7 defined functions to update the agent's mental state. We could simply complement this definition to contain function *Appraise*, capable of inserting emotions in the emotion base (Υ), similar to the update *insert* for the belief base (Σ). This fits well in the modular approach suggested by Marsella et. al. [13], where models can implement parts of a complete cycle of emotional reasoning. For example, one could add a module capable of using emotions to guide the agent's decision making (e.g., what action to perform in the environment, or when to decrease the utility of a desire as a type of coping behaviour). The framework presented in this paper thus provides a modular, domain-independent, and consistent implementation for the computation of emotions for cognitive agent programming frameworks, thus facilitating the development of intelligent virtual agents with affective abilities.

Acknowledgements

This research is done for the PAL (a Personal Assistant for a healthy Lifestyle)-project. PAL is funded by Horizon2020 grant nr. 643783-RIA.

References

[1] C. Adam, A. Herzig, and D. Longin. A logical formalization of the occ theory of emotions. *Synthese*, 168(2):201–248, 2009.

[2] J. Bates et al. The role of emotion in believable agents. *Communications of the ACM*, 37(7):122–125, 1994.

[3] R. Beale and C. Creed. Affective interaction: How emotional agents affect users. *International Journal of Human-Computer Studies*, 67(9):755–776, 2009.

[4] R. H. Bordini, J. F. Hübner, and M. Wooldridge. *Programming multi-agent systems in AgentSpeak using Jason*, volume 8. John Wiley & Sons, 2007.

[5] J. Broekens, D. Degroot, and W. A. Kosters. Formal models of appraisal: Theory, specification, and computational model. *Cognitive Systems Research*, 9(3):173–197, 2008.

[6] J. Dias, S. Mascarenhas, and A. Paiva. Fatima modular: Towards an agent architecture with a generic appraisal framework. In *Emotion Modeling*, pages 44–56. Springer, 2014.

[7] J. Dias and A. Paiva. Feeling and reasoning: A computational model for emotional characters. In *Progress in artificial intelligence*, pages 127–140. Springer, 2005.

[8] M. S. El-Nasr, J. Yen, and T. R. Ioerger. Flamefuzzy logic adaptive model of emotions. *Autonomous Agents and Multi-agent systems*, 3(3):219–257, 2000.

[9] C. D. Elliott. The affective reasoner: A process model of emotions in a multi-agent system. 1992.

[10] J. Gratch and S. Marsella. A domain-independent framework for modeling emotion. *Cognitive Systems Research*, 5(4):269–306, 2004.

[11] K. V. Hindriks. Programming rational agents in goal. In *Multi-Agent Programming:*, pages 119–157. Springer, 2009.

[12] R. S. Lazarus. *Emotion and adaptation.* Oxford University Press, 1991.

[13] S. Marsella, J. Gratch, and P. Petta. Computational models of emotion. *A Blueprint for Affective Computing-A sourcebook and manual*, pages 21–46, 2010.

[14] S. C. Marsella and J. Gratch. Ema: A process model of appraisal dynamics. *Cognitive Systems Research*, 10(1):70–90, 2009.

[15] A. Ortony, G. L. Clore, and A. Collins. *The cognitive structure of emotions.* Cambridge university press, 1990.

[16] A. Pokahr, L. Braubach, and W. Lamersdorf. Jadex: A bdi reasoning engine. In *Multi-agent programming*, pages 149–174. Springer, 2005.

[17] A. Popescu, J. Broekens, and M. van Someren. Gamygdala: an emotion engine for games. *Affective Computing, IEEE Transactions on*, 5(1):32–44, 2014.

[18] W. S. Reilly. Believable social and emotional agents. Technical report, DTIC Document, 1996.

[19] R. Reisenzein. Appraisal processes conceptualized from a schema-theoretic perspective: Contributions to a process analysis of emotions. 2001.

[20] R. Reisenzein. Emotions as metarepresentational states of mind: Naturalizing the belief–desire theory of emotion. *Cognitive Systems Research*, 10(1):6–20, 2009.

[21] R. Reisenzein. What is an emotion in the belief-desire theory of emotion? 2012.

[22] R. Reisenzein. Social emotions from the perspective of the computational belief-desire theory of emotion. In *The Cognitive Foundations of Group Attitudes and Social Interaction*, pages 153–176. Springer, 2015.

[23] P. Rizzo. Why should agents be emotional for entertaining users? a critical analysis. In *Affective interactions*, pages 166–181. Springer, 2000.

[24] S. Russell, P. Norvig, and A. Intelligence. A modern approach. *Artificial Intelligence. Prentice-Hall, Egnlewood Cliffs*, 25:27, 1995.

[25] K. R. Scherer. Appraisal theory. *Handbook of cognition and emotion*, pages 637–663, 1999.

[26] K. R. Scherer. Appraisal considered as a process of multilevel sequential checking. *Appraisal processes in emotion: Theory, methods, research*, 92:120, 2001.

[27] B. R. Steunebrink, M. Dastani, and J.-J. C. Meyer. The occ model revisited. In *Proc. of the 4th Workshop on Emotion and Computing*, 2009.

Self-Explaining Robots: The Role of Goals versus Beliefs in Robot-Action Explanation for Children and Adults

Frank Kaptein[1], Joost Broekens[1], Koen Hindriks[1], and Mark Neerincx[1]

***Abstract*— A good explanation takes into account the user that is receiving the explanation. We aim to get a better understanding of the difference between children and adults that receive explanations from a robot. We implement a robot as a belief-desire-intention (BDI)-based agent, and explain its actions in two different explanation styles. Both are based on how humans explain actions amongst each other. One communicating the *beliefs* that give context information on why the agent performed the action; and, one communicating the *goals* that inform the user of the agent's purpose for performing the action. We investigate the preference of children and adults for goal- versus belief-based action explanations. From this, we learned that adults have a significantly higher tendency to prefer goal-based action explanations. This is a vital step in addressing the challenge of building explanations for different users (i.e., providing *personalized explanations*), in human-robot and human-agent interaction.**

I. Introduction

Explainable Artificial Intelligence (AI) increases a user's trust and understanding in the intelligent system they are working with [3], [14], [22]. Intelligent systems become increasingly complex, making it difficult for the user to understand the system's actions [4]. Explainable AI (XAI) is important in areas such as, medical support [22], fire-fighting [9], [10], and education [3].

A theoretical approach towards explaining behaviour is the *intentional stance*. When adopting the intentional stance, behaviour is explained by means of beliefs and goals (desires adopted for active pursuit) [5]. This way of explaining ones behaviour is referred to as *folk psychology* [2], [17], and has been used in developing XAI for BDI-based (belief, desire intention) agents [1], [8], [9], [11].

A good explanation is *personalized*, i.e., takes the user that is receiving the explanation into account. As humans mature, we improve our capabilities to create folk psychology based explanations for each other's behaviour. For example, young children (4 years old) are not very good at realizing someone may have a belief that is false [19]. Furthermore, in pedagogical environments it is known that adults, more than children, have the desire to know the objectives (goals) of instructional material [13], [15]. We therefore hypothesize that XAI, especially when based on folk psychology, is received differently by children and adults.

The context of this paper is the PAL (a Personal Assistant for a Healthy Lifestyle) project. This project helps children (aged 7-14) to cope with diabetes mellitus type 1. In this project, we develop an agent that controls a Nao-robot or its virtual avatar that interacts with the children and their parents. The system autonomously interacts with the children for prolonged periods of time. It gives them advise on how to cope with medical health issues. Therefore, it is important that the different users trust and understand the actions of the PAL-agent. To facilitate this, we develop the capability to explain these actions to the different users.

There are often many beliefs and goals that precede a BDI-based agent's decision to perform an action. Earlier work has confirmed that explanations should not be too long [8], [12]. For example, when the PAL-robot tells the child that there is a hypo when blood glucose measurement is below 4.0 mmol/L, then we may explain this by saying the robot wants to teach the child how to detect and treat having a hypo. Other reasoning that drove the robot may be that it thinks the child does not know when one has a hypo, that it thinks the child is in a good mood to learn new information, and that it wants to be a good diabetes assistant. However, providing *all* of this information to the user may be too much [12]. We need to determine what information a BDI-based agent should communicate to its users. This may further differ for different user groups, e.g., children might prefer different explanations than adults.

To address this question, we compare two different explanation styles on two different user groups. We construct goal-based, and belief-based robot-action explanations. We then ask both children and adults what explanation best helped them to understand the different actions. We compare if these different explanation styles significantly differ on this metric, taking user group into account as a factor.

We will first review related work in the field of XAI, and in particular work that focused on explaining agents, in section II. Then, in section III, we describe a generally applicable representation of a BDI-based agent's decision making, and how we derive belief-based and goal-based explanations from this. We explain the set-up of our experiment in section IV. Finally, we present the results and discuss them.

II. Motivation for Research Conducted

Previous studies have shown that Explainable AI (XAI) facilitates user trust and understanding towards the intelligent systems that they are working with [3], [14], [16], [22], [23]. For example in the MYCIN project [22], the system performed medical diagnosis of infectious blood diseases, and provided therapeutic advice. In this project, researchers found that explanations helped to increase comprehension

*This work is supported by Horizon2020 grant nr. 643783-RIA.

[1] All authors are with department of Intelligent Systems, Delft University of Technology, 2628 CD Delft, The Netherlands F.C.A.Kaptein@tudelft.nl

and confidence of the user, which contributed to the acceptance rate of the system. Furthermore, Lim et. al. [16] test four explanation strategies on over two hundred users. They found that explanations that inform the user of the agent's reasoning, create better understanding and stronger feelings of trust in the user.

When adopting the intentional stance, one assumes the other agent is pursuing certain goals and has certain beliefs about the environment [5]. Recent work on XAI has started examining how one can automatically generate explanations for BDI-based agents [1], [8], [9]. However, an explanation will fail if it provides too much information [12]. Therefore, when one adds additional information to a single explanation, then this does not necessarily improve the explanation. The developer of the explaining agent thus needs to decide what information to provide the user with.

In [18], Malle compares the use of goals and beliefs for explanations. The goals inform of the agent's aim. They answer the questions, 'To what end?'; or, 'For what purpose?' The agent's beliefs provide information on why the agent chose a certain action over another. They give information about the context and the circumstances. Goals are easier to infer from the action itself, whereas beliefs provide information specific to the particular agent that performed the action. Malle writes that in order to infer an agent's belief, one needs to take the *perspective* of this particular agent (i.e., taking the agent's point of view). [18]

Initially adults and children are equally well at perspective taking; however, adults are faster at adjusting when they learn their initial perspective is incorrect [6]. Furthermore, children and adults alike are better at perspective taking when their motivation to perform well is higher [6].

Personalized explanations take the user that is receiving the explanation into account. Previous work on explanations provided by intelligent systems already take user knowledge into account [7], [20]. By classifying a user as a beginner or expert, one can provide explanations that better fit the individual user's preferences. But, more elaborate user models are required for good personalized explanations. For example, folk psychology based explanations develop as a child *matures* [6], [12], [18].

A. Hypothesis

In this paper, we define two explanation algorithms. One that always provides the triggering condition (belief) that caused the agent to perform the action, and one that always provides the parent goal that the agent is trying to achieve. We test what explanation algorithm is preferred by adults and children, when given example actions explained by these algorithms. Our hypothesis is:

Hypothesis. *Adults have a stronger preference than children for goal-based over belief-based explanations.*

There is psychological support for this hypothesis. Behaviour explanations based on folk psychology change as humans mature [12], [18]. For example, young children (4 years old) are not good at understanding someone might belief something that is false [19]. Furthermore, children and adults alike are inclined to belief that others have similar beliefs and knowledge as we do [12]; however, adults have accumulated a vast amount of knowledge to which they link new information [15]. Furthermore, when educating adults, they strongly desire (more than children) to know the goals they are pursuing when provided with instructional material [13], [15].

III. Goal Hierarchy Trees

A goal hierarchy tree is a representation of an agent's reasoning. It shows the goals an agent aims to achieve depending on the beliefs an agent has, and it shows what actions the agent will perform given the state of its beliefs and goals. A goal hierarchy tree can be used to develop a high level design of an agent's decision making process. Previous work on explaining agents uses such a representation to construct explanations for the agent's actions [1], [8], [9]. In this section we describe the structure of a goal hierarchy tree, and how we can construct explanations from it.

A. The Structure of a Goal Hierarchy Tree

Figure 1 shows the structure of a goal hierarchy tree (GHT). A GHT shows the goals the agent is pursuing (square nodes); the actions the agent performs (darker shaded square nodes); and, beliefs that cause the agent to adopt new goals, or perform a certain action (circular nodes). Based on this GHT, we can generate explanations for performed actions.

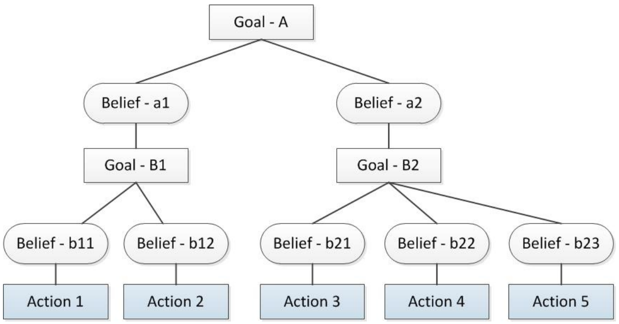


Fig. 1. The square nodes are goals the agent adopts. The top node is the agent's main goal. When following the edges, sub-goals are represented that the agent adopts in order to achieve the main goal. The triggering conditions (beliefs), that determine whether the agent should adopt a sub-goal, are represented in rounded square nodes. The agent's actual actions are the shaded nodes (the leaves) of the tree.

A GHT is constructed as follows. The agent's main goal becomes the top node of the GHT (Goal-A). When an agent should adopt a goal in order to achieve a goal, then the new goal becomes a sub-goal (Goal-B, and Goal-C are sub-goals of Goal-A). When an agent should perform an action to achieve a (sub-)goal, then it is defined as action below that goal (Action 1-5). Conditions that cause an agent to adopt one sub-goal or another, or perform a particular action, are beliefs (circular nodes).

A GHT does not model what external events can occur and how events and agent-actions cause the agent to update its beliefs. Rather, the GHT shows a high level design of the agent's *reasoning*, i.e., what action it should perform given a current state of agent beliefs and goals. This is sufficient for our purpose of generating explanations based on an agent's reasoning. However, if one want to run a BDI-based agent that acts in accordance to the GHT, then this additional modelling is required.

A BDI-based agent that runs in accordance to a GHT chooses actions as follows. Based on the agent's current goals and beliefs it chooses an action to perform. If multiple actions are applicable, then the agent randomly chooses one. When no actions are applicable then the agent remains idle, until its beliefs change. Which can cause it to adopt new goals, and can make new actions become applicable.

B. *Goal-based and belief-based agent-action explanations*

According to Malle, one can explain an action by means of the goal one aims to achieve, and by explaining why it was possible to perform the action (belief) [17]. However, Malle does not inform us on the specific beliefs and goals that humans choose in their explanations. When explaining an agent-action, we might choose any belief or goal that lies above it in the goal hierarchy (following the edges), or all of them. The latter is not advisable according to [8], [12]. Explanations should not be too long, and especially in larger GHTs the amount of beliefs and goals leading to the decision to perform an action can become large. Therefore, we need to choose a subset of beliefs and goals to put in the explanation. In this paper we use the following explanation algorithms.

The **belief-based explanation algorithm** selects the belief directly above the action (triggering condition). For example, action-2 is explained by Belief-b1c2, and action-3 by Belief-b2c1 (figure 1).

The **goal-based explanation algorithm** selects the goal directly above the action (parent goal). For example, action-2 is explained by Goal-B1, and action-3 by Goal-B2 (figure 1).

We use these simple algorithms to test the hypothesis of section II-A. Both explanation algorithms consist of one element in the goal hierarchy tree, a belief and a goal respectively. Thus, they resemble the explanation strategies proposed by Malle [17]. Furthermore, they are not very long and thus unlikely to overflow the receiver of the explanation with too much information [8], [12]. Therefore, we can use these algorithms to compare a preference between belief-based and goal-based agent-action explanations in a general way.

IV. Experimental Set-up

We developed a GHT within the context of the PAL-project, and set-up an experiment using the explanation algorithms from section III-B. We tested whether goal-based or belief-based action explanations would be better received by the participants (children and adults). We tested for a significant difference in preference, within and between these user groups.

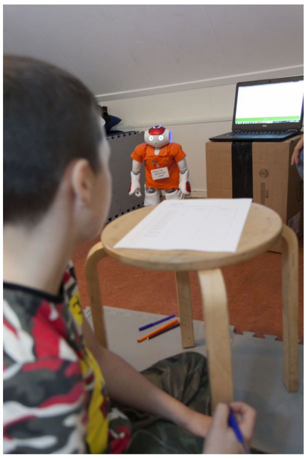


Fig. 2. Set-up of the experiment. The NAO verbally presents example scenarios and provides two explanations for each of these. The screen textually shows what the NAO is saying, so the child can always read-back what happened on the screen. The child then puts a mark at the most preferred explanation.

A. *Participants*

The participants for this experiment were selected from a diabetes camp for children. Here the NAO robots (the embodiments of the PAL-agent) explain to the participants that they need their input to learn how to be a good diabetes assistant. Children and their parents were asked to participate in this experiment.

There were 21 children and 20 parents present in the camp. 1 child did not participate in the experiment. 1 child did participate, but was looking over his friend's shoulder while filling in answers. 1 adult (parent) did not fill in the initial sheet asking for data like age, gender, and education. This left us with 19 children, and 19 adults. Children had 12 male 7 female, and adults had 8 male and 11 female participants. Children were between 8 and 11 years old. Adults were between 35 and 48 years old.

B. *Designing a Goal Hierarchy Tree*

Figure 3 presents our design of a GHT for our agent. Based on this GHT, the agent makes its decisions to perform different actions that should support a child in diabetes management.

The GHT consists of two styles of support. The agents aims to educate the child when the child is in a good mood to learn new things. The agent aims to sheer up the child when the child is sad. We call this *cognitive support*, and *affective support*, respectively. The way that this agent provides these type of supports is defined by means of ontologies that were developed in cooperation with experts in the field, i.e., the caregivers [21]. The here developed GHT resembles the treatment plan provided by these experts.

C. *Set-up & Materials*

The GHT shown in section IV-B has nine different robot actions. These actions can all be explained by using the belief-based explanation algorithm or by using the goal-based explanation algorithm. The Nao-robot presented all the

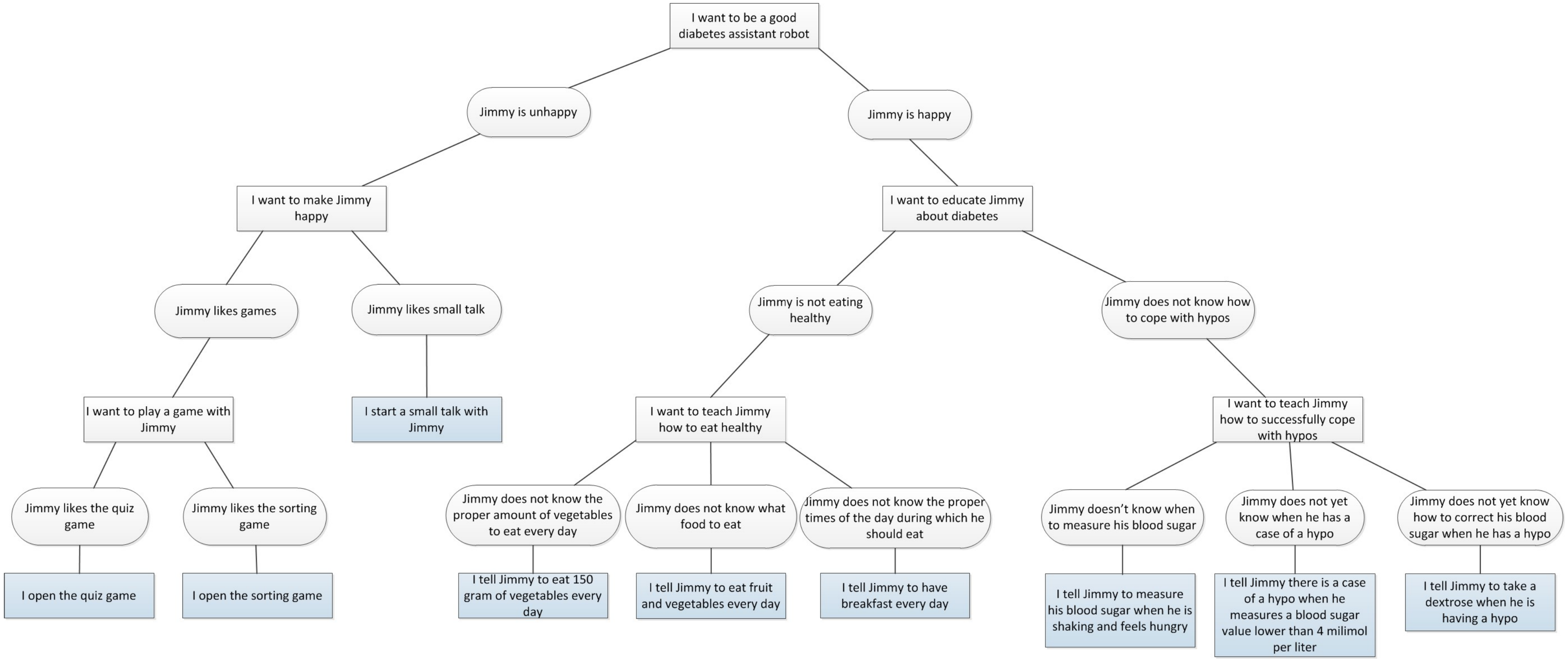

I want to be a good diabetes assistant robot
Jimmy is unhappy
Jimmy is happy
I want to make Jimmy happy
I want to educate Jimmy about diabetes
Jimmy likes games
Jimmy likes small talk
Jimmy is not eating healthy
Jimmy does not know how to cope with hypos
I want to play a game with Jimmy
I start a small talk with Jimmy
I want to teach Jimmy how to eat healthy
I want to teach Jimmy how to successfully cope with hypos
Jimmy likes the quiz game
Jimmy likes the sorting game
Jimmy does not know the proper amount of vegetables to eat every day
Jimmy does not know what food to eat
Jimmy does not know the proper times of the day during which he should eat
Jimmy doesn't know when to measure his blood sugar
Jimmy does not yet know when he has a case of a hypo
Jimmy does not yet know how to correct his blood sugar when he has a hypo
I open the quiz game
I open the sorting game
I tell Jimmy to eat 150 gram of vegetables every day
I tell Jimmy to eat fruit and vegetables every day
I tell Jimmy to have breakfast every day
I tell Jimmy to measure his blood sugar when he is shaking and feels hungry
I tell Jimmy there is a case of a hypo when he measures a blood sugar value lower than 4 milimol per liter
I tell Jimmy to take a dextrose when he is having a hypo

Fig. 3. In this figure the goal hierarchy tree of the PAL agent is shown. The square nodes are goals the agent adopts. The top node is the agent's main goal. When following the edges, sub-goals are represented that the agent adopts in order to achieve the main goal. The triggering conditions (beliefs), that determine whether the agent should adopt a sub-goal, are represented in rounded square nodes. The agent's actual actions are the shaded nodes (the leaves) of the tree.

TABLE I

DISTRIBUTION OF CHILDREN AND PARENTS OVER THE 4 CONDITIONS

	Random Seed of Scenarios			
	Normal Order of Explanations		Reversed Order of Explanations	
	Normal Order of Scenarios	Reversed Order of Scenarios	Normal Order of Scenarios	Reversed Order of Scenarios
Children	6	5	4	4
Adults	5	4	5	5

actions to the participants. For each action, it proposed two explanations obtained from the algorithms. The participants had a forced choice to prefer either one of the proposed explanations.

The robot was located in front of the participants and next to a laptop screen. Figure 2 shows a photo of the experiment. For every action and corresponding explanations presented by the Nao-robot, the laptop screen showed the action performed and explanations provided. In this way the participants can read back what the robot said. For example, the screen could look like:

Action: ‘I tell Jimmy to take dextrose when he is having a hypo.’
Explanation 1: ‘Jimmy does not yet know how to correct his blood-sugar when he has a hypo.’
Explanation 2: ‘I want to teach Jimmy how to successfully cope with hypos.’

The presentation of an action including the two explanations is henceforth called a *scenario*. The robot verbally presented nine scenarios, one for each action, and the screen showed the scenarios in text. The verbal presentation of a scenario starts with the robot saying: ‘I performed action $action$’. Where $action$ is exactly one of the actions as shown in the GHT (Figure 3), and exactly as shown on the screen. Then, the robot says: ‘How should I explain this? One: $explanation$-1; or 2, $explanation$-2’. Where explanation 1 and 2 are initialized by the belief-based algorithm and the goal-based algorithm (in random order), and exactly as shown on the screen.

By using the robot, the children experience the experiment as a fun activity rather than a chore. Which helps to keep their attention to the experiment. We chose to also use a screen since Nao-robots do not always pronounce words very well. In this way the participants can always read-back what the robot said.

Due to the camp setting we were not able to do the experiment with every user separately. We were forced to have the participants do the experiment in small groups. This meant that these groups of users would see the same order of scenarios. Thus, we counterbalanced the conditions. We produced a single random seed of scenarios, i.e., the actions are put in random order and for every action the system randomly chooses explanation 1 to be belief-based and explanation 2 goal-based, or vice versa. We counter balanced among the participants, meaning there were 4 conditions (i.e., going through the scenarios in normal or reversed order, and flipping the order of belief, and goal-based explanations within the scenarios). The participants were evenly distributed over these 4 conditions (see Table I).

D. Procedure

In small groups, the participants were asked to enter the room, and were located in front of the robot and laptop. The researcher says that he will remain present during the experiment, but that the robot will guide the experiment. Only if there are any additional questions, then these can be directed to the researcher.

The NAO robot starts the experiment with a small presentation. Here, it tells the participants that it wants to learn how to explain its behaviour to them, and that it needs their input. It then explains that it will provide example scenarios where it helped a fictional child ‘Jimmy’ to deal with diabetes. In this starting presentation the robot says it sometimes plays a game with Jimmy, and sometimes tries to educate Jimmy concerning diabetes management. The robot says that in all the example scenarios it wants to explain its action, and always considers two possible explanations. The participants are then asked to select the explanation that best helped them to understand why the PAL-robot performed that action.

V. RESULTS

To test the preference towards the different explanation styles we counted the percentage of cases where the participants preferred a goal-based explanation. A one-sample Wilcoxon signed rank test shows that the median of preferring goal-based, rather than belief-based explanations is significantly above 50%, for children ($med = 0.667, p = .007$), and parents ($med = 0.778, p < .000$). So both user groups significantly prefer goal-based explanations over belief-based explanations. Figure 4 shows the distribution of preferring goal-based explanations for children and parents.

Furthermore, a Mann-Whitney test indicated that the preference towards goal-based, rather than belief-based explanations was greater for adults than for their children, U = 112.5, p = .042, r = .33. Parents significantly prefer goal-based explanations more than their children. Thus, the medians differ significantly (0,667 for children, and 0,778 for parents).

VI. DISCUSSION

The results in the previous section show that there is a significant preference towards goal-based explanations for both user groups. However, it would be premature to state

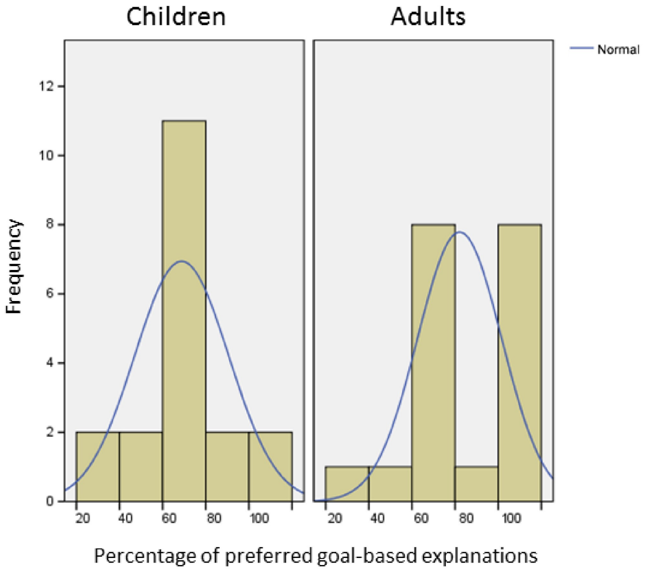


Fig. 4. Histogram showing the distribution of preferring goal-based explanations over belief-based explanations. The x-axis shows the percentage of scenarios where the subject preferred the goal-based explanation over the belief-based explanation. The y-axis shows the number of participants that had such a percentage of preferring goal-based explanation. Adults have a significantly higher preference towards goal-based action explanations (Median = 0,778), over children (Median = 0,667).

that goal-based explanations would always be preferable. Previous studies have shown contradicting findings on this subject. The work in [8] analyses three studies that provide explanations based on a goal hierarchy tree. Two studies in the firefighting domain [9], [10], (one with experts and one with Laymen), and one in a cooking domain [1] (where users were perceived as experts, since they all knew how to cook). In the non-expert domain the participants showed a preference towards belief-based explanations, in the expert domains the preference is towards goal-based explanations. It is hypothesized in [8] that this difference may be due to this expert level. Since both the children (who have diabetes mellitus themselves), and their parents are familiar with the domain, it can be expected that these users would prefer goal-based explanations. Future work should further explore this by testing this domain on layman (adult & child) users, and compare the results with the here presented findings.

Another finding in the results is that parents significantly prefer goal-based explanations more than their children. This is related to a known phenomenon in adult learning psychology. Adults need to know the objectives the instruction is aiming to achieve [13], [15]. They are goal-oriented learners that rely on their vast personal experience. They need to know how the material helps them to enhance their existing abilities, rather than children that learn under the assumption that all material will help them sometime in the future [13]. Parents thus strongly desire knowing the objectives (goals) that the personal assistant aims to achieve, while this robot performs actions that support diabetes mellitus management.

Another explanation of the results is that children are more motivated to understand a robot character. In order to infer an agent's belief, one needs to take the *perspective* of this particular agent (i.e., taking the agent's point of view) [18]. Children and adults alike are better at perspective taking when their motivation to do this properly is higher [6]. A higher motivation would then correspond to a better understanding of belief-based explanations. However, adults are better adapted to perspective taking than children. Adults are faster at adjusting when they learn their initial perspective is incorrect [6]. Further research is required to properly test for the influence of motivation when comparing belief-based and goal-based explanations.

The experiment presented was well controlled. Despite the many similarities between our user groups, (i.e., they both face the problem of managing diabetes, they work with the same caregivers, at the same hospitals, they even share the same genes) we still found that there was a significant difference between adults and children in their preference towards the explanation styles. This is strong evidence for child/ adult being the only factor responsible for this difference.

VII. Conclusion

In this paper we compared the preference for goal-based versus belief-based agent-action explanations between two user groups. We presented children and adults a set of example robot actions, and provided two possible explanations for these. One explanation communicating the triggering condition (belief) preceding the decision to perform an action. The other explanation provided the parent goal, that the action was aiming to achieve. The users were asked to choose the explanation that *best helped them to understand why the PAL-robot performed this action*. We found that adults have a significantly *higher* preference towards goal-based explanations than children.

These findings will help us to improve explanations provided by BDI-based agents, and robots implemented as such agents. Furthermore, the work here confirms that different user types (child/ adult) have different preferences towards explanation styles. This shows that explanation techniques that take user-type into account (i.e., personalized explanations) are needed and possible.

References

[1] J. Broekens, M. Harbers, K. Hindriks, K. Van Den Bosch, C. Jonker, and J. J. Meyer. Do you get it? user-evaluated explainable bdi agents. In *Multiagent System Technologies MATES*, volume 6251 LNAI, pages 28–39. Springer Berlin Heidelberg, 2010.

[2] P. M. Churchland. Folk psychology and the explanation of human behavior. *Philosophical Perspectives*, 3:225–241, 1989.

[3] C. Conati and K. VanLehn. Providing adaptive support to the understanding of instructional material. In *Proceedings of the 6th international conference on Intelligent user interfaces*, pages 41–47. ACM, 2001.

[4] M. G. Core, H. C. Lane, M. Van Lent, D. Gomboc, S. Solomon, and M. Rosenberg. Building explainable artificial intelligence systems. In *Proceedings of the national conference on artificial intelligence*, volume 21, page 1766. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2006.

[5] D. C. Dennett. Three kinds of intentional psychology. *Perspectives in the Philosophy of Language: A Concise Anthology*, pages 163–186, 1978.

[6] N. Epley, C. K. Morewedge, and B. Keysar. Perspective taking in children and adults: Equivalent egocentrism but differential correction. *Journal of Experimental Social Psychology*, 40(6):760–768, 2004.

[7] S. Gregor and I. Benbasat. Explanations from intelligent systems: Theoretical foundations and implications for practice. *MIS quarterly*, pages 497–530, 1999.

[8] M. Harbers, J. Broekens, K. Van Den Bosch, and J.-J. Meyer. Guidelines for developing explainable cognitive models. In *Proceedings of ICCM*, pages 85–90. Citeseer, 2010.

[9] M. Harbers, K. Van den Bosch, and J.-J. Meyer. Design and evaluation of explainable bdi agents. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on*, volume 2, pages 125–132. IEEE, 2010.

[10] M. Harbers, K. van den Bosch, and J.-J. C. Meyer. A study into preferred explanations of virtual agent behavior. In *International Workshop on Intelligent Virtual Agents*, pages 132–145. Springer, 2009.

[11] K. V. Hindriks. Debugging is explaining. In *International Conference on Principles and Practice of Multi-Agent Systems*, pages 31–45. Springer, 2012.

[12] F. C. Keil. Explanation and understanding. *Annual review of psychology*, 57:227, 2006.

[13] M. S. Knowles et al. *The modern practice of adult education*, volume 41. New York Association Press New York, 1970.

[14] D. N. Lam and K. S. Barber. Comprehending agent software. In *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, pages 586–593. ACM, 2005.

[15] S. Lieb and J. Goodlad. Principles of adult learning, 2005.

[16] B. Y. Lim, A. K. Dey, and D. Avrahami. Why and why not explanations improve the intelligibility of context-aware intelligent systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2119–2128. ACM, 2009.

[17] B. F. Malle. How people explain behavior: A new theoretical framework. *Personality and social psychology review*, 3(1):23–48, 1999.

[18] B. F. Malle. How the mind explains behavior. *Folk Explanation, Meaning and Social Interaction. Massachusetts: MIT-Press*, 2004.

[19] H. W. H. Mayringer. False belief understanding in young children: Explanations do not develop before predictions. *International Journal of Behavioral Development*, 22(2):403–422, 1998.

[20] G. Milliez, R. Lallement, M. Fiore, and R. Alami. Using human knowledge awareness to adapt collaborative plan generation, explanation and monitoring. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*, pages 43–50. IEEE Press, 2016.

[21] M. A. Neerincx, F. Kaptein, M. A. van Bekkum, H.-U. Krieger, B. Kiefer, R. Peters, J. Broekens, Y. Demiris, and M. Sapelli. Ontologies for social, cognitive and affective agent-based support of childs diabetes self-management. *Artificial Intelligence for Diabetes*, page 35.

[22] E. H. Shortliffe, R. Davis, S. G. Axline, B. G. Buchanan, C. C. Green, and S. N. Cohen. Computer-based consultations in clinical therapeutics: explanation and rule acquisition capabilities of the mycin system. *Computers and biomedical research*, 8(4):303–320, 1975.

[23] L. R. Ye and P. E. Johnson. The impact of explanation facilities on user acceptance of expert systems advice. *Mis Quarterly*, pages 157–172, 1995.

Ontologies for social, cognitive and affective agent-based support of child's diabetes self-management

Mark A. Neerincx[1,2], **Frank Kaptein**[2], **Michael A. van Bekkum**[1], **Hans-Ulrich Krieger**[3], **Bernd Kiefer**[3], **Rifca Peters**[2], **Joost Broekens**[2], **Yiannis Demiris**[4] and **Maya Sapelli**[1]

Abstract. The PAL project is developing: (1) an embodied conversational agent (robot and its avatar); (2) applications for child-agent activities that help children from 8 to 14 years old to acquire the required knowledge, skills and attitude for adequate diabetes self-management; and (3) dashboards for caregivers to enhance their supportive role for this self-management learning process. A common ontology is constructed to support normative behavior in a flexible way, to establish mutual understanding in the human-agent system, to integrate and utilize knowledge from the application and scientific domains, and to produce sensible human-agent dialogues. This paper presents the general vision, approach, and state of the art.

1 Ontologies in Cognitive Engineering

In Europe, an increasing number of about 140,000 children (<14 year old) have Type 1 Diabetes Mellitus (T1DM) [1]. The PAL project develops an Embodied Conversational Agent (ECA: robot and its avatar) and several applications for child-agent activities (e.g., playing a quiz and maintaining a timeline with the agent) that help these children to enhance their self-management (PAL, Personal Assistant for healthy Lifestyle, is an European Horizon-2020 project; www.pal4u.eu). In addition, it develops dashboards for caregivers (like diabetes nurses and parents) to enhance their supportive role. The general objective is to establish a smooth transition of the diabetes care responsibility from caregiver to the developing child, so that the child will have the required knowledge, skills, and attitude for adequate self-management at adolescence.

PAL is part of a joint, cognitive system, in which humans and agents share information and learn to improve self-management. The required sharing of (evolving) knowledge in the envisioned "blended care" setting has four important challenges:

1. To address the values & norms of both the caregivers in their different hospitals (e.g., diabetes regimes), and the caretakers in their different contexts (e.g., privacy, literacy).
2. To establish mutual understanding (a) within and between the different stakeholders of the PAL system (e.g., the end-users like children and caregivers and research & developers like academics and engineers), and (b) between the humans and PAL-agents.
3. To continually acquire, utilize and deploy knowledge about child's self-management support.
4. To produce natural, flexible, personalized human-agent interactions that are sustainable in the long term as well as allow to extract data about the user from these interactions.

To meet these four challenges, we are developing an ontology as an integrated part of system development, i.e., in a systematic, iterative, and incremental cognitive engineering process. First, available ontologies and approaches are assessed and, possibly, improved and integrated for our purposes (section 2). Second, relevant theories and models of the concerning scientific research fields are identified and formalized for adoption in the ontology (section 3). Third, the ontology is implemented in an artefact or prototype (i.e., the PAL system) and, subsequently, tested and refined (section 4).

2 Models for Diabetes Self-Management

Because PAL covers a large domain of interest, we have developed ontology models as high-level building blocks for smaller, separate areas of interest (frames). First, appropriate frames were selected from existing (global) libraries and, if needed, tailored to the PAL purposes. Second, for the missing elements, frames were modeled by constructing a new ontology. Subsequently, the individual frame models were related (interlinked) in an integrated PAL model. Because most existing ontologies provide "only" a partial fit to the intended scope of PAL, we needed to adapt these models by extending them (e.g., when concepts were lacking), or by selectively downsizing them (e.g., when there were too many details or concepts in the model). The frames we have identified and modeled so far are among others: (1) human and machine roles involved in self-management; (2) emotions and sentiments that cover the emotional responses of both robot and child to interaction as well as the general state of mind of the child; (3) tasks that include among other things: learning and self-management tasks, associated goals, and objects; (4) issues related to medical examinations (e.g., lab values); and (5) dialogue management through a combination of dialogue acts and shallow semantics. A more elaborate PAL ontology will also include interaction and behavior models of robot and avatar, a model for privacy of information of self-management activities and a model to cover the agreements and social contracts between child and ECA.

Figure 1 provides a simple example of the task frame (cf. [2]). An Agent, such as a child or avatar, is an entity that performs a certain task, like an educative quiz game. An associated goal

[1] TNO, The Netherlands, mark.neerincx@tno.nl
[2] Delft University of Technology, Netherlands, F.C.A.Kaptein@tudelft.nl
[3] DFKI, Germany, krieger@dfki.de
[4] Imperial College, United Kingdom, y.demiris@imperial.ac.uk

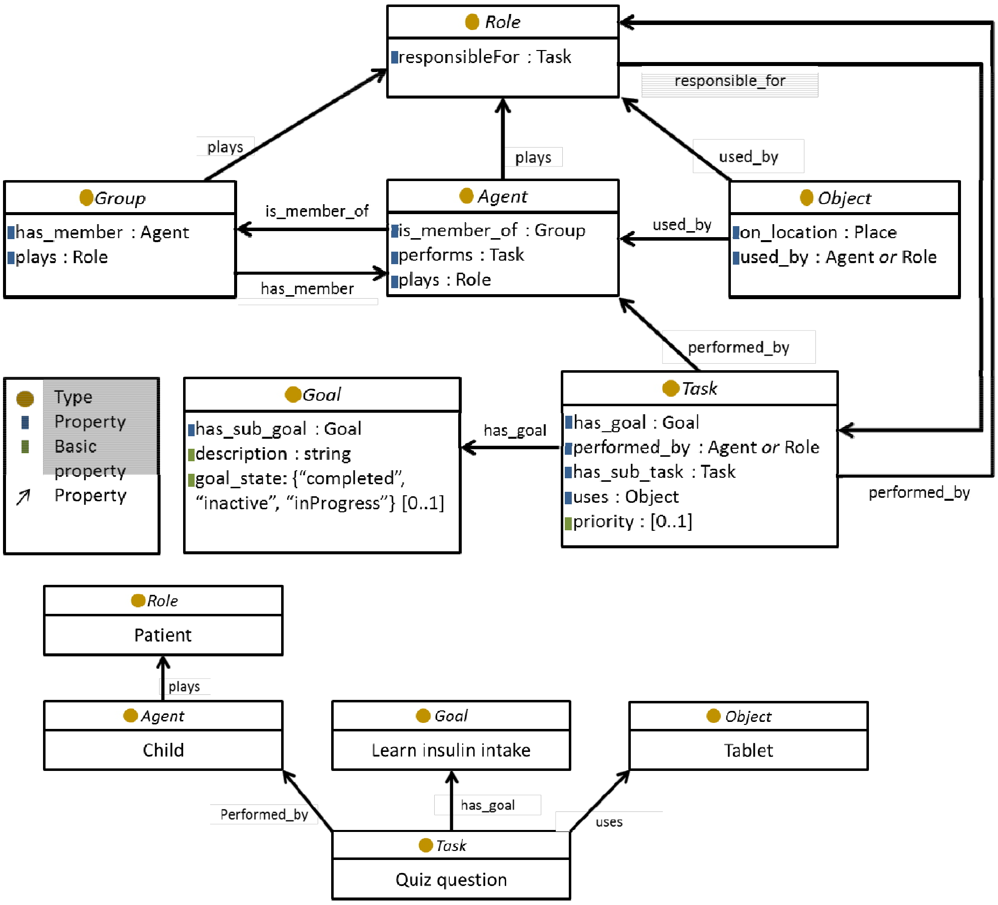


Figure 1: Simple example of the general task frame at the top and an instantiation at the bottom.

(e.g., learning about Insulin taking) can be attained by performing the related task (e.g., answering related questions correctly while playing the quiz). Objects such as a tablet device, are typically used when performing the task. The agent has a role while performing the task (e.g., patient) and can be part of a group of agents (e.g., parents).

Important objectives of the PAL ontology are norm-compliancy, shared understanding, interpretation, reasoning, and generation of verbal utterances. The ontology is based on a uniform representation of an application semantics that uses dialogue acts and frames that are defined in an extended RDF and OWL ontology [3]. In addition, all data that influence multimodal utterance generation are specified in the ontology (e.g., user data), which facilitates access and combination of the different bits of information. We heavily extended existing processing components, e.g., the reasoning engine HFC from DFKI and its database layer [4], which make information available to the interaction management and analysis. We defined a new formalism for the specification of dialogue policies that combines dialogue rules, transaction time-based knowledge representation [5], and dialogue history in a unique way. One important part of the PAL ontology combines dialogue acts using the DIT++ standard [6] and semantic frames, loosely based on thematic relations [7], used in today's frameworks VerbNet, VerbOcean, or FrameNet. Below, we show a simplified version of the combined representation, built for the sentence: "I could show you a picture of the last football game".

Offer(Showing, theme=Picture, sender=I MYSELF,
addressee=NAO ROBOT, topic=Football).

3 Integrating Relevant Theories

In the PAL project, dedicated studies of models in the concerning scientific research areas are being conducted. For supporting the *social processes* that are involved in self-management learning, PAL models relationships in terms of familiarity or intimacy, liking, attitude and benevolence [8]. Particularly, the child-ECA bonding process is being supported by the Dyadic Disclosure Dialog Module (3DM) that supports the mutual child-agent self-disclosures. The PAL ontology distinguishes three main classes for these dialogues: disclosure, prompt and closer. In addition to valence and topic, each disclosure has an intimacy level according to the 4-level Disclosure Intimacy Rating Scale (DIRS). Burger et al. (2016) provide more detailed information on the 3DM of PAL and its theoretical foundations [9].

For supporting the *cognitive processes*, the diabetes knowledge and corresponding learning goals have been modeled to monitor and reason about progress (e.g., on diabetes regimes, self-control, food, physical exercises, and stress coping). Goal attainment is an important indicator of the changes in behavior of children [10], and can be supported by personalized feedback of the ECA. Figure 2 provides a simplified sketch of a dialogue instantiation in the PAL system. Answering a quiz question is an example of a task (Fig 1). Answering correctly (partly) fulfills one or more (learning) goals. Note that the same goal can be satisfied by another task too, such as a sorting game. The different goals have specific difficulty levels (0-3). The caregivers decide what goals are currently relevant and achievable for a child. Together with caregivers, a child selects the specific goals to attain: <child:URI> <hasGoal> <goal:URI>. Since the system will only suggest tasks that can achieve the child's current goals, these tasks are implicitly following these same difficulty levels. For example, a quiz question that satisfies a level 3 goal will be more difficult than a question satisfying a level 0 goal. Goal attainment is an important aspect of self-management. PAL will monitor the goal attainment progress: <Goal:URI> <hasProgress> float. For every goal, the ontology defines what tasks, and (sub-)goals should be achieved to achieve the goal itself. GoalProgress is function of goal:neededForAsClass and goal:requiresAsClass. By computing the percentage of tasks, subtasks, and sub-goals currently achieved, the system computes a current progress on this goal. This is recorded with a time stamp, so that progress over time can be calculated.

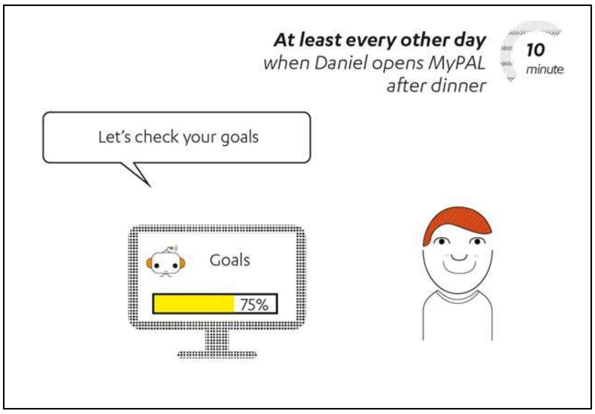


Figure 2: Simplified situated speech act of the avatar.

For supporting the *affective processes*, the PAL system introduces several methods to model the affective state of a child. First, sentiment mining technology is applied to estimate child's mood in the child-PAL textual dialogues [11]. Second, in the tablet application, the child can further self-report on the experienced emotions and moods for activities the child performed during the day. Third, the child model will estimate emotions experienced by the child resulting from activities proposed by the ECA. For example, the ECA can propose to play a quiz with the child, and predict joy when the child did well during the quiz. This child model is based on the belief desire theory of emotions [12, 13], in which emotions are a direct consequence of beliefs and desires of an individual. For example, if one beliefs X and desires X, then one is happy about X. This way, the child model can reason about the child's beliefs and desires. The model improves over time. If the child self-reports positive emotions during an activity while the child model estimates negative ones, then the child model updates the beliefs-desire assumptions concerning the child. The PAL ontology will represent complex affective states. Emotions are directed at objects, or events, and are short intense episodes. Moods are undirected and less intense, but linger for a prolonged period of time. Emotions are stored with the activity that had this emotion as a consequence. Moods contain a timestamp, indicating when it was measured. This representation makes it possible to find correlations between activities and affect over a prolonged period of time.

4 Implementation and Evaluation

The PAL system consists of several modules with dedicated support objectives. For example, the dialogue manager aims at engaging conversations between child and the ECA, the action-selection module HAMMER [14] learns over time what the best actions are (e.g., proposing to play a quiz, or starting a new dialogue) to improve the child's knowledge of diabetes while maintaining a positive emotional state for the child, and the child model aims at estimating the emotional states.

Figure 3 shows the data flows of the PAL system with an extendable set of modules that communicate through a common Nexus. When a module has new information to share with other modules (e.g., action selection proposes to play a quiz) then this information is posted on the Nexus. Any module can read and use this new information. The application can then read this proposal and start a quiz on the tablet, and/or the dialogue manager can start a small dialogue by asking the child whether he/she wants to play a quiz. The PAL ontology provides the shared knowledge representations, defined in the extended HFC reasoner and allowing for testing and refining.

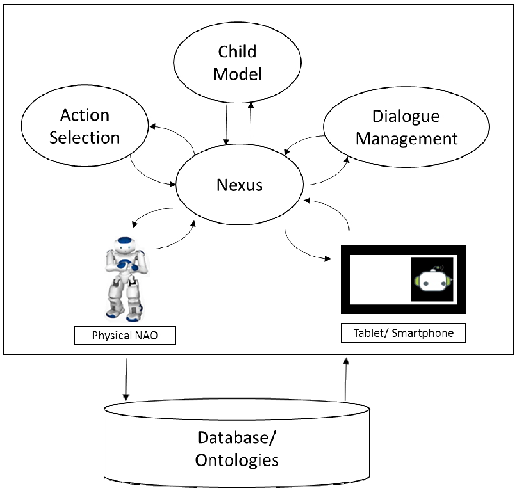


Figure 3: The PAL system.

Currently, we are analyzing the first data sets of children and caregivers that used the PAL system in diabetes camps, hospitals and at home (in Italy and in the Netherlands) from a few days to 4 weeks. Based on the ontological concepts, we can identify meaningful patterns in the data that will be used to improve the intelligence of PAL, e.g. concerning the goal attainment progress (i.e., enhance the knowledge base with refined ontology and reasoning mechanisms). Furthermore, the data analysis will help to refine the ontology substantially. For example, parents' relationship (cohabit or divorce) seems to affect child's PAL usage (quantity and regularity) substantially. These concepts with their mutual relations are being added to the ontology to "feed" mitigating support functions. A second example concerns the identified cultural differences in Italian and Dutch children for the wealth and directness of their multimodal interactions with the robot [11]. Among other things based on these results, the child and robot models will be enriched to establish adaptive —personalized and culture-harmonized— child-robot interactions.

5 Discussion

The PAL project develops personalized support for children, helping them to acquire the required attitude, knowledge and skills for adequate diabetes self-management. It applies a situated Cognitive Engineering (sCE) methodology to design and test: (1) an ECA for children, (2) several (educative) child-ECA activities, and (3) dashboards for caregivers. This methodology includes an ontology engineering component to establish a system's knowledge base that is univocal, theoretically sound, coherent, consistent and transparent [15]. The resulting common ontology is used to establish mutual understanding in the human-agent system, to integrate and utilize knowledge from the application and scientific domains, and to produce sensible human-agent dialogues. For the first version of the PAL ontology, a network of connected ontologies ("frames") have been constructed, each consisting of general concepts and their relations. The "dialogue management frame" was worked out in more detail, i.e., the specification of the data structures to be used by the dialogue specifications, dialogue history, and information state. Furthermore, the reasoning components were adapted, so that this knowledge source can be used efficiently once the formalism specification is fully implemented.

The PAL project entails multi-disciplinary research and design of a "blended care" system with the involvement of a large diversity of stakeholders. In general, the ontology construction helped to identify (interrelated) key concepts that should be univocally addressed in the design (e.g., requirements), implementation (e.g., dialogues) and evaluations (e.g., goal attainment). Furthermore, it enforces the systematic integration of relevant theories on social, cognitive and affective processes into the support system (e.g., on bonding, goal-driven learning and emotion). In line with the general iterative development process, the ontology will be refined for enhanced self-management support in the next versions of the PAL system.

It is interesting to note that the PAL ontology can be viewed as a frame-based ontology in terms of Minsky [16] and Hoekstra [17]: An explicit, structured, and semantically rich representation of declarative knowledge like psychological theories of human cognition use, distinguishing "frames" or "classes" (upper level) from "instantiations" (lower level). This approach seems therefore particularly appropriate for representing knowledge involved in learning [15], e.g., learning to cope with a chronic disease.

Acknowledgements

PAL is funded by Horizon2020 grant nr. 643783-RIA.

References

[1] Freeborn, D., Dyches, T., Roper, S.O. and Mandleco B. (2013). Identifying challenges of living with type 1 diabetes: child and youth perspectives, *Journal of clinical nursing*, vol. 22, no. 13-14, pp. 1890–1898.

[2] Van Welie, M., Van der Veer, G.C. and Eliëns, A. (1998). An ontology for task world models. *Eurographics Workshop on Design Specification and Verification of Interactive Systems*, pp.3–5

[3] ter Horst, H. J. (2005). Completeness, decidability and complexity of entailment for RDF Schema and a semantic extension involving the OWL vocabulary. *Journal of Web Semantics*, 3:79–115.

[4] Krieger, H.-U. (2013). An Efficient Implementation of Equivalence Relations in OWL via Rule and Query Rewriting. *Proceedings of the 7th International Conference on Semantic Computing (ICSC).*

[5] Krieger, H.-U. (2016). Capturing Graded Knowledge and Uncertainty in a Modalized Fragment of OWL. *Proceedings of the 8th International Conference on Agents and Artificial Intelligence (ICAART).*

[6] Bunt, H., Alexandersson, J., Choe, J.-W., Fang, A.C., Hasida, K., Petukhova, V., Popescu-Belis, A., and Traum, D. (2012). ISO 24617-2: A semantically-based standard for dialogue annotation. *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC).*

[7] Fillmore, C.J. (1977). The Case for Case Reopened. In: Grammatical Relations. *Syntax & Semantics*. Academic Press.

[8] Altman, I. and Taylor, D. (1973): *Social penetration theory*. Holt, Rinehart & Mnston, New York.

[9] Burger, F., Broekens, J. and Neerincx, M.A. (2016). A Self-Disclosing Companion Agent for Children. *Proc. of the Intelligent Virtual (IVA) 2016 Conference*. Springer Int. Publishing

[10] Kleinrahm, R. Keller, F., Lutz, K. Kölch, M. and Fegert, J.M. (2013). Assessing change in the behavior of children and adolescents in youth welfare institutions using goal attainment scaling. *Child & Adolescent Psychiatry & Mental Health*, 7:33 (11 pages).

[11] Neerincx, A., Sacchitelli, F., Kaptein, R., van der Pal, S., Oleari, E., & Neerincx, M. A. (2016). Child's Culture-related Experiences with a Social Robot at Diabetes Camps. Eleventh ACM/IEEE International Conference on Human Robot Interaction (pp. 485-486). IEEE Press.

[12] Reisenzein, R. (2001). Appraisal processes conceptualized from a schema-theoretic perspective: Contributions to a process analysis of emotions.

[13] Reisenzein, R. (2009). Emotions as metarepresentational states of mind: Naturalizing the belief–desire theory of emotion. *Cognitive Systems Research*, 10(1), 6-20.

[14] Demiris, Y. and Khadhouri, B. (2006). Hierarchical Attentive Multiple Models for Execution and Recognition (HAMMER), in *Robotics and Autonomous Systems*, 54:361-369.

[15] Peeters, M. M., Bosch, K. V. D., Neerincx, M. A., & Meyer, J. J. C. (2014). An ontology for automated scenario–based training. *International Journal of Technology Enhanced Learning*, 6(3), 195-211.

[16] Minsky, M. (1975). *A framework for representing knowledge*. The Psychology of Computer Vision, McGraw-Hill, New York.

[17] Hoekstra, R. (2009). *Ontology Representation: Design Patterns and Ontologies that Make Sense*, IOS Press.