Evaluating Cognitive and Affective Intelligent Agent Explanations in a Long-Term Health-Support Application for Children with Type 1 Diabetes

Frank Kaptein[1], Joost Broekens[2], Koen Hindriks[3], and Mark Neerincx[1,4]

Abstract—**Explanation of actions is important for transparency of-, and trust in the decisions of smart systems. Literature suggests that emotions and emotion words - in addition to beliefs and goals - are used in human explanations of behaviour. Furthermore, research in e-health support systems and human-robot interaction stresses the need for studying long-term interaction with users. However, state of the art explainable artificial intelligence for intelligent agents focuses mainly on explaining an agent's behaviour based on the underlying *beliefs and goals* in *short-term experiments*. In this paper, we report on a long-term experiment in which we tested the effect of cognitive, affective and lack of explanations on children's motivation to use an e-health support system. Children (aged 6-14) suffering from type 1 diabetes mellitus interacted with a virtual robot as part of the e-health system over a period of 2.5 - 3 months. Children alternated between the three conditions. Agent behaviours that were explained to the children included why 1) the agent asks a certain quiz question; 2) the agent provides a specific tip (a short instruction) about diabetes; or, 3) the agent provides a task suggestion, e.g., play a quiz, or, watch a video about diabetes. Their motivation was measured by counting how often children would follow the agent's suggestion, how often they would continue to play the quiz or ask for an additional tip, and how often they would request an explanation from the system. Surprisingly, children proved to follow task suggestions more often when *no explanation* was given, while other explanation effects did not appear. This is to our knowledge the first long-term study to report empirical evidence for an agent explanation effect, challenging the next studies to uncover the underlying mechanism.**

Index Terms—**Explainable AI, Long-term human-agent interaction, Goal-based XAI, Emotions in explanations**

I. Introduction

Humans are increasingly supported by Artificial Intelligence (AI), for example, at home using virtual assistants, in health care settings, and in education [1]. Transparency of why such systems provide particular advice or choose certain actions, as well as user trust in such systems, is important [2]–[4]. Therefore, the ability to provide explanations to motivate the reasoning behind the AI's decisions, i.e., eXplainable AI (XAI), becomes increasingly important. This trend is supported by the recent General Data Protection Regulation (GDPR) law, which states that users have the right to explanations [5].

Current XAI for agents is often based on folk psychology, i.e., how humans in their everyday lives explain their decisions amongst each other [6]. Such explanations are based on the beliefs and goals of the system. For example, 'I suggest you watch this video about diabetes because I **think** (a system belief) it contains valid information about proper blood sugar levels, and I **want** (a system goal) you to learn when your blood sugar level would be too low'. Using beliefs and/or goals for explaining intentional behaviour is common in both human-human communication and in XAI [7]–[11]. We refer to this as providing *cognitive explanations*.

Literature suggests that emotions and emotion words - in addition to beliefs and goals - are used in human explanations of behaviour [12]–[14]. Humans explain their decisions also based on their emotions. For example, 'I called the hospital because I was **scared** (emotion) that I might have a hypo (too low blood sugar level)'. As such, explanations of agents based on beliefs and/or goals may not always be sufficient and emotions may be required as part of the explanations in human-agent interaction [15].

Furthermore, research in e-health support systems and human-robot interaction stresses the need for studying long-term interaction with users [1], [16]–[18]. However, state of the art of XAI for intelligent agents has focussed mainly on explaining an agent's behaviour based on the underlying *beliefs and/or goals* in *short-term experiments* [4], [8], [11], [19].

In this paper, we report on a long-term experiment in which we tested the effect of cognitive, affective and lack of explanations on children's motivation to use an e-health support system. Children (aged 6-14) suffering from Type 1 Diabetes Mellitus (T1DM) interacted with a virtual robot as part of the e-health system over a period of 2.5 to 3 months. Children alternated between the three conditions. Agent behaviours that were explained to the children included why 1) the agent asks a certain quiz question; 2) the agent provides a specific tip (a short instruction) about diabetes; or, 3) the agent provides a task suggestion, e.g., play a quiz, or, watch a video about diabetes. Their motivation was measured by counting how often children would follow the agent's suggestion, how often they would continue to play the quiz or ask for an additional tip, and how often they would request an explanation from the system.

This work has been supported by the PAL project, Horizon2020 grant nr. 643783-RIA.

[1]TU Delft, Van Mourik Broekmanweg 6, 2628 XE Delft, The Netherlands Frank_Kaptein@hotmail.com

[2]Leiden University, LIACS, Niels Bohrweg 1, 2333 CA Leiden, The Netherlands

[3]VU, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands

[4]TNO, Postbus 23, 3769 ZG Soesterberg, The Netherlands

978-1-7281-3888-6/19/$31.00 ©2019 IEEE

II. Motivation, Related Work, and Hypothesis

First, we motivate why intelligent agents in consequential domains, such as health-care, must be able to explain their behaviour. As computer systems become more powerful, more complexity is introduced in their decision making [4]. To maintain trust in a system in the long-term, the system must be clear about the task it is trying to achieve [1]. Lack of trust in a behaviour change system causes users to not rely on the given advice [20], and can cause them to misuse or even abandon the system [21]. XAI has been shown to have a positive impact on a user's trust in several studies [2], [3], [22], [23]. Indeed, such consequential domains often include explainable AI for transparency and intelligibility [6].

Now we motivate why emotions need to be considered in the generation of explanations. XAI is typically based on how humans explain their behaviour amongst each other, i.e., on *folk psychology* [12], [13], [24]. This refers to the use of beliefs, goals and emotions to explain behaviour [13], [14]. Explanations using beliefs and goals (which we call *cognitive explanations*) are often used in XAI [7]–[11]. However, using emotions and emotion words for explanations (which we call *affective explanations*) has not yet been properly tested in XAI. Still, synthetic emotions expressed by agents have the potential to influence user attitudes and behaviour [25], and explanations of agents based on beliefs and goals may not always be sufficient, emotions may be required as part of the explanations in human-agent interaction [15].

Finally, we motivate why long-term experiments are essential. Explanations are typically done by using the agent's beliefs and/or goals and in short-term experiments [4], [6], [8], [11], [19]. However, the importance of testing long-term effects has been stressed in human-robot interaction and e-health [1], [16]–[18]. Long-term interaction typically has more repetition of information and interaction patterns, and such systems need to overcome novelty effects. Related work shows that reasons to stop using a robot change over time [18]. In the short-term, the robot must be enjoyable and easy to use, in the long-term it must be functionally relevant.

The context of this work is the PAL project (a Personal Assistant for a healthy Lifestyle). Here we develop a support application with a (Nao) robot and virtual avatar thereof that helps children (aged 6-14) with T1DM to cope with their illness. The child sets personal learning goals with the caregiver, such as, recognise hypo and correct blood sugar accordingly. The PAL agent then shapes the activities to support the child to achieve these goals. For example, during the quiz PAL might ask the child what the child should do when (s)he suddenly starts shaking and is feeling very hungry. The child can then ask PAL why the agent asks the child this question. The XAI module developed and reported upon here enables the system to respond along the lines of: I would be happy for you if you learn how to recognise that you have a hypo, and learn what you should then do'.

Because agent explanation of action is important for trust and motivation, because emotions need to be considered as part of the explanations, and because XAI needs to be evaluated in such long-term experiments, we address the following question:

What is the effect of cognitive, affective and lack of explanations on the motivation of children to use an e-health support system in long-term interaction?

We look at several motivational effects of explanation style and split our research question into four hypotheses. First, we want to know if children appreciate and use explanations. We assess this by measuring the total number of requested explanations.

Hypothesis 1: There is a difference in total number of requested explanations induced by explanation style (cognitive versus affective explanations).

Second, we expect explanations to have an effect on the usage of the system. People desire to know the goals they are pursuing when being educated [26], [27]. Explanations may help a user to better understand why an action is proposed, thereby understanding the learning goal. In previous work it was found that adults, more than children, prefer goal-based over belief-based explanations [11]. Here we are interested in the effect of cognitive versus affective explanations.

Hypothesis 2: There is a difference in the average number of questions in a quiz before children close it given the explanation style (cognitive versus affective versus lack of explanation).

Hypothesis 3: There is a difference in how often children request an additional tip given the explanation style (cognitive versus affective versus lack of explanation).

Finally, to directly assess the motivational value of explanations, we look at how often a task suggestion by the system is followed. We expect such task suggestions to be followed more often when they are explained because in general people are more motivated to learn something when they know why they should learn it [26], [27].

Hypothesis 4: There is a difference in how often children follow a task suggestion after they received an explanation, induced by explanation style (cognitive versus affective versus lack of explanation).

III. Implementation of a Model for Explainable AI

In our model, explanations consist of some raw *content* and a *presentation* of the content. The content of the explanation is the goal that the agent is pursuing with its behaviour. The presentation is the resulting set of sentences generated. We consider two different *styles* in which these sentences can be formulated, (i.e., *cognitive* and *affective* explanations).

A. Explainable Actions

We explain three different types of actions shown by the PAL agent. 1) Asking the user a quiz question (e.g., 'What should you do when you are experiencing a hypo whilst doing sports?'). 2) Giving the user a *tip of the day* or shortly a *tip* (e.g., 'When your blood sugar level is below 4.0 mmol/L then you have a hypo'). 3) Suggesting a task to do (e.g., 'play the quiz' or 'watch this video').

Quiz questions and tips are activities that the child can do within the system. A child can play a quiz as often and for as many questions as they like. When a child requests a tip, then (s)he can request *next tips* as often as (s)he likes. Suggesting an activity happens when the child is shown a list of four possible activities ('tasks') to do in the system. This always happens when the application starts. Additionally, the child can request a (new) list of possible tasks at any moment. The PAL agent then always suggests that the top-most task would currently be the best task to do. It can potentially motivate this suggestion further by explaining *why* it thinks the child should do that activity (See also figure 1). The text used to suggest a task is chosen randomly from a set of pre-made sentences. For example, in figure 1.a the text is 'Let's do the first activity' and in figure 1.c the text is 'I think you should do this first activity'. In the cognitive and affective styles, the explanation and task suggestion texts are concatenated in a single text balloon.

B. Content of explanations

The content of the explanations is the goal that the agent is trying to pursue. Which is a common approach in XAI [7]–[11]. However, an action often pursues multiple goals. For example, (a proposal for) watching a video can be valuable for a large list of unrelated learning goals (like, 'recognise hypo', 'be able to talk with friends about diabetes', and 'start eating more vegetables'). This is a complicating factor since an explanation loses its value when it becomes too long [28], so we should not mention all the goals in an explanation.

Within PAL we chose a simple solution for this problem. We pick a random goal as content for the explanation to show the child why an activity is beneficial for the child's self-management. Clearly, we are not claiming that this is the best way of selecting content for the explanation. However, we do believe that this is a valid way that fits our purposes (i.e., measure the effect of explanations on a child's motivation to use the system in long-term interaction).

C. Presentation of explanations

With the content of the explanation being a single goal, we still need to share this information with the child. So, we need a way to transform it into some natural language sentence. We do this partly by automation and partly by annotation. The learning goals are annotated with a natural language sentence that describes them, e.g., 'how to recognise that your blood sugar level might be too high (hyper), and what you should then do'. We can then automatically put a sentence in front of that that completes the explanation, e.g., 'I want you to learn..'. And, we can add a sentence behind to refer to the explained action, like, 'That is why I ask you this question', or 'And that is why I gave you this tip (of the day)'. So a full explanation can be: 'I want you to learn how to recognise that your blood sugar level might be too high (hyper), and what you should then do. That is why I ask you this question.'

We differentiate the sentences before and after the description to prevent repetitiveness in sentences. For example, 'I want you to learn' can be interchanged with 'my aim is that you learn', and 'That is why I ask you this question' can be interchanged with 'So, remember the answer to this question well!'. We have 3 different sentences to precede the goal description and 5 different sentences for every explainable action to follow it. We implemented the explanations in three languages (English, Dutch, Italian), which is a strong proof of concept that similar implementation is possible in at least a large set of languages.

In addition, this implementation allowed us to differentiate the *style* of the explanation. We consider *cognitive explanations* and *affective explanations*. The cognitive explanations are phrased like above, affective explanations use emotion words in the phrasing of the explanations. For example, we can exchange the sentence 'I want to' with 'It would make me happy if you'. In that way, the full explanation becomes: 'It would make me happy if you learn how to recognise that your blood sugar level might be too high (hyper), and what you should then do. That is why I ask you this question.' This shows that this implementation enables providing, with a very simple manipulation of the sentence generation, explanations in different styles.

IV. Method

We evaluated the different explanation styles in a long-term (2.5 - 3 months) experiment.

A. Participants

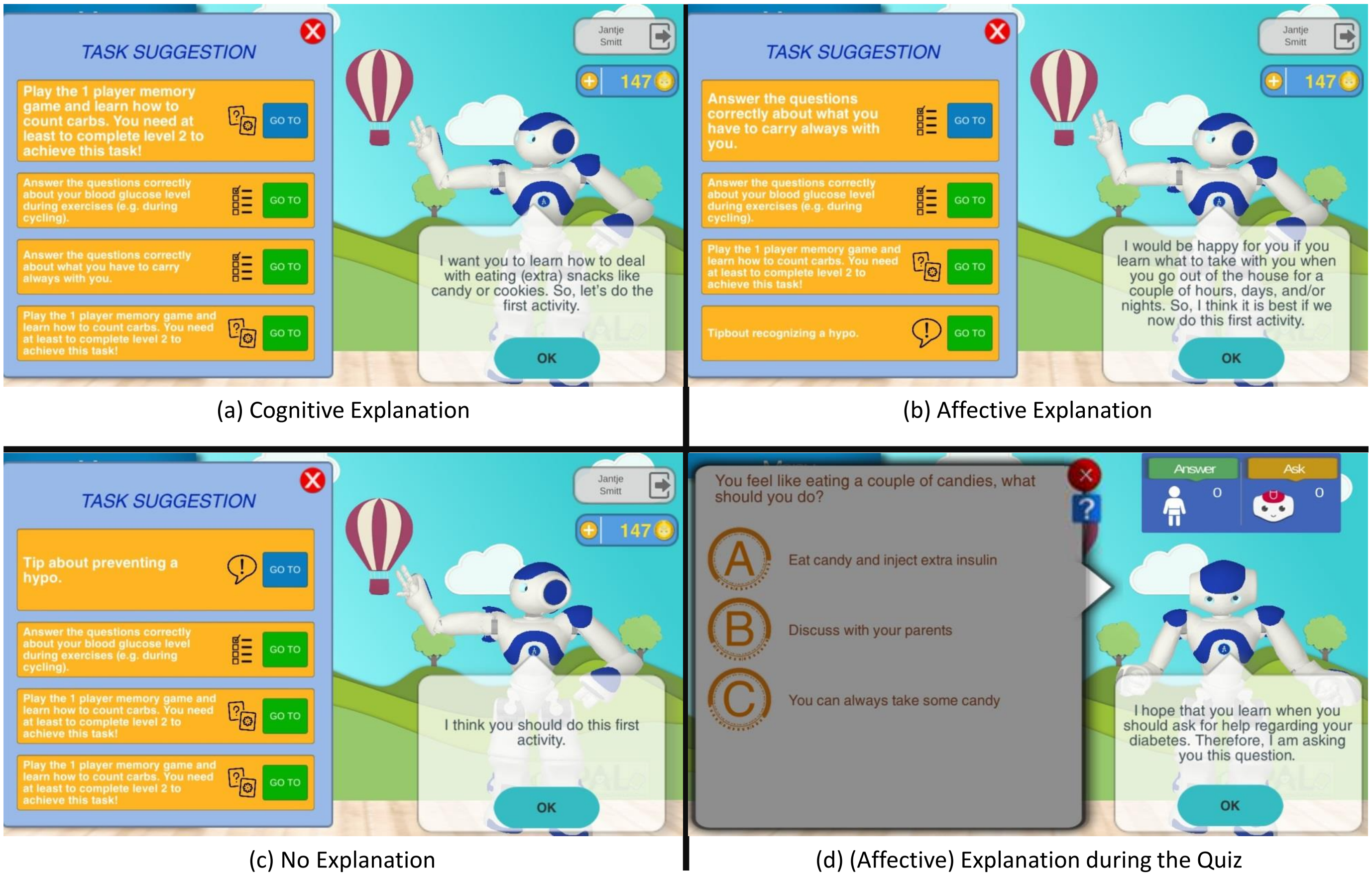
In total there were 48 (25 Dutch and 23 Italian) children with T1DM aged 6-14. The children were recruited via hospitals in the Netherlands and in Italy. There were no consequences to dropping out intermediately.

B. Experimental Design

When a child logs into the system (s)he is set to an initial experimental condition randomly. There are three possible conditions, *Cognitive Explanations*, *Affective Explanations*, and *No Explanations*. The children rotate between the three conditions (within-subjects testing).

It was not possible to test our hypothesis between subjects in this particular experiment. This experiment is part of a larger project where multiple experiments have been tested simultaneously. A requirement was therefore that all children would see the same content in the system. This meant that it was not possible to distribute the conditions randomly over the children and then keep them in that condition.

There were two phases of the experiment. The system had some small differences in the two phases. Task suggestions are only given in the second phase. Quizzes and tips were given in both phases. Furthermore, there were minor changes between the phases in activities without explanations. The experimental conditions switched per week in the first phase and per log-in in the second phase. We changed this in the second phase because many children used the system actively for only one or



Play the 1 player memory game and learn how to count carbs. You need at least to complete level 2 to achieve this task!

Answer the questions correctly about your blood glucose level during exercises (e.g. during cycling).

Answer the questions correctly about what you have to carry always with you.

Play the 1 player memory game and learn how to count carbs. You need at least to complete level 2 to achieve this task!

I want you to learn how to deal with eating (extra) snacks like candy or cookies. So, let's do the first activity.

(a) Cognitive Explanation

Answer the questions correctly about what you have to carry always with you.

Answer the questions correctly about your blood glucose level during exercises (e.g. during cycling).

Play the 1 player memory game and learn how to count carbs. You need at least to complete level 2 to achieve this task!

Tipbout recognizing a hypo.

I would be happy for you if you learn what to take with you when you go out of the house for a couple of hours, days, and/or nights. So, I think it is best if we now do this first activity.

(b) Affective Explanation

Tip about preventing a hypo.

Answer the questions correctly about your blood glucose level during exercises (e.g. during cycling).

Play the 1 player memory game and learn how to count carbs. You need at least to complete level 2 to achieve this task!

Play the 1 player memory game and learn how to count carbs. You need at least to complete level 2 to achieve this task!

I think you should do this first activity.

(c) No Explanation

You feel like eating a couple of candies, what should you do?

A Eat candy and inject extra insulin

B Discuss with your parents

C You can always take some candy

I hope that you learn when you should ask for help regarding your diabetes. Therefore, I am asking you this question.

(d) (Affective) Explanation during the Quiz

Fig. 1. Four screen-shots of the PAL system. Screen-shots (a-c) show task suggestions, screen-shot (d) shows the quiz. During a task suggestion the user is always shown a list of four possible tasks. The top-most task is then suggested by the PAL agent as being the 'best' to do at the current time. In screen-shot (a) the PAL agent explains why it is a good task to do by providing a cognitive explanation, in (b) it provides an affective explanation, and in (c) it provides no explanation for its suggestion. Finally, screen-shot (d) shows an example of an affective explanation given during the quiz.

two weeks, which causes them to not have enough exposure to the different conditions. Children that participated in the first phase were allowed to do so again in the second phase. 4 children (Dutch) and 9 children (Italian) did both phases.

Finally, both cognitive and affective explanations can be offered to the children in two different ways. 1) On the initiative of the PAL agent. Meaning the PAL agent simply gives the explanation for its behaviour. 2) On the initiative of the child. Meaning the system shows a question mark. The child can choose to press the question mark of his/her own accord. See figure 1.d for an example during the quiz.

Task suggestions are *always* explained when the child is in the cognitive or affective condition, and they are always explained on the initiative of the PAL agent. For the quiz and the tips the PAL agent provides explanations automatically 20% of the time. The other cases the child is shown a question mark. There is an exception to this. When the quiz is opened through the task suggestions rather than manually, then all questions in the quiz are for the same underlying goal which has already been mentioned during (the explanation for) the task suggestion itself. The explanations for the questions would always have the exact same content. Questions during a quiz opened in this way always only show a question mark.

C. *Measures and Variables*

For hypothesis 1, we test how often children request explanations of their own accord. We count how often children press the question marks (visible during the quiz and the tips) given an explanation condition (cognitive or affective). There is no measure in the no explanation condition since children cannot request explanations in that condition.

For hypothesis 2, we count the number of questions a child answered before closing the quiz. We then compute the average quiz length in the different styles for that child.

For hypothesis 3, the number of times the child manually request a 'next tip'. When the child receives a tip of the day, then the child can choose to either close the screen or press the 'next tip' button. We compute the average of next tip presses in the different styles for that child.

For hypothesis 4, we test whether children are more inclined to follow task suggestions in the different conditions (cognitive, affective, and lack of explanations). When presented with a task suggestion, the child can accept the suggestion by pressing the top-most task in the screen (see figure 1), or the child can reject the suggestion by either closing the screen or choosing another task in the list. We log the child's decision and measure the percentage of times the child actually chooses the suggested task given the explanation condition.

D. Material & Set-Up

There are two main locations where children interact with the PAL system, at home and at the hospital. At the hospital, the children interact with a physical Nao robot from Aldebaran and the PAL system. There they interact with a Health-Care Professional (HCP) and a researcher present. At home, they get a tablet with a virtual avatar of the robot and the same health-care applications (quiz, sorting game, etc.). At home, they interact with the system individually.

E. Procedure

Children were first invited to come to a hospital. There they were introduced to the PAL agent and system. Together with the HCP, they set some specific goals to advance their self-management of their diabetes (e.g., 'learn to recognise when you might have a hypo'). The system shapes the activities and task suggestions to work towards those goals. At the end, the children were given the tablet with the avatar to take with them to their houses. For 2.5 to 3 months they could play with the PAL system as often and long as they wanted. At the end of the period, they were invited to the hospital again.

V. RESULTS

One child (out of 48) was excluded from analyses due to a glitch in the data caused by a system error. The remaining 47 children had an average of 19 log-ins (STD = 12.9, minimum = 1, maximum = 55). Only three children requested an explanation in both the cognitive and the affective style. In section VI, we discuss possible improvements on our method for addressing the first hypothesis in future work.

For the second hypothesis, a one-way within subjects (or repeated measures) ANOVA was conducted to compare the effect of (IV) explanation style (cognitive, affective, and no explanations) on (DV) the average length of the quiz measured by the number of questions. There was no significant effect of the IV explanation style, Wilks' Lambda = 0.88, F (2,19) = 1.319, p = .291.

For the third hypothesis, a one-way within subjects (or repeated measures) ANOVA was conducted to compare the effect of (IV) explanation style on (DV) how often children request another tip. There was no significant effect of the IV explanation style, Wilks' Lambda = 0.93, F (2,45) = 1.772, p = .182.

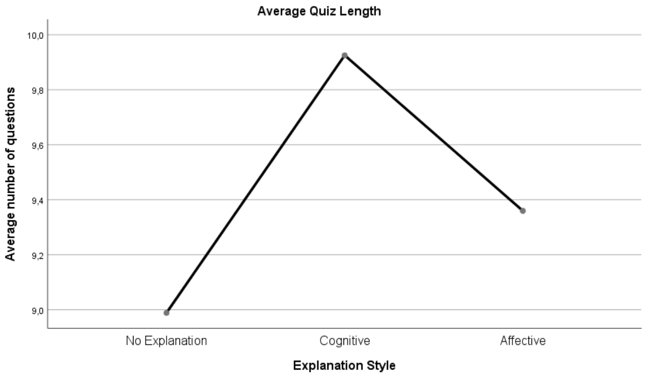


Fig. 2. The average number of questions per child and per style before children close the quiz in the different explanation styles.

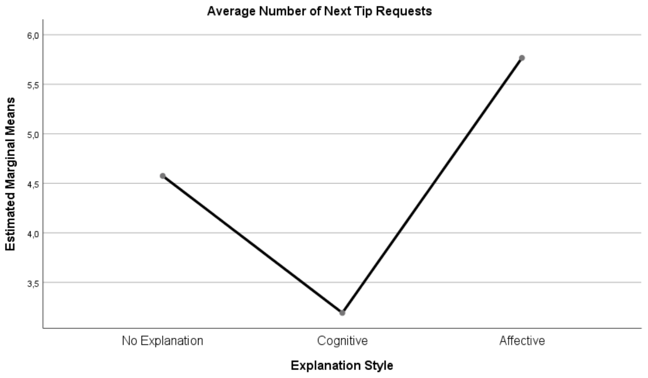


Fig. 3. The average number of times per child and per style that children requested a next tip in the different explanation styles.

Finally for the fourth hypothesis, a one-way within subjects (or repeated measures) ANOVA was conducted to compare the effect of (IV) explanation style on (DV) the percentage of task suggestions followed by the children. There was a significant effect of the IV explanation style, Wilks' Lambda = 0.60, $F(2, 13) = 4.285, p = .037$. In addition, three paired samples t-tests were used to make post hoc comparisons between conditions. A first paired samples t-test indicated that there was a significant difference in the percentage of task suggestions followed for no explanations ($M = 23\%$, $SD = 28\%$) and cognitive explanations ($M = 7\%$, $SD = 15\%$) conditions; $t(14) = 2.204, p = 0.045$. A second paired samples t-test indicated that there was a significant difference in the percentage of task suggestions followed for no explanations ($M = 23\%$, $SD = 28\%$) and affective explanations ($M = 11\%$, $SD = 27\%$) conditions; $t(14) = 2.505, p = 0.025$. A third paired samples t-test indicated that there was no significant difference in the percentage of task suggestions followed for cognitive explanations ($M = 7\%$, $SD = 15\%$) and affective explanations ($M = 11\%$, $SD = 27\%$) conditions; $t(14) = -0.501, p = 0.624$. With a LSD test these values are significant; however, if we consider a Bonferroni correction then the significance threshold is 0.0167. So, the ANOVA test shows that explanation style has an effect on the percentage of task suggestions followed by the children; however, the post hoc tests are inconclusive concerning the effect's direction.

We did an additional test where we combined the cognitive and affective conditions and compared the combined (any explanation) group against the no explanation group. A paired samples t-test indicated that there was a significant difference in the percentage of task suggestions followed for no explanations ($M = 23\%, SD = 28\%$) and any explanations ($M = 9\%$, $SD = 16\%$) conditions; $t(14) = 2.950, p = 0.011$. This final test indicates that providing *no explanations* for task suggestions correlates with children following the suggested tasks more often.

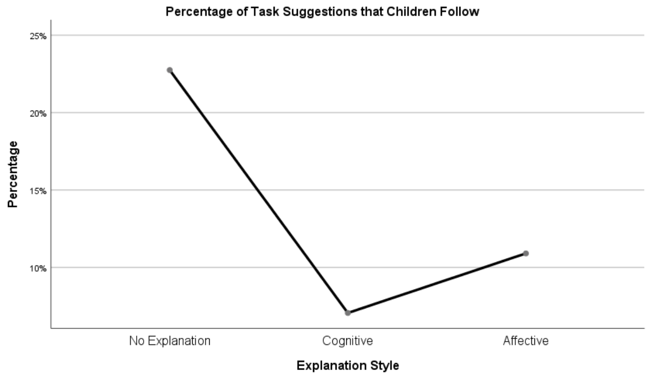


Fig. 4. The percentage of task suggestions that children follow in the different explanation styles.

VI. Discussion

The results come from a long-term 'in the wild' study. We recruited children aged 6-14 diagnosed with with T1DM. We are dealing with a real-world system (PAL) which is far more representative than a lab experiment could have been. However, this also means that the experiment was difficult to control. Children could stop the interaction with the system at any point in time. They could potentially request an explanation and close the application before the avatar could present it. Still, the system and the explanations were running robustly during the period of three months.

We found that explanation style influences how often children follow task suggestions. We found no further significant effects. This might be because the exposure of explanations during task suggestions was high. Every time children login the system the first thing they saw was a task suggestion which (in the cognitive and affective conditions) is always explained. During the quiz and the tip the explanations were not often explained in a forced manner. Most of the time, the children would only see a question mark that they could press of their own accord. The results show that children did not press the question marks often. Since children already see an explanation in 20% of the cases, a case might be added in future work where children get *no* forced explanations to prevent potential saturation effects.

We did not expect that the *no explanation* condition would correlate with task suggestions being followed more often. We offer three possible explanations for this. 1) A straightforward explanation is that children simply do not read the longer texts in explained task suggestion (see also figure 1 for examples of differently explained task suggestions). This would result in more randomly chosen tasks from the menu. This would mean that the in literature suggested length of explanations [8] is still too long when applying explanations in a long-term experiment with child users. 2) Another possibility is that children *do* read and understand the explanations but they sometimes think they already know what the task is supposed to teach them. For example, if the PAL agent says the child should do a quiz because it teaches the child how to recognise when one might have a hypo, and if the child thinks (s)he already knows this, then the child is more likely to choose another task instead. This would relate to literature about teaching and learning, where it is suggested that explaining the importance of educational material helps students to orient/ plan their behaviour better themselves [29]. This would imply there *is* a positive effect of explanation style on the child's behaviour in the system. 3) The child might sometimes get stubborn from the explanation. Thinking something along the lines of 'I don't feel like practising / doing that!'. Which causes them to choose different tasks.

Future work should determine the underlying mechanism of why certain explanation styles change the users' behaviour in long-term interaction. A possible approach is to (sometimes) 'ask' the users why they chose a particular task after their selection. This was not possible in the here presented work due to limitations imposed by the project; however, it is our recommendation for future long-term experiments in this area. Secondly, the work here indicates that there is insufficient knowledge on when and how affective explanations should be used. When varying the style (but not the content) of your explanations, then (in the long-term) this may only trigger subtle differences in the users. A formal model of *when a particular type of explanation is preferred* is beneficial for further research in this area as this enables testing such a model against randomly chosen styles.

VII. Conclusion

In this paper, we presented results from a long-term (2.5 - 3 months) experiment on the effect of explanations on the motivation of children to use an e-health system involving interaction with a virtual robot. We considered cognitive explanations (based on the beliefs and goals of the agent), affective explanations (also using emotions of the agent for generating the explanation), and no explanations (providing no explanations at all for the agent's behaviour). The explanations were implemented in an in-the-wild autonomous health-support application for children (aged 6-14) suffering from T1DM. We found that explanation style influences how often children follow task suggestions. Specifically, the results indicate that children follow the suggestions more often when *no explanation* is given. We found no other significant effects of explanations in this study. Although no effect was found of cognitive versus affective explanations, this is to our knowledge the first evidence that explanations impact long-term human-agent interaction and system usage. Our results also show that counter-intuitive effects of agent explanations may be expected when used with children, and, that more research is needed to understand why lack of explanations seems to correlate with following task suggestions.

References

[1] Iolanda Leite, Carlos Martinho, and Ana Paiva. Social robots for long-term interaction: a survey. *International Journal of Social Robotics*, 5(2):291–308, 2013.

[2] Dung N Lam and K Suzanne Barber. Comprehending agent software. In *Autonomous Agents and Multiagent Systems*, pages 586–593, 2005.

[3] Brian Y Lim, Anind K Dey, and Daniel Avrahami. Why and why not explanations improve the intelligibility of context-aware intelligent systems. In *Human Factors in Computing Systems*, pages 2119–2128, 2009.

[4] Steven R Haynes, Mark A Cohen, and Frank E Ritter. Designs for explaining intelligent agents. *International Journal of Human-Computer Studies*, 67(1):90–110, 2009.

[5] Peter Carey. *Data protection: a practical guide to UK and EU law*. Oxford University Press, Inc., 2018.

[6] Sule Anjomshoae, Amro Najjar, Davide Calvaresi, and Främling Kary. Explainable agents and robots: Results from a systematic literature review. In *Autonomous Agents and Multi-agent systems*, 2019 (in press).

[7] Mark G Core, H Chad Lane, Michael Van Lent, Dave Gomboc, Steve Solomon, and Milton Rosenberg. Building explainable artificial intelligence systems. In *Innovative Applications of Artificial Intelligence*, pages 1766–1773, 2006.

[8] Maaike Harbers, Joost Broekens, Karel Van Den Bosch, and John-Jules Meyer. Guidelines for developing explainable cognitive models. In *International Conference on Cognitive Modeling*, pages 85–90, 2010.

[9] Joost Broekens, Maaike Harbers, Koen Hindriks, Karel Van Den Bosch, Catholijn Jonker, and John-Jules Meyer. Do you get it? user-evaluated explainable bdi agents. In *Multiagent System Technologies*, pages 28–39. Springer, 2010.

[10] Sandy H Huang, David Held, Pieter Abbeel, and Anca D Dragan. Enabling robots to communicate their objectives. *Autonomous Robots*, pages 1–18, 2017.

[11] Frank Kaptein, Joost Broekens, Koen Hindriks, and Mark Neerincx. Personalised self-explanation by robots: The role of goals versus beliefs in robot-action explanation for children and adults. In *Robot and Human Interactive Communication (RO-MAN), 2017 26th IEEE International Symposium on*, pages 676–682. IEEE, 2017.

[12] Paul M Churchland. Folk psychology and the explanation of human behavior. *The future of folk psychology: Intentionality and cognitive science*, pages 51–69, 1991.

[13] Bertram F Malle. *How the mind explains behavior: Folk explanations, meaning, and social interaction.* MIT Press, 2004.

[14] Sabine A Döring. Explaining action by emotion. *The Philosophical Quarterly*, 53(211):214–230, 2003.

[15] Frank Kaptein, Joost Broekens, Koen Hindriks, and Mark Neerincx. The role of emotion in self-explanations by cognitive agents. In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, pages 88–93. IEEE, 2017.

[16] Jingting Wang, Yuanyuan Wang, Chunlan Wei, Nengliang Yao, Avery Yuan, Yuying Shan, and Changrong Yuan. Smartphone interventions for long-term health management of chronic diseases: an integrative review. *Telemedicine and e-Health*, 20(6):570–583, 2014.

[17] Jamy Li, René Kizilcec, Jeremy Bailenson, and Wendy Ju. Social robots and virtual agents as lecturers for video instruction. *Computers in Human Behavior*, 55:1222–1230, 2016.

[18] Maartje De Graaf, Somaya Ben Allouch, and Jan Van Dijk. Why do they refuse to use my robot?: Reasons for non-use derived from a long-term home study. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pages 224–233. ACM, 2017.

[19] Ning Wang, David V Pynadath, and Susan G Hill. Trust calibration within a human-robot team: Comparing automatically generated explanations. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*, pages 109–116. IEEE Press, 2016.

[20] John D Lee and Katrina A See. Trust in automation: Designing for appropriate reliance. *Human factors*, 46(1):50–80, 2004.

[21] Bonnie M Muir. Trust in automation: Part i. theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics*, 37(11):1905–1922, 1994.

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

[23] Cristina Conati and Kurt 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.

[24] Daniel C Dennett. Three kinds of intentional psychology. In R. Healey, editor, *Reduction, Time and Reality*, pages 37–61. Cambridge University Press, Cambridge, 1981.

[25] Russell Beale and Chris Creed. Affective interaction: How emotional agents affect users. *International journal of human-computer studies*, 67(9):755–776, 2009.

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

[27] Stephen Lieb and John Goodlad. Principles of adult learning, 2005.

[28] Frank C Keil. Explanation and understanding. *Annual Review of Psychology*, 57(1):227–254, 2006.

[29] Jan D Vermunt and Nico Verloop. Congruence and friction between learning and teaching. *Learning and instruction*, 9(3):257–280, 1999.