# The situated Design Rationale
# of a social robot for child's disease self-management

Looije, Rosemarijn

**DOI**
10.4233/uuid:1de25f4d-c0cb-42d6-9685-d75b300c0aad

**Publication date**
2019

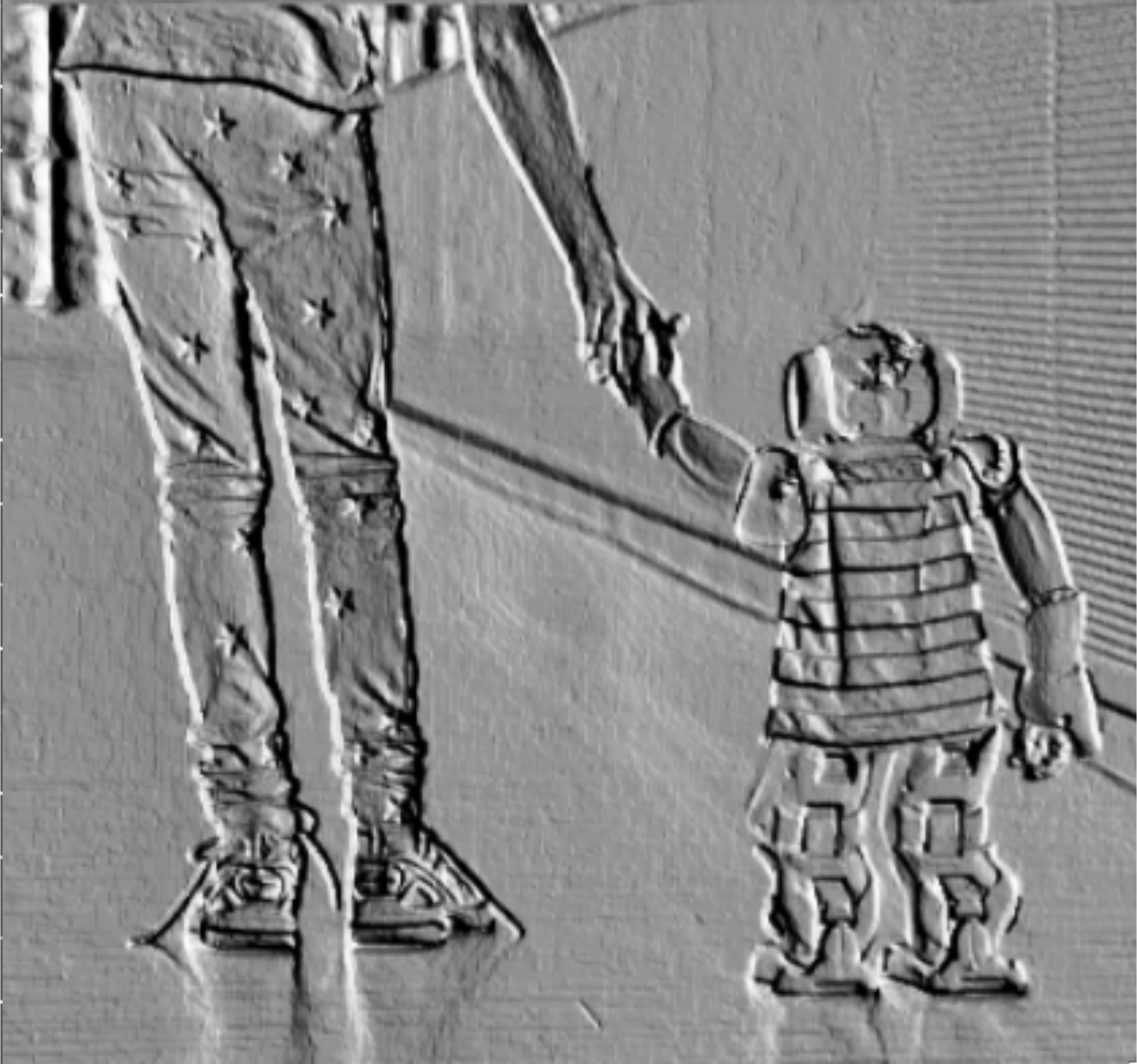
**Document Version**
Final published version

**Citation (APA)**
Looije, R. (2019). *The situated Design Rationale: of a social robot for child's disease self-management*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:1de25f4d-c0cb-42d6-9685-d75b300c0aad

# The situated design rationale

of a social robot for child's disease self-management

Rosemarijn Looije



# Propositions

accompanying the dissertation

**THE SITUATED DESIGN RATIONALE**
OF A SOCIAL ROBOT FOR CHILD'S DISEASE SELF-MANAGEMENT

by

**Rosemarijn LOOIJE**

1. A situated Design Rationale supports design and evaluation of behavioral change support systems (this thesis)
2. Emotional expressivity of a social robot supports engagement, motivation and performance (this thesis)
3. A social robot that adapts its interaction to the child is motivating and increases performance (this thesis)
4. A social robot that discloses information and expresses emotions supports openness of the children it interacts with (this thesis)
5. A robot pal for children reduces parental stress (experiments)
6. A social robot stimulates interaction between robot and child (feedback)
7. The rise of the robots will free time of people to interact with each other
8. A falling robot divides humans in two groups, those with and those without empathy
9. We can't ever build a robot that will be even as good as a human being in anything that counts, let alone better (Asimov)
10. Parenthood changes your (research) perspective, and is totally worth it!

These propositions are regarded as opposable and defendable, and have been approved as such by the promotor prof. dr. M.A. Neerincx.

# The situated Design Rationale

## of a social robot for child's disease self-management

# THE SITUATED DESIGN RATIONALE

## OF A SOCIAL ROBOT FOR CHILD'S DISEASE SELF-MANAGEMENT

**Proefschrift**

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof.dr.ir. T.H.J.J. van der Hagen,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op
donderdag 27 juni 2019 om 15:00

door

**Rosemarijn LOOIJE**

doctorandus in de Kunstmatige Intelligentie,
Rijksuniversiteit Groningen, Groningen, Nederland
geboren te Haarlem, Nederland

Dit proefschrift is goedgekeurd door de promotoren.

Samenstelling promotiecommissie bestaat uit:

| | |
|---|---|
| Rector Magnificus, | voorzitter |
| Prof.dr. M.A. Neerincx | Technische Universiteit, promotor |
| Prof.dr. K.V. Hindriks | Technische Universiteit Delft, promotor |
| | |
| *Onafhankelijke leden:* | |
| Prof.dr. C.M. Jonkers | Technische Universiteit Delft |
| Prof.dr.ir. R.H.M. Goossens | Technische Universiteit Delft |
| Prof.dr. P. Markopoulos | Technische Universiteit Eindhoven |
| Prof.dr. V. Evers | Universiteit van Twente |
| Prof.dr. T. Belpaeme | Universiteit van Gent, België |

*Keywords:* Social robot, Cognitive engineering, Design rationale, Diabetes, Children

*Printed by:* GVO drukker

*Front & Back:* Beeld uit film van Maaike Broos https://vimeo.com/105283635
Bewerking Rosemarijn Looije



ISBN 978-94-6366-157-7

An electronic and interactive version of the situated Design Rationale visualization is available at
https://bit.ly/2RXxWNd.
An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

*We can't ever build a robot that will be even as good as a human being in anything that counts, let alone better. We can't create a robot with a sense of beauty or a sense of ethics or a sense of religion. There's no way we can raise a positronic brain one inch above the level of perfect materialism.*
*We can't, damn it, we can't. Not as long as we don't understand what makes our own brains tick. Not as long as things exist that science can't measure. What is beauty, or goodness, or art, or love, or God?*

Elijah Baley in "The caves of steel, a robot novel." (Isaac Asimov, 1954).

# Contents

# Summary

A young boy with type 1 Diabetes Mellitus is supported by a social robot on the road to self-management. The robot has knowledge on the goals that the boy needs to reach, as discussed with his health care professional. The robot also knows the boy's activity options and preferences. It suggests activities based on this knowledge, but also encourages the boy to try new approaches. For parents, such a social robot means that they can be less teacher and more parent and for the health care professionals it means they can focus on the emotional aspects instead of the knowledge aspects during visits. Finally, the boy sees the robot as something that is fun and a peer in contrast to someone/something with a higher authority. The robot supports relatedness and a feeling of competence, the different activities provide a feeling of autonomy, and less budding of the parents reduces stress for the whole family. This all supports that the boy sees diabetes as his own responsibility and feels that he has enough competence and autonomy to take care of diabetes himself. In support of this vision we look in this thesis at the design and evaluation of a social robot.

This design and evaluation needs to be done in a coherent and precise manner. What concepts are relevant and how are these related to each other? In chapter 2 we describe the situated Design Rationale (sDR), which is an extension of the situated Cognitive Engineering method [172]. sDR makes it possible to keep track of decisions during the design phase, in which decisions on objectives and methods are related to functions and interaction design patterns which in their turn are related to the expected effects, that contribute to the objectives, and instruments to measure these.

The objectives and methods were chosen based on knowledge of children, diabetes, and behavior change, and of course from interactions with the main stakeholders (children, parents and health care professionals). Self Determination Theory (SDT) [74]was the best fit for the objectives. It is used for behavior change and education, and for the same age group as we are interested in: 7-10. It says a solution should support a feeling of competence (having enough knowledge to deal with problems), a feeling of autonomy (having the opportunity to choose for themselves), and a feeling of relatedness with the "teacher". These three objectives were used as a starting point to decide on methods that support these and from there choose the functions and related interaction design patterns that can be programmed in the robot and (partially) implement the methods. Of course, the successfulness of the implementation in relation to the expected effects due to the chosen objectives and methods should be evaluated.

First, we evaluated the effect of emotions and embodiment on emotion recognition, engagement, motivation and performance. Emotions are an important part of social behavior that is needed to make the robot succeed in the relatedness objective. The three evaluations we did, showed the importance of keeping track of your design decisions. The choices about which interaction design patterns are used influence the expected results and provide feedback on improving the shape of the functions.

Second, we looked at more complex interactions related to adaptivity and self-disclosure by the robot. Adaptivity to the child on an educational game to support competence, adaptivity of emotional expressions to express more social behavior and contribute to relatedness. Finally, the self-disclosure by the robot was also expected to increase self-disclosure by the child and therefore support relatedness. Results supported our hypotheses.

Third, two evaluations were performed in which the robot engaged in several activities with the child, are presented. These showed that the sDR had an added value in showing the relations between the different functions and use cases, but that there was a too large variety between the participating children to conclude anything. This is a problem that will be difficult to overcome within this user group.

This thesis shows that the sDR method could be applied in all evaluations and that within a project the main objectives and methods stay the same and therefore also the expected effects and instruments. This supports the re-usability. The main differences between evaluations are in the use cases, functions and interaction design patterns. By presenting each evaluation in the same format and being able to concatenate them, the differences and similarities can be found in an easy manner which contributes to finding missing parts and theory forming; i.e. when different evaluations look at the same method, function relation and the results support each other.

When applying the sDR method to other user groups the methods and therefore functions might change, but the objectives can stay the same. On the other hand when the whole application changes the functions can still be used, but with a complete different foundation (objectives and methods). By iteratively evaluating a complete system (robot) that is adapted to the evaluation results, with the expected end users in their own environment, we develop a social robot that supports our vision.

# Samenvatting

Een jongetje met Type 1 Diabetes Mellitus wordt ondersteund door een sociale robot in zijn tocht naar zelfmanagement. De robot heeft kennis over de doelen die het jongetje moet bereiken. De doelen zijn besproken met de zorgprofessional. De robot weet ook activiteit opties en voorkeuren van het jongetje. Het stelt activiteiten voor gebaseerd op deze kennis, maar stimuleert het jongetje ook om nieuwe aanpakken te proberen. Voor ouders betekent zo'n sociale robot dat ze minder docent en meer ouder kunnen zijn. Voor zorgprofessionals betekent het dat ze op de emotionele aspecten kunnen focussen tijdens afspraken in plaats van op kennis aspecten. Tot slot, de jongen ziet de robot als iets leuks en gelijkwaardigs in plaats van als iemand/iets met een hogere autoriteit. De robot ondersteunt het hebben van een band en een gevoel van competentie, de verschillende activiteiten geven een gevoel van autonomie, en minder bemoeienis van de ouders vermindert de stress van de hele familie. Dit allemaal ondersteunt de jongen zodat hij inziet dat de diabetes zijn eigen verantwoordelijkheid is en dat hij voelt dat hij genoeg weet en kan om zelf zorg te dragen voor zijn diabetes. Om deze visie te ondersteunen kijken we in dit proefschrift naar het ontwerp en de evaluatie van een sociale robot.

Het ontwerp en de evaluatie moet precies gebeuren en op samenhangende wijze. Welke concepten zijn relevant en hoe hangen deze samen? In chapter 2 beschrijven we de situated Design Rationale (sDR), dit is een uitbreiding van de situated Cognitive Engineering method [172]. sDR maakt het mogelijk om beslissingen bij te houden tijdens het ontwerpproces. Beslissingen over doelen en methodes zijn gerelateerd aan functies en interactie ontwerppatronen, die op hun beurt weer gerelateerd zijn aan de verwachtte effecten. Deze effecten dragen bij aan de doelen, en er zijn instrumenten om deze te meten.

De doelen en methodes zijn gekozen op basis van kennis over kinderen, diabetes, gedragsverandering, en natuurlijk onze gesprekken met de belangrijkste belanghebbenden (kinderen, ouders en zorgprofessionals). Self Determination Theory (SDT) [74] paste het best bij onze doelen. Het is toegepast voor gedragsverandering en onderwijs, en ook in dezelfde leeftijdscategorie als waar wij naar kijken: 7-10. De theorie zegt dat een oplossing een gevoel van competentie moet ondersteunen (genoeg kennis hebben om te kunnen omgaan met de problemen), een gevoel van autonomie moet geven (de kans hebben om zelf te kiezen), en een gevoel van band met de önderwijzer"moet ondersteunen. Deze drie doelen werden gebruikt als startpunt om methodes te kiezen die deze ondersteunen en vanuit daar de functies en gerelateerde ontwerppatronen te kiezen die in een robot geprogrammeerd kunnen worden en (gedeeltelijk) de methodes implementeren. Natuurlijk moet het succes van de implementatie in relatie tot de verwachtte effecten door de gekozen doelen en methodes geëvalueerd worden.

Als eerste evalueerden we het effect van emoties en fysieke vorm op emotieherkenning, betrokkenheid, motivatie en prestatie. Emoties zijn een belangrijk deel van sociaal gedrag dat nodig is om te slagen in het doel om een band op te bouwen met de robot. De drie evaluaties die wij hebben gedaan, lieten zien dat het belangrijk is de ontwerpbeslissingen bij te houden. De keuzes over welke interactie ontwerppatronen zijn gebruikt beïnvloeden de verwachtte resultaten en geven feedback over het verbeteren van de vorm van de functies.

Daarna keken we naar meer complexe interacties die gerelateerd waren aan adaptiviteit en praten over zichzelf door de robot. Adaptiviteit naar het kind tijdens een educatief spel om het gevoel van competentie te ondersteunen, adaptiviteit van emotionele expressies om meer sociaal gedrag te uiten en bij te dragen aan een gevoel van band. Tenslotte, het was de verwachting dat een kind meer over zichzelf zou praten en een groter gevoel van band met de robot zou krijgen als de robot ook over zichzelf praatte. De resultaten ondersteunden onze hypotheses.

Als laatste beschrijven we twee experimenten waarbij de robot meerdere activiteiten deed met het kind. Deze laten zien dat sDR een toegevoegde waarde heeft in het laten zien van de verschillende functies en gebruiksscenario's. Maar ook dat er een te grote variatie was tussen de kinderen om conclusies te kunnen trekken. Dit is een probleem dat moeilijk op te lossen is met deze gebruikersgroep.

Dit proefschrift laat zien dat de sDR methode toegepast kan worden in alle evaluaties en dat binnen een project the hoofd doelen en methodes hetzelfde blijven, en hierdoor ook de verwachtte effecten en instrumenten. Dit ondersteunt de herbruikbaarheid. The grootste verschillen tussen de evaluaties zitten in de gebruiksscenario's, functies en interactie ontwerppatronen. De verschillen en overeenkomsten tussen verschillende evaluaties kunnen gemakkelijk gevonden worden door iedere evaluatie op dezelfde wijze te presenteren. Dit draagt bij aan het vinden van missende onderdelen en het vormen van theorieën. Bijvoorbeeld wanneer verschillende evaluaties naar dezelfde methode kijken dan ondersteunen de functie relatie en de resultaten elkaar.

Wanneer we de sDR methode toepassen bij andere gebruikersgroepen dan kunnen de methodes en daardoor de functies veranderen, maar de doelen kunnen hetzelfde blijven. Aan de andere kant, wanneer de hele toepassing verandert dan kunnen de functies misschien nog gebruikt worden, maar met een compleet andere achtergrond (doelen en methodes). Door iteratief het hele systeem te evalueren en aan te passen aan de evaluatie resultaten, met de verwachtte eindgebruikers in hun eigen omgeving, kunnen we een sociale robot ontwikkelen die onze visie ondersteunt.

# 1 | Introduction

If you look at the media you would expect that social robots are already everywhere for the general public or will be in the near future. Robots are envisioned as (play) companions[1], or as actors that can change or support behavior. Examples are education [236], support of older adults with daily tasks and regular exercises, so they can live longer at home[2], and support of children with, for instance, autism [204] or diabetes [18].

There is much scientific research on social robots, with conferences specifically on this topic (Ro-Man, HRI, ICSR, New Friends), next to sessions and papers in other conferences. Furthermore, there are multiple journals that present social robot papers independently from a specific domain (e.g. IJHCS, IJSR, JHRI, IEEE Transactions on SMC Part A), and of course social robot papers are presented in journals related to the research domain (e.g. education, behavior change, autism journals). Notwithstanding the efforts to bring social robot research together, the research itself is still fragmentary due to the plethora of used robots, objectives, methods and application domains.

Both for the general public and research community the expectation is that social robots will fulfill a need, in for instance behavior change as this is one of the big societal challenges. Behavior change to improve self-management is important for many lifestyle related illnesses as obesity, diabetes, and asthma. Social robots have been shown to have positive effects on changing behaviors with autistic children [204]. Education, feeling of competence and relatedness are important components for changing behavior [214] and social robots have been shown to have a positive effect on education [236]. A persistent change requires that the human-robot interaction supports a feeling of autonomy, competence and relatedness with the robot. But although there are ideas on where social robots can contribute, we actually do not know how to provide this support. We don't know the exact support needs and also don't have a clear idea yet on how the needs we already know can be precisely implemented in the current generation of social robots. There is a lot of potential in the social robots that are now available, but the last years show that it is hard to reach their full potential. As many see the possible promises of social robots, there are many initiatives in creating them, making it a complex market. Some social robots can be bought off-the-shelf, but they do not yet come with all the required behavior, which is up to e.g. research institutes to develop (e.g. Aldebaran's NAO, Pepper). Others are completely developed at research institutes (e.g. Kaspar (University of Hertfordshire), Simon (Georgia Tech)). And there are even others that are developed at commercial companies as research projects surrounded with quite some secrecy on functionalities (e.g. Asimo from Honda). This complicates the implementation of functionalities that adhere to the different support needs of different user groups and different applications.

We don't have a systematic overview of what functions different social robots have,

[1] http://2machines.com/185217/

[2] http://www.robots.nu/robot-lea-wint-livewire-award-2015/

what their effects are on the users and how this compares to other robots. To get from loosely coupled papers to a theory of social robotics, there is a need for consistent, coherent design and evaluation of social robots. As mentioned there are many objectives for which social robots can be used. The best design of functions in one domain might not be the best solution for another domain. The aim of this thesis is to contribute to a more unified and comprehensive theory of robot design that will inform us what works for specific domains or across domains, with a specific focus on behavior change.

# Vision

Daniel is 8 and has Type 1 Diabetes since he was 6. His mother changed jobs, so that she could take care of Daniel during the day. Because of the huge dependency on his mother, all these years Daniel is not stimulated to take care of his diabetes by himself. In addition, as she is used to do it all, his mother finds it really hard to trust Daniel. Both know there is a need for less hoovering and more indepency, but how to go about this?

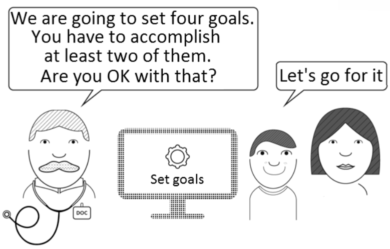



Figure 1.1: Setting objectives

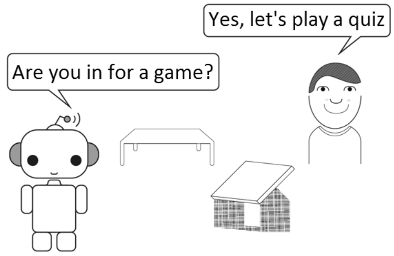



Figure 1.2: Play with the robot

Daniel, his mother and his health care professional sit together to set objectives related to Daniel his self-management that he should work on (Figure 1.1). The health care professional then shows his mother what she can see in the program when they are at home, e.g. progress on the objectives, blood glucose measurements, but also information about diabetes. Meanwhile, Daniel is off to interact with a robot at the hospital. The robot knows about the objectives (e.g. learn about high/low carbohydrate foods) and suggests to play a quiz to support reaching these objectives (Figure 1.2). As the robot is not easy to take everywhere at home and at school, the robot's avatar will support Daniel on a phone or tablet by filling his glucose values and play for instance a quiz with him (Figure 1.3). His glucose values are imported from his meter to the tablet/phone and he can add events and discuss his day with the avatar. Next to this, he can do different activities with the avatar, similarly to the ones he can do with the robot (e.g. quiz). The avatar stimulates him to do relevant activities to reach his objective. It, for instance, suggests to play a game in which foods have to be ordered by number of carbohydrates to improve his knowledge on carbohydrates and provides positive feedback about his progress towards the objectives (Figure 1.4). Furthermore, the avatar adapts to Daniel by getting to know his preferences so it can suggest activities preferred by Daniel or motivate to do something else by acknowledging his preference (e.g. "I know you would prefer to do the quiz, we can do that, but it would be better if you first fill in the timeline to achieve your objective"). By getting to know Daniel and interacting regularly the robot becomes a pal.

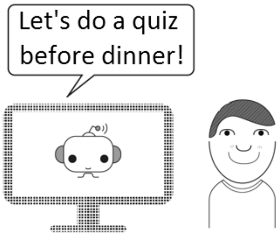



Figure 1.3: Play with avatar

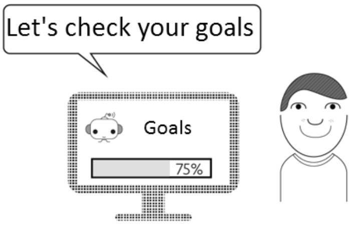



Figure 1.4: Visible, up-to-date, progress

Meanwhile, his mother can see the glucose values and, if approved by Daniel, other information as activities, diary input and progress on the objectives (Figure 1.5). This should support the parent in letting go.
In between visits, objectives can be refined or changed. During visits the progress is discussed and new objectives are set. There is also time to continue interaction with his pal (e.g. while his mother speaks with the health care professional), but now in robot embodiment. (Figure 1.6).


Figure 1.5: Information for parents

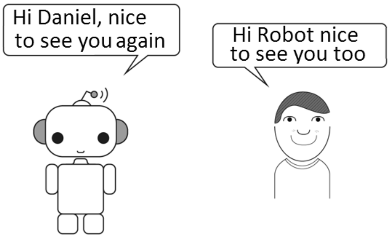



Figure 1.6: Return visits

Over time Daniel and his mother balance the self-management tasks and both get more confidence that everything will be all right when Daniel takes care of himself. Daniel is more secure on what he knows and can do because he reached the objectives. This is further encouraged by the increased trust his mother has in Daniel's capabilities to perform self-management tasks by himself.

Adherence is very important when you have a chronic illness, and oftentimes this means to change your behavior. Examples of lifestyle related diseases in which behavior change is of importance are obesitas, Type 2 Diabetes Mellitus and Type 1 Diabetes Mellitus (T1DM). Children with T1DM, like Daniel (see Figure 1.1), need to take actions at least every time around mealtime, each day, the whole year round until their death. What makes it especially hard is that not taking the best care of T1DM does not have immediate effects, but can lead to complications (e.g. eye disease, cardiovascular disease) later in life and/or early death. This is not only stressful for the children but also for the parents, who find it hard to trust their children with their own health regime because they also know about the possible complications. It is important though that people with a lifestyle related disease are responsible themselves, because then the intrinsic motivation can get and stay high enough to follow the regime their whole life. To reach this level of intrinsic motivation, support provided by the surroundings is indispensable, this support should be provided by parents, health care providers, sport coaches etc., but we think a social robot can also play a part in this. Children are more prone to trust a robot and accept it as a pal that can support them than adults.

In this dissertation we will thus focus on a social robot that supports children with a lifestyle related disease, T1DM, by being their pal and implementing functions derived from behavior change methods that contribute to the Self Determination Theory (SDT) behavior change objectives; autonomy (put the child in charge of themselves - interaction with the robot and related to their illness), competence (by having the robot educating and the health care professional providing personal objectives) and relatedness (by the robot being a pal). The improved self-management of the child is expected to improve the trust of the parents in their child and reduce their stress.

The following section will provide an overview of the state of the art with respect to behavior change methods (subsection 2.2.1) and social robots (subsection 2.2.2). Following this, an iterative design and evaluation methodology that is used as a basis in this dissertation is explained (subsection 2.2.4).

## 1.1. BACKGROUND

### 1.1.1. TYPE 1 DIABETES MELLITUS

There are two types of diabetes, Type 1 and Type 2 [109]. Type 1 Diabetes Mellitus (T1DM) is a result from destruction of the insulin-producing cells in the pancreas by the autoimmune system. Type 1 typically presents itself at a young age. Type 2 diabetes is a metabolic disorder where the body still makes insulin, but not enough and it's not absorbed well. Type 2 diabetes often occurs at a later age.

We will focus on T1DM, because that is the type that is most prevalent in children and the incidence is rising [187]. For these children it is very important to keep their blood glucose levels as steady as possible. To attain this objective, children and their social environment (parents, teachers, siblings, friends etc.) need to have knowledge and skills to manage the disease. Examples are: Regularly measure blood glucose levels, counting of carbohydrates, calculating needed insulin and injecting with a pen or setting the bolus provision of the pump. During these example activities they need to take into account the (interactive) effects of food intake, physical exercise, mental stress and hormones.

Furthermore, a child and his or her environment need to be able to recognize symptoms of high and low blood glucose to act accordingly. Even when managed properly, a child will have periods of high imbalance due to for instance illness, hormones or growth spurts.

Research suggests that high family stress negatively affects glycemic control [246]. To lower family stress it is important that children learn to manage their illness at a young age and that parents let them do this. For this the behavior of child and family should change and behavior change theories can provide clues on what aspects (e.g. feeling of competence) are relevant for children with T1DM.

The effects of T1DM, even with our modern treatment, are quite severe. More than 50% of the children develop complications with regard to major organs, like the heart and blood vessels, 12 years after diagnosis [63]. The life expectancy of children diagnosed by age 10 is 19 years shorter than of healthy children [169]. There are also effects on psychological well-being, on feelings of embarrassment and on school performance [190].

The lifetime costs of children with diabetes are much higher than those of healthy children. In the US, for instance, the lifetime costs of children diagnosed between the ages of 3-9 are an estimated $746 million in medical costs and $1208 million in income loss[3]. These numbers exclude the costs related to parents and siblings, e.g. parents taking other jobs/quitting to take care of their child, increased stress [33], and chance on burnouts[4].

### 1.1.2. BEHAVIOR CHANGE

The domain in which we want to support behavior change, children in the age of 7-14 with T1DM, is used to focus the behavior change study. Behavior change is a large research field, in which the choice of behavior change theory guides decisions on functionality. We have chosen Self-Determination theory (SDT) [74] as behavior change theory as several aspects that are seen as important in social robot interaction are also seen as important in SDT (e.g. trust and likeability). Furthermore, SDT is not exclusively used in the behavioral change domain but also in education [175], for children in the relevant age group (7-11) [223] and in games, where it showed to be a predictor of enjoyment and future game play [215]. In chapter 2 this decision is substantiated more elaborately.

SDT [74] is a motivational theory that supports a continuum of motivation, from external regulation (completely extrinsic) via more and more internally motivated to finally reach intrinsic motivation [213]. The motivation can be influenced by supporting three basic psychological needs (see Figure 1.7): (1) autonomy, (2) competence and (3) relatedness. Autonomy is about the willingness and opportunity to do a task, competence is the need for challenge and feeling of ability, and relatedness refers to the connection with others [73]. Long-term interaction with the interaction partner (e.g. therapist, social robot) is seen as a prerequisite for behavior change in the long run and, therefore, several behavior change methods state that there is a need for a bond with the interaction partner (e.g. Motivational Interviewing [161]).

Within SDT different behavior change techniques (BCTs) are used. Many of these

[3] http://outpatient.aace.com/type1-diabetes/the-burden-of-type-1-diabetes
[4] https://www.dvn.nl/dvn/nieuws/2061/ouders-van-kinderen-met-diabetes-onder-zware-druk

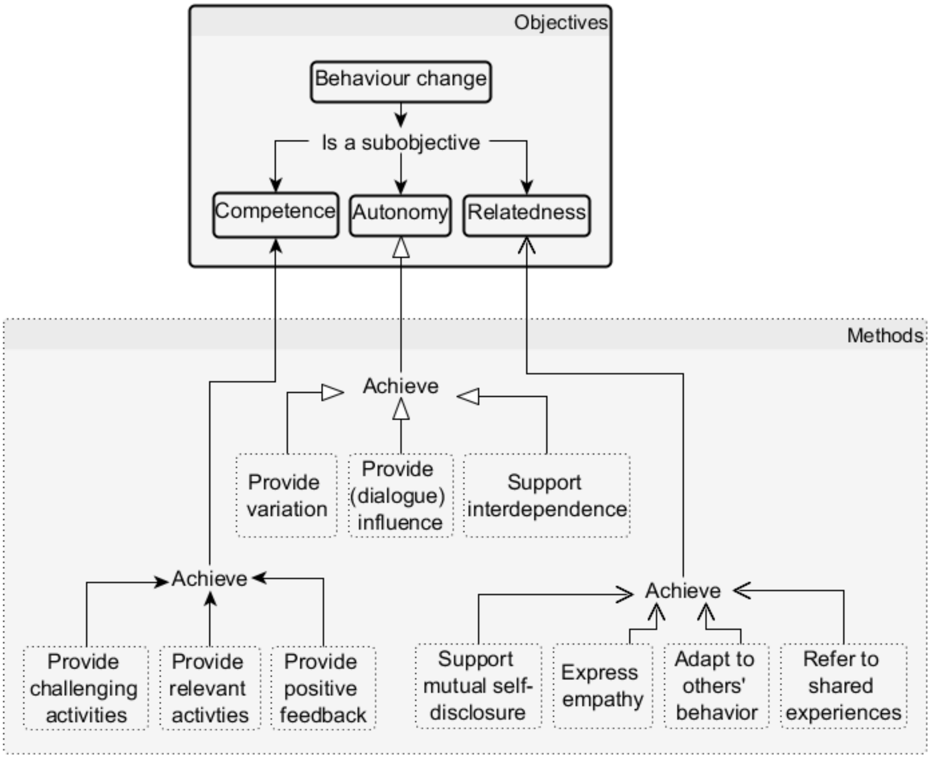



Figure 1.7: SDT and related methods.

techniques are also used in Motivational Interviewing as methods by the therapist (e.g. express empathy, provide positive feedback (see Figure 1.7)). It is important to identify these specific BCTs because it enables identification of effective methods, interactions between methods [61] and comparison between studies that use different interventions, but the same BCTs. In Michie et al. [160], a hierarchically structured taxonomy of behavior change techniques (BCTs) is construed with the help of 55 experts in delivering and/or designing behavior change interventions from different countries. This resulted in 93 BCTs that were clustered in 16 groups. An example of a group is “Reward and Threat” covering seven BCTs (e.g. material reward, threat, incentive).

A selection of BCTs can be implemented in a social robot where the social robot is used to complement a human. The robot can be viewed as a technological artifact of a behavior change support system (BCSS). BCSS is defined by Oinas-Kukkonen [179] as a socio-technical information system with psychological and behavioral outcomes designed to form, alter or reinforce attitudes, behaviors or an act of complying without using coercion or deception. A BCSS provides functions that are derived from theories of behavior change and persuasive technology.

The behavior change literature provides objectives and methods that can be used to specify these functions for a social robot. An example of such a robot is given in section 1 where Daniel is supported in his feeling of competence by the positive feedback of the robot/avatar and the selection of relevant activities to reach his objectives, which are also relevant for him.

It is important to guide the process of creating a BCSS by relating the derived functions (provide compliments) back to behavior change techniques (positive feedback) and always keep the intended outcome, to comply to the needs, in mind (increase competence). Currently, the derivation is hardly formalized and does not pinpoint the effects of a BCSS to specific functions. This is also explicitly indicated by Oinas-Kukkonen who sees this as one of the open questions on the BCSS research agenda [179].

### 1.1.3. SOCIAL ROBOTS

One of the first definitions of a social robot was provided by Bartneck and Forlizzi [9] (2004, page 2):

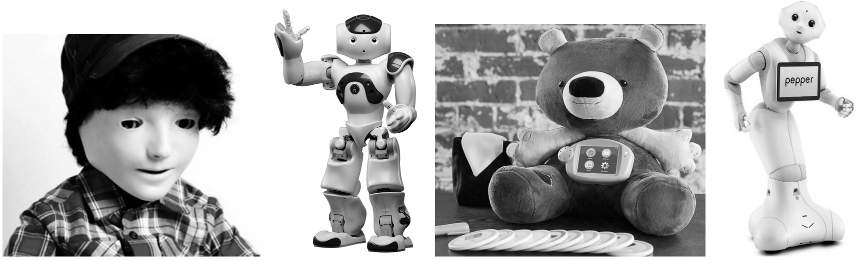
> "An autonomous or semi-autonomous robot that interacts and communicates with humans by following the behavioral norms expected by the people with whom the robot is intended to interact."

This definition ensures that robots act as humans expect, but these expectations can vary wildly between different stakeholders. In our case the robot has to act according the norms and expectations of children, but also of their formal and informal caregivers.

We will focus on social robots that aim at behavior change and can thus be seen as a BCSS. A social robot for changing the behavior of and/or educating children is not new as is shown by applications for autistic children [204, 6], for general child education [236, 239, 126], to acquire a healthy lifestyle [227] and even already for children with diabetes [46, 41]. Most children find interacting with a robot fun and in this way the diabetic children have something pleasant connected to their illness. Furthermore, aspects of behavior change and motivational theories can be implemented, dependent on the features and form of the robot, on the robot and applied to improve self-management. A not all-knowing robot that provides educational materials that are challenging but in the reach of the children, might support the children in getting relevant knowledge and skills to increase their self-efficacy.

As we want to provide something additional to the current care package, we further specify the robot to have a non-hierarchical relation with the child unlike an (in)formal caregiver. To be able to address functionalities as described by Motivational Interviewing, it is necessary that the robot has basic sensor and modalities to react appropriately on a child interacting with it. An overview of functionalities necessary for a social robot to be accepted as a communication partner are described by Fong, Nourbakhsh & Dautenhahn [91]:

1. Express and/or perceive emotions
2. Communicate with high-level dialog
3. Learn/recognize models of other agents
4. Use natural cues (gaze, gestures, etc.)
5. Exhibit distinctive personality and character
6. May learn/develop social competencies




(a) (b) (c) (d)

Figure 1.8: Kaspar (a), NAO (b), Jerry (c) and Pepper (d)

A robot that has (part of) these functionalities, has a non-hierarchical relation with the user, and is used to implement behavior change is defined by us as being a "PAL-robot". Below we discuss several platforms that can be used as PAL-robots and relate their functionalities to the characteristics defined in [91].

## PLATFORMS

**Kaspar** KASPAR (see Figure 1.8a) is a child-sized humanoid research robot designed to help teachers and parents support children with autism, developed by the University of Hertfordshire (UK). According to its website[5], it exhibits several of the characteristics identified in [91]. It can express simplified emotions, use natural cues like gaze, turn-taking behavior and react on touch (characteristics 1 (partially) and 4). There are several play scenarios in which the autistic children can practice social interactions. The robot has autonomous behaviors but is mostly used as a “hand puppet” of the researcher, clinician or children themselves[6] thereby being able to fulfill more of the characteristics. The intelligence is not in the robot but in its controller.

**NAO** The NAO (see Figure 1.8b) is a commercially available research platform and widely used in a range of social robot domains. August 2016, the NAO is in it’s 5th generation and 9000 have been sold world wide. The NAO has cameras, sonar, microphones and touch sensors that make it possible to get external input and speakers, led lights and 25 degrees of freedom in its body to provide output[7]. The robot comes with some predefined behaviors for emotions. Research institutes and some commercial companies (e.g. QBMT [8]) develop software that enables the NAO to reason about its inputs and respond with reasonable outputs. The different institutes are developing many programs with different and overlapping functionalities that are hardly ever compared with each other. The high level dialog is limited due to the current state of speech recognition and dialog management. So, although there is work on all 6 characteristics as defined in [91], it is

[5] http://www.herts.ac.uk/kaspar/introducing-kaspar
[6] http://www.herts.ac.uk/kaspar/research/technical-specifications
[7] https://www.ald.softbankrobotics.com/en/cool-robots/nao/find-out-more-about-nao
[8] http://zorarobotics.be/

hard to say how good it is at all these characteristics. This is due to the distributed development and because the NAO is most often deployed in a hybrid autonomous/human operated manner.

**Jerry the Bear** Jerry the Bear (see Figure 1.8c) is a commercially available robot for children with diabetes[9]. The ultimate goal of the company is to design Jerry to support open-ended play, and encourage kids to build a relationship with and care for Jerry in the same way that they care for themselves. According to the company's CEO Jerry the Bear is available in 25% of pediatric endocrinologists' offices in the U.S. and has been sold to 4% of newly diagnosed children (2016)[10]. Children need to take care of Jerry as he has diabetes, like themselves. They need to help Jerry accomplish his goals and in this way learn about their own diabetes. Jerry comes with 21 interactive storybooks and a selection of accessories (e.g. food cards) it can react to. The main characteristic it adheres to is having a distinctive personality (nr 5).

**Pepper** The Pepper robot (see Figure 1.8d) is from the same company as NAO[11] and like NAO, it is a commercially available platform. Unlike NAO it comes with some intelligence, like interpreting emotions and reacting on this (characteristic nr 1). This makes it more feasible for companies like Softbank and Nestle to buy it and try it.

Next to these robots there are also many start ups looking into the niche of social robots. Examples are Jibo[12], Buddy[13] and Personal Robot[14] (see Figures 1.9a, 1.9b and 1.9c). As none of these are thoroughly tested at this moment and it's not clear what functionalities they exactly have of their own or after being programmed we will not discuss them. One thing becomes clear from all these projects: Reaching a commercial standard for social robots is harder then expected as can be seen by the delayed delivery times and changes in capabilities (e.g. Jibo not being delivered outside the US and Canada due to technical and ethical problems).

We use the NAO as this was one of the few commercial and affordable platforms available when we began our research. We did some research with the iCat [127] (see Figure 1.9d), but the iCat was discontinued by Philips. The NAO provides a stable platform on which it is relative easy to implement it's behavior and connect to other services.

### SOCIAL ROBOT RESEARCH

Social robot research is booming and the movement of deep learning based on big data is supporting advancements in perceptual aspects like vision and recognition of speech[15],

[9] https://www.jerrythebear.com/
[10] http://www.mmm-online.com/technology/how-a-startup-is-educating-kids-with-diabetes%2Dwith-a-teddy-bear/article/502792/
[11] https://www.ald.softbankrobotics.com/en/cool-robots/pepper
[12] https://www.jibo.com/
[13] http://www.bluefrogrobotics.com/en/buddy-your-companion-robot/
[14] https://www.autonomous.ai/personal-robot
[15] http://www.inc.com/kevin-j-ryan/internet-trends-7-most-accurate-word-recognition-platforms.html

(a) (b) (c) (d)

Figure 1.9: Jibo (a), Buddy (b), Personal Robot (c) and iCat (d)

emotion [120], dialog and context [247]. The related actions are also improving, for instance moving through a room, answering questions (e.g. Amazon's Alexa, Apple's Siri, Google Now and Microsoft Cortana), and learning things like picking things up or know user preferences.

The advances in most of these fields are based on the availability of large datasets and a common interest among users. Also, with Alexa and consorts, users accept faults, but when such a system is used to answer emotionally loaded questions, faults will probably be less accepted. In our case, we look at a robot for children with diabetes. If we just focus on speech recognition, we already see that speech recognition rates for children are much lower than for adults [94]. The expectation is that this will improve because there is a lot of interest in educational applications in which speech recognition is necessary. A thing that will be harder to deal with is that (young) children often tell things out of context. For example a child in an experiment answered the question "How are you feeling" with a whole description of a museum visit from the week before. To react appropriately to these kinds of interactions is important, but on the other hand we can also put some extra effort in making users (children and adults) more aware of the limitations of the system.

Another aspect of using robots in the wild, as is the purpose of all robots mentioned in section 1.1.3, is that they need to keep the user motivated to keep using it. People expect a lot of variation and adaptation over time, but current social robots are lacking in reasoning skills to store and use data over a longer period of time and use semantic knowledge to communicate about it [25]. Most of the research performed with social robots is also missing the long-term interaction that makes it possible to get the data to reason over time and adapt to specific users [142].

To improve this we should do two things, 1) develop robots and their interaction for specific contexts, and 2) just start using the robots and see what are the aspects that work and don't work. Putting robots in society, perhaps without a lot of empirical research, will help to make the end users aware of the things a robot can and cannot do. This in turn will support acceptation and implementation of the robots in the long term [240].

Evaluating robots in the wild provides advantages in the sense that the complete system with all its interactions is used. The disadvantage is that it is hard to impossible to distinguish which design decisions lead to which results. Why do children have for instance a certain level of trust in the system, which functions or design decisions result in this outcome and how should these be adapted to improve the level of trust? In [262] a method is proposed to select hypotheses that are testable and have the highest empirical value. This ensures that the interconnections between functions remain manageable without losing too much of the empirical value. Although this method makes sure that the empirical evaluation is still feasible, the added knowledge and value of evaluating in the wild a complete system with all its interactions is lessened. These two different approaches are defined in [188] as design validation, that focuses mainly on user experience and model validation, that tests claims.

It will remain hard to compare research outcomes of different research projects. At the moment research is fragmentary and done with many different platforms, which will not change in the near future. Also, the fact that especially the interesting long term research is often performed with small and specific user groups, makes generalizability hard. To support comparability in this complex environment we should define the implemented functions on a level of design patterns and the instruments that measured the expected effects, so that we can compare functionalities and their effects over different robots and contexts.

### 1.1.4. Situated cognitive engineering

Oinas-Kukkonen [179] provides guidelines and core components as support for designing a Behavior Change Support System, but this is not enough for systematic design and evaluation of a complex system. Part of these guidelines are addressed, like taking knowledge of context and user into account, in the situated Cognitive Engineering (sCE) [172] method, while other guidelines are not addressed in sCE, for instance explicitly relate methods to specific software functions. The sCE method is developed to incrementally design and evaluate complex systems. Its main strength lies in the analysis of three system development components: the foundation, specification and evaluation (see Figure 1.10). It has been applied, for example, in the domain of behavior change [32] and robots [138]. In sCE functions are incrementally developed. It can be viewed as a refinement of classical cognitive engineering methods [108, 176, 197], addressing the reciprocal adaptive behaviors of both human and machine (i.e., emergent human-machine cooperation patterns).

The classical methods are mostly focused on a thorough domain and task analysis (e.g a scenario or vision), but the sCE method explicitly adds technology and human factor knowledge (methods, instruments) to establish a sound *foundation*. Technology is added for two reasons. First, it provides focus in the process of specification and generation of ideas, in our case a social robot. Second, the effects of technology are made explicit and are integrated into the development and thereby the evaluation process. The explicit use of human factors knowledge, e.g. knowledge on developmental age, behavior change, education and so forth, supports the development and embedding of functions and experimental results in theories. Moreover, the sCE method is situated in a domain that is made explicit in use cases that contextualize the (robot) functions. The explication from foundation (e.g., tasks analysis) to specification is guided by use cases.

The *specification* component encompasses, among other things, functions (requirements) that provide a high level description of the robot behavior, interaction design patterns, use cases and expected effects (claims). Key (recurring) functions are shaped in interaction design patterns (i.e., the "look-hear-and-feel" of robot behaviours) and applied in specific use cases (i.e., contexts). The functions are justified by the expected effects.

In the *evaluation* components, experiments test the expected effects (claims) and provide guidelines about what to use and when to use it. As such, the results of the evaluation also provides input for theory development.

Our research focuses on the development of a social robot with the *objective* to enhance child's self-management by applying different behaviour change *methods* as the theoretical foundation, and to establish the empirical foundation via sound evaluation *instruments* that show how far this *objective* has been achieved. We have to explicitly relate the sCE concepts to these objectives, methods and instruments in order to reason about the design decisions made. Part of this is already suggested by [180], which suggest to explicitly relate methods to specified software functions. The sCE method does insufficiently support this type of reasoning. For instance, it does not specify explicit relations between a specific method and the related objective and functions. Of course use cases take the objectives into account, but the relations are not well (or completely) modeled. Furthermore, the expected effects are related explicitly to the functions and instruments, but the interrelations between expected effects and functions are not made explicit. One function can have multiple effects, an effect can be related to different functions, multiple instruments can be used to measure the same effect, but it can also happen that one instrument measures multiple effects. These relations need to be explicated so that we can disambiguate the design and evaluation as much as possible by refining it, e.g. by using instruments that are related to specific effects as much as possible. Disambiguation will not always be feasible, but explicating all relations makes it possible to see where there are still ambiguous relations. Knowing these ambiguities can guide further design and evaluation.

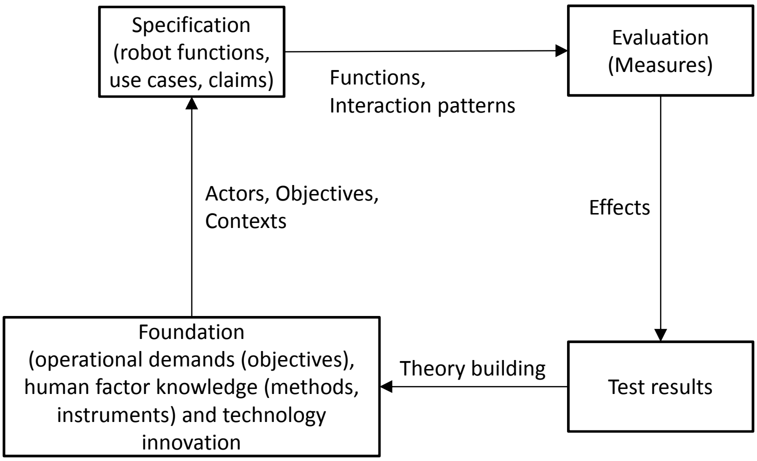



Figure 1.10: sCE.

1

## 1.2. PROBLEM STATEMENT

In the previous sections we introduced one of the main problems of current social robot research. The design and evaluation is not done in a systematic manner, connecting decision to theory and evaluation. This thesis provides an effort to fill this void. The main research question of this thesis is: Which robot functionalities and behaviors support the motivation and competencies for self-management of diabetes by children in the age of 7-12 with diabetes type 1 (T1DM)?

## 1.3. RESEARCH QUESTIONS, HYPOTHESES AND OUTLINE

In the previous section 1.2 the problem statement is described. To answer the main research question we had to define the core concepts of (1) children's diabetes self-management, and (2) the behavior support methods for motivation and competencies that can be integrated into robot's functionalities and behaviors. The relations between these concepts had to be defined also. The corresponding research question is:

> Research Question: Which knowledge structure can capture the core design and evaluation concepts of behavior change support robots for children's diabetes self-management?

The answer on this research question contributes to the refinement of the situated Cognitive Engineering methodology [172] in the form of the situated Design Rationale (sDR). The sDR provides a concise and coherent specification of the design rationale grounded in the context.

In parallel with this we performed experiments to fill, define and refine the concepts and its relations (Part I, Part II and Part III). Each part focused on a different aspect of the sDR.

The first part focuses on interaction design patterns and particularly on emotions. Describing interaction design patterns and their expected effects are of importance because not only the functionalities that you choose influence the system effects, but also the design of these functionalities. An example of this is that each mobile phone needs to have the functionality to answer a call, but for some mobiles this can be done by sliding, others have virtual or physical buttons for "accept" and "decline". These possible solutions are also called interaction design patterns. An interaction design pattern is a formal way to document a design solution for a common problem (e.g. answering the phone, showing emotions by a robot). We propose to formalize designs even if they have not proven themselves yet, and then we can refine or change the pattern based on found effects. These design choices influence the expected (use) effects. As we want to be able to pinpoint effects to choices made during the design it's important to make interaction design patterns part of the sDR.

### PART I: EMOTIONS IN DIFFERENT EMBODIMENTS

In Part I we focus on the interaction design patterns related to emotion expression by the robot (virtual and physical).

Emotions are a part of natural human-human interaction and we think it also has an added value for child-robot interaction[44]. To make this work there is a need for an emotion model that is perceived by the children as intended and also invokes desired behavior. To develop and evaluate such a model we need to take into account several aspects. Robotic platforms have different capabilities to exhibit emotions (e.g. face, sound, body) and it is also dependent on their presence in the real or virtual world. Next to this a decision needs to be made amongst the multiple emotion models, most noticeable distinction being discreet (e.g. Ekman [80]) or dimensional (e.g. Russell [210]). Leading to the following challenge:

I. Part I: Develop recognizable emotions for different embodiments that are perceived as intended and invoke desired behavior of the children.
   To address this challenge two design questions and three related hypotheses are outlined:

   I.1. chapter 3: Multi-modal emotions of a facial expressive robot
      - I.1.1. Design question: How to model the four Ekman emotions of anger, fear, happy and sad in the face and speech of the iCat, so that they are recognizable for children?
      - I.1.2. Hypothesis: Children will show better understanding, acceptance, trust, fun, empathy and performance when interacting with an iCat that expresses multi-modal emotions (i.e., increasing in the following order: no emotions, facial, facial-and-vocal).

   I.2. chapter 4: Bodily expressive robot versus facial expressive robot
      - I.2.1. Design Question: How to model the five Ekman emotions of anger, fear, happy, sad and surprise in the postures and LEDs of the NAO, so that they are recognizable for children?
      - I.2.2. Hypothesis: Three factors influence the recognition rate of robot's emotions: (1) the recognition rate differs between robot embodiments, (2) the rate is higher when the emotions are expressed in a congruent context (compared to no context), and (3) the recognition improves over time.

   I.3. chapter 5: Physical versus virtual embodiment of a robot
      - I.3.1. Hypothesis: Children's performance, attention, trust, enjoyment and preference in quiz task are higher, when interacting with a physical NAO compared to a virtual NAO.

Models for emotion expression by the iCat in face and vocal and in posture for the NAO were developed and evaluated on recognizability with positive results. The emotion expression also stimulated favorable behavior by the children, contributing to the objective of relatedness. An interesting result is that vocal emotion expression looses in understandability, which negatively effects the trust children feel toward the robot chapter 3, underscoring the necessity of formalizing the design decisions and their relations to functions and effects.

#### PART II: ROBOTS FOR COMPETENCE AND RELATEDNESS

Where the first part focused on emotional expression to support relatedness, the second focuses on adaptivity and/or expressivity of a robot that supports multiple factors of the Self Determination Theory. As with emotions, a model for adaptivity needs to be developed before it can be evaluated. The same holds with moments for self-disclosure. Leading to the following challenge.

II. Part II: Develop a set of behaviors for a robot that invoke feelings of competence and relatedness of the children interacting with the robot.
To tackle this challenge three design questions and related hypotheses are discussed in chapter 6, chapter 7 and chapter 8 each focusing on one specific use case (math game, quiz and diary).

II.1. chapter 6: Increasing motivation by adapting difficulty

II.1.1. Design Question: How to challenge children, aged 9-10, within their dynamic individual capabilities (c.f. Zone of Proximal Development [256] and Optimal Challenge [62]) in a math and memory game with a robot?

II.1.2. Hypothesis: Child's motivation to play a math game with a robot is higher when the game is adapted to his or her dynamic individual capabilities.

II.2. chapter 7: Reciprocal emotion elicitation

II.2.1. Design question: How to model robot's emotional expressions that represent: robot's current performance, match child's intro-extroversion trait, and adapt to child's performance and emotional state?

II.2.2. Hypothesis: A robot with adaptive emotional expressions will "score higher" on relatedness factors in both behaviors (emotional expressivity of the child) and opinion (fun, acceptance, empathy, trust, preference and recognized emotional expressivity) in comparison to a robot without adaptive emotional expressions.

II.3. chapter 8: Stimulating mutual self-disclosure

II.3.1. Design question: How to design, within the context of a diabetes diary, self-disclosure and empathetic behavior by a robot based on mutual self-disclosure (e.g. [202]) and empathy theories [67]?

II.3.2. Hypothesis: Empathetic behaviors and self-disclosure of a video-conferencing robot improve children's adherence to fill out their diabetes diary.

These questions and hypotheses resulted in models and behaviors for: adaptivity based on theories of challengingness, reciprocal emotion expression and reciprocal self-disclosure. The intended behaviors were evaluated and positive results were seen. Furthermore, a specific measure to evaluate intrinsic motivation (free choice period) was tested positively.

### PART III: ROBOTS FOR AUTONOMY

In Part I the focus was on interaction patterns, in Part II the robot was evaluated in separate use cases. In Part III a combination of use cases is made and the robot, including its activities, is evaluated as an integrated system (over multiple sessions). The main foci here are 1) to contribute to a feeling of autonomy, while keeping the functions that attribute to relatedness and competence, and 2) to evaluate an integrated system (in the wild). Leading to the following challenge.

III. Part III: Develop a set of behaviors and activities that support a feeling of autonomy with the children and evaluate them in an integrated manner over time. Two design questions, a hypothesis and two research questions were derived from this challenge.

- III.1. chapter 9: Behaviors for the iCat to display different roles
  - III.1.1. Design question: How to create behaviors for a moderate expressive [253] iCat robot based on Motivational Interviewing [207] techniques?
  - III.1.2. Hypothesis: Text, virtual and physical robot are for children, in an incremental order, increasingly motivating and educating. This can be explained by the incremental number of motivational interviewing techniques that can be implemented in the different interfaces.
- III.2. chapter 10: Evaluating in the wild
  - III.2.1. Design question: What does experimentation in the wild add over controlled experiments that test isolated components of the robot one-by-one in a lab environment?
  - III.2.2. Research question: Is the complete system is appreciated by children with diabetes, after multiple interactions, on the factors; autonomy, competence and relatedness.
  - III.2.3. Research question: Does performing an experiment in the hospital with the real target users increases acceptation of all involved (children, parents and health care professionals)?

Evaluating an integrated system complicates the analysis afterwards, due to the correlated function and thereby effects. This is further complicated when evaluating in the wild, where there are many confounding variables. Nevertheless, it provides valuable contributions on how children (and their surroundings) respond to a robot that behaves according to its role and elicits the intended behavior.

Figure 1.11 shows the setup of this thesis. The current section will be followed by chapter 2 which provides a detailed description of the sDR methodology, which is subsequently supported by the three challenges that are addressed in three separate Parts. Each Part contains several chapters addressing the design questions and hypotheses as described in this section. This is concluded in chapter 11 discussing the contributions and limitations.

In the introductions of the parts we present the different studies that support the challenge of the part and their respective individual design questions and hypotheses ending with a short conclusion on how these studies fit in the larger sDR puzzle. This finally results in an overall sDR (see https://bit.ly/2RXxWNd) of the ALIZ-e project that is used as an example project throughout the thesis.

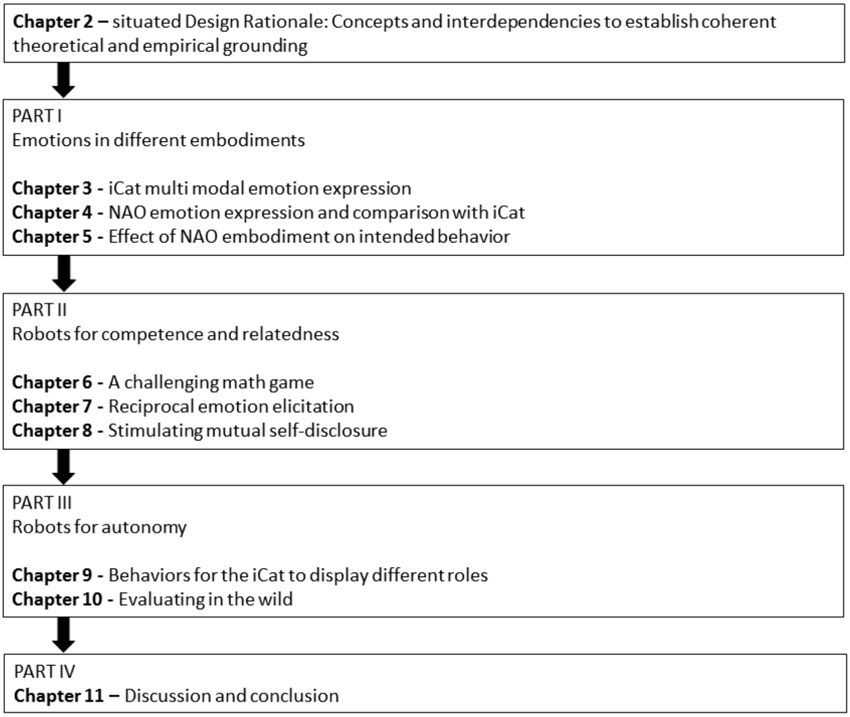



Figure 1.11: Thesis Overview

# 2 | Specifying and testing the design rationale of social robots for behavior change in children[1]

**Abstract**

*We are developing a social robot that helps children with diabetes Type 1 to acquire self-management skills and routines. There is a diversity of Behavior Change Techniques (BCTs) and guidelines that seem to be useful for the development of such support, but it is not yet clear how to work out the techniques into concrete robot support functions and behaviors. The situated Cognitive Engineering (sCE) methodology provides guidance for the design and evaluation of such functions and behaviors, but doesn't provide a univocal specification method of the theoretical and empirical justification. This paper presents an extension of sCE: a formal template that describes the relations between support objectives, behavior change theory, design specifications and evaluation outcomes, called situated Design Rationale (sDR) and the method to get this. As test case, the European ALIZ-e project is used to instantiate this design rationale and to evaluate the usage. This case study showed that sDR provides concrete guidance (1) to derive robot functions and behaviors from the theory and (2) to designate the corresponding effects with evaluation instruments. Furthermore, it helps to establish an effective, incremental and iterative, design and evaluation process, by relating positive and negative evaluation outcomes to robot behaviors at the task and communication level. The proposed solution for explicating the design rationale makes it possible for others to understand the decisions made and thereby supports replicating experiments or reusing parts of the design rationale.*

## 2.1. Introduction

There is a need for social robot design methods, which provide theoretically and empirically founded implementations that can be systematically reused, compared and built upon progressively (cf., [91]). Current design methods do not (yet) meet these needs, holdings back the coming of age of the research field.

This paper focuses on the development of robots for behavior change. Although there is a substantial amount of research in social robots and behavior change techniques, it is hard to compare the results of studies due to a lack of agreement on (1) the (definitions of) relevant theoretical concepts, (2) the design specifications, (3) the

[1]This is the text from the author copy of the published paper [148] http://dx.doi.org/10.1016/j.cogsys.2016.07.002

methods for validation (or evaluation), and (4) the approach to relate these concepts, specifications and methods. Literature from the social robot domain on classification of robots (e.g. [65]) and evaluation (e.g. [260]) provides valuable information for design specifications and their evaluation. However, it is unclear how they relate and can be linked to behavior change theories. On the other hand, for behavior change techniques there is a taxonomy in development [160] which supports disambiguation of results, and therefore validation of effective techniques, but it does not relate these to design specifications (such as use contexts). Use contexts are taken into account in the research of Behavior Change Support Systems (BCSS), for instance in the persuasive systems design (PSD) model [181]. This model emphasizes the translation between method and design patterns for functionalities related to the context. Although method, requirement, design and implementation are related in PSD, it does not model the correlations and interrelations between different implementations.

An open question remains: "How can we conduct experiments in such a manner that it will be really possible to pinpoint a change to have been caused by a BCSS, or even more precisely, by a specific software feature in it?" [179] . Our social robot is in essence a BCSS and the question we want to answer is quite similar:

- How can we design and evaluate in such a manner that a) robot behaviors are derived from theory and b) evaluation effects can be designated to specific robot behaviors?

The situated Cognitive Engineering (sCE) methodology [171] can partially answer this question. sCE has been used in different domains, amongst which to systematically design and evaluate robot systems [138]. Although sCE supports iterative and incremental design and evaluation, it does not provide precise and concise translations and relations between the theory, functionalities of the system, hypotheses and instruments to evaluate (i.e. the concepts).

The situated Design Rationale (sDR) was developed as a refinement of the sCE methodology. This formal template supports the design of functionalities, the planning and performance of evaluations, and makes it possible to reason about the evaluation effects and decisions afterwards. To come to this formal template, we distinguish three subquestions all in the context of the development of a social robot for supporting behavior change:

1. Which minimal set of concepts is needed to describe the what, when and why of design decisions?
2. How do these concepts relate to each other?
3. What is an adequate, concise and coherent, representation for describing the concepts and its relations for the design and evaluation process?

The research took place in the context of the development of a social robot that provides self-management support for children with diabetes (i.e., the European ALIZ-e project[2]). The structure of this paper is as follows: First in section 2.2, we provide background on diabetes, social robotics, behavior change and situated Cognitive Engineering. Second in section 2.3, we describe the sDR template, that describes the concepts

[2] www.aliz-e.org

and it relations, followed by the instantiation of sDR in section 2.4. In section 2.5 the use of the sDR is further exemplified with an experiment performed within the ALIZ-e project. And we finish with the conclusions and discussions on future work in section 2.6.



## 2.2. BACKGROUND

Type 1 diabetes has an enormous impact on the daily life of children with this illness as we will discuss in section 2.2.1. There is a need for support of self-management and behavior change. A social robot might provide this support for this user group (age 7-12) (section 2.2.2). The behavior of the robot should be based on knowledge from behavior change theories and systems (section 2.2.3), and the design of the robot should be based on a state-of-the-art design methodology (section 2.2.4). Based on this background we can conclude what is lacking to come to a concise and precise situated Design Rationale.

### 2.2.1. TYPE 1 DIABETES MELLITUS

To understand why we want to develop a social robot to support children with diabetes to increase their self-management it is necessary to understand what diabetes is and what this means for the life of the children, and their environment. There are two types of diabetes, Type 1 and Type 2 [109]. Type 1 typically presents itself at a young age, while Type 2 often occurs at a later age. Where Type 1 Diabetes Mellitus (T1DM) is a result from destruction of the insulin-producing cells in the pancreas by the autoimmune system, Type 2 is a metabolic disorder where the body does not make and absorb enough insulin. We will further focus on T1DM, because that is the type that is most prevalent in children and the incidence is rising [187]. For these children it is very important to keep their blood glucose levels as steady as possible. To reach this objective, children and their social environment (parents, teachers, siblings, friends etc.) need to have knowledge and skills to manage the disease. Examples of these are: Regularly measuring of blood glucose, counting of carbohydrates, calculating needed insulin and injecting (when pen is used) or bolusing (when pump is used) accordingly, and discounting the (interactive) effects of food intake, physical exercise, mental stress and hormones. Furthermore, a child and his or her environment need to be able to recognize symptoms of high and low blood glucose to act accordingly. Even when managed properly, a child will have periods of high imbalance due to for instance hormones or growth spurts. The effects of T1DM, even with our modern treatment, are quite severe. More than 50% of the children develop complications with regard to major organs like the heart and blood vessels 12 years after diagnosis [63]. The life expectancy of children diagnosed by age 10 is 19 years shorter than that of healthy children [169]. There are also effects on psychological well-being, feelings of embarrassment and on school performance [190]. The effects on psychological well-being are not limited to the children themselves, but also their parents are hugely influenced, because they understand the long-term effects better than a (young) child [33]. Other research suggests that high family stress negatively affects glycemic control [246]. To lower family stress it is important that children learn to manage their illness at a young age and that parents let them do this. A social robot can support in this, because it has a non-hierarchical relation with the child unlike a (in)formal caregiver. A social robot for changing the behavior of and/or educating children is not new as is shown by [204, 6] where they are applied for autistic children, [236, 239, 126] for education and [227] to acquire a healthy lifestyle. Aspects of behavior change and motivational theories can be, dependent on the features and form of the robot, implemented on the robot and applied to improve self-management.

### 2.2.2. Social robots

Below we provide a short overview of design and evaluation methods that are used in the field of personal social robotics on context, behaviors, appearances, and effects. We exclude work-oriented human-robot interaction (e.g., human-robot teamwork; [269, 183, 234]), because we focus on (non-work) social settings of the child. Robots can be classified according to their appearance (from mechanical to human-like for instance [266]) and their behavior. Bartneck et al. [9] for instance classify social robots on five factors: Form (abstract - anthropomorphic), modality (unimodal - multimodal), social norms (no knowledge on social norms - full knowledge on social norms), autonomy (no autonomy - full autonomy) and interactivity (no causal behavior - fully causal behavior). [91] provide a more elaborate classification specifically for socially interactive robots, robots for which social interaction plays a key role. First they identify two primary approaches to build socially interactive robots, biologically inspired or functionally designed. Decisions on the design and evaluation need to take the context into account. Fong et al. further identify other aspects that can be used to classify robots, e.g. embodiment, emotion, dialogue, personality, perception of humans, user modeling, socially situated learning and intentionality. It is meant as support for people designing socially interactive robots to make decisions on the form and behavior of the robot for the use in a specific context. This is further explained by providing different applications and examples of robots used in every application and a short indication of what aspects of the classification they adhere to. [64] looks at the aspect of consistency of design and behavior. Examples are provided what happens when it is not consistent (e.g. very humanlike appearance of robots induces the uncanny valley effect, because it cannot perform as expected), but reaching consistency seems to be a matter of trial and error. With the design space provided it is possible however to place robots on the two dimensional axis of appearance (machine like vs. human like) and behavior (non-social and non-interactive vs. social and interactive). [231] uses the axis of machine to human like, next to an indication of toy like, body and facial realism to categorize and evaluate 3 robots (iCat, NAO and Nabaztag) and a human-like avatar. These different ways of classifying (social) robots shows that designers of robot systems make many choices, and these choices should be formalized to understand why these choices were made and also decide on the validity of the choices after evaluation.

It is important for comparability between different robot designs to measure the same type of effects and preferably also use the same measures. [260] propose to use the following evaluation factors: Usability, social acceptance, user experience and societal impact. Which factor to use depends on the hypotheses. Furthermore, they propose, for the evaluation of hypotheses, to use a mix of interdisciplinary evaluation methods: Expert evaluation, user studies, (standardized) questionnaires (e.g. unified theory of acceptance and use of technology (UTAUT) questionnaire [252]), physiological measures, focus groups and interviews. [10] provides an instrument toolkit to measure how users perceive a robot on five factors relevant for HRI: Anthropomorphism, animacy, likeability, perceived intelligence and perceived safety. They developed five validated questionnaires for these five factors. These questionnaires are all relevant for evaluating the design of a social robot, but do not provide measures that are related to the objective of the robot use, e.g. education.



### 2.2.3. BEHAVIOR CHANGE

Behavior change is a large research field. We will focus on two topics: A taxonomy developed to describe behavior change methods and a model to design persuasive systems for behavior change. The taxonomy is interesting, because it is an effort to describe components of a behavior change method in a way to derive effectiveness in a similar way we want to describe the components of the robot. The persuasive systems model is of added value, because we also want to create a persuasive system, where we use the robot as ICT component.

Interventions to change behavior are complex and have many interacting components [61]. Therefore, the same problems occur as in social robot research: Research outcomes are hard to replicate, to implement in practical applications and to use for building theory [160]. We therefore need a better understanding of which components are effective within a behavior change intervention.

A first step is to get a common understanding of the components in an intervention. This helps in recognizing overlap between different interventions and identifying effective components. In Michie et al. [160] a hierarchically structured taxonomy of behavior change techniques (BCTs) is construed with the help of 55 experts in delivering and/or designing behavior change interventions from different countries. This resulted in 93 BCTs that were clustered in 16 groups of which 26 were used 5 or more times in different interventions. An example of a group is “Reward and Threat” covering seven BCTs (e.g., material reward, threat, incentive).

A selection of BCTs can be implemented in a social robot where the social robot is used instead of, or as a complement of, a human. The robot can be viewed as the IT artifact of a behavior change support system (BCSS). BCSS is defined by [179] as a socio-technical information system with psychological and behavioral outcomes designed to form, alter or reinforce attitudes, behaviors or an act of complying without using coercion or deception. A BCSS is a complex system that is developed using theories of behavior change and persuasive technology by explicating functionalities of a system.

To support the design of a BCSS, Oinas-Kukkonen suggests the use of the Persuasive Systems Design (PSD) process. The design of a BCSS takes postulates from User Centered Design which are also used in persuasive design (e.g., ease of use), uses these in context (intent, event and strategy) and then a decision on the design of software features needs to be made. During the context step the intended outcome is decided on, using the outcome & change design matrix, which also influences the strategies. The combination of the PSD process and the outcome & change matrix provides a way of defining the system, context and intent clearly. This is necessary because these influence the outcomes, e.g., different IT systems will be able to implement persuasive strategies on different levels.

The behavior change literature provides objectives and methods that can be used to guide implementation of a social robot for behavior change. The PSD model guides the design of a BCSS by relating functions to behavior change techniques and always keeping the intended outcome in mind. The design thus takes as a starting point the intended outcome, but due to a lack of formalization between design decisions and evaluation measures the PSD model cannot pinpoint the effects to specific functions. This is also explicitly indicated by Oinas-Kukkonen who sees this as one of the open questions on the BCSS research agenda [179].

### 2.2.4. SITUATED COGNITIVE ENGINEERING

The situated Cognitive Engineering (sCE) [172] methodology has its main strengths in the analysis of three system development components: the foundation, specification and evaluation. It has been applied, for example, in the domain of behavior change [32] and robots [138]. In sCE functions are incrementally developed. It can be viewed as a refinement of classical cognitive engineering methods [108, 176, 197], addressing the reciprocal adaptive behaviors of both human and machine (i.e., emergent human-machine cooperation patterns).

The classical methods are mostly focused on a thorough domain and task analysis, but the sCE method adds explicitly technology and human factor knowledge (methods, instruments) to establish a sound *foundation*. Technology is added for two reasons. First it provides focus in the process of specification and generation of ideas. Second, the effects of technology are made explicit and are integrated into the development and thereby evaluation process. The explicit use of human factors knowledge, e.g. knowledge on developmental age, behavior change, education and so forth, supports the development and the embedding of functions and experimental results in theories. Moreover, the sCE method is situated in a domain that is made explicit in use cases that contextualize the (robot) functions. The explication from foundation (e.g., tasks analysis, value sensitive design) to specification is guided by use cases.

The *specification* component encompasses, among other things, functions (requirements), interaction design patterns, use cases and expected effects (claims). Key (recurring) functions are shaped in interaction design patterns (i.e., the "look-hear-and-feel" of robot behaviours) and applied in specific use cases (i.e., contexts). The functions are justified by the expected effects.

In the *evaluation* components, experiments test the expected effects (claims) and provide guidelines about what to use and when to use it. As such, the results of the evaluation also provides input for theory development.

Our research aims the development of a social robot with the *objective* to enhance child's self-management by applying different behaviour change *methods* as the theoretical foundation, and to establish the empirical foundation via sound evaluation *instruments* that show how far this *objective* has been achieved. We have to explicitly relate the sCE concepts to these objectives, methods and instruments (see Figure 2.1) in order to reason about the design decisions made. The sCE method does insufficiently support this type of reasoning. There are for instance no explicit relations between a specific method and therefore objective and a function. Of course use cases take the objectives into account, but the relation is not made explicit. Furthermore, the expected effects are related explicitly to the functions and the instruments, but the interrelations between expected effects and functions are not made explicit. One function can have multiple effects, an effect can be related to different functions, multiple instruments can be used to measure the same effect, but it can also happen that one instrument measures multiple effects. These relations need to be explicated so that we can disambiguate the design and evaluation as much as possible by refining it, e.g. more specificity in instruments. Disambiguation will not always be possible, but explicating all relations makes it possible to see where there are still ambiguous relations. Knowing these ambiguities can guide further design and evaluation.



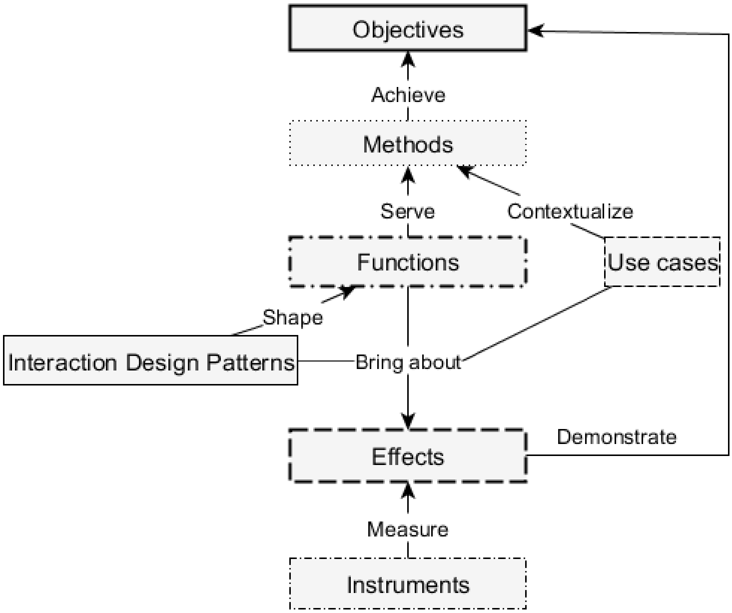



Figure 2.1: Generic concept map of the situated Design Rationale (sDR).

## 2.3. SITUATED DESIGN RATIONALE

To create a situated Design Rationale (sDR) that specifies the relations between functional aspects and expected effects in a manner such that we can reason about the design decisions made and the interactions between effects and functions, we extend the sCE method. The concepts come from the sCE method and some of the relations also, but we add relations to make all relevant relations explicit in an sDR.

### 2.3.1. CONCEPTS

In the previous section we distinguished the relevant concepts that have to be related to each other. The first relevant concept is *objectives* (e.g. support the forming of a relation between robot and user), second are *methods* (e.g. adapt the robot to the user's behavior) that are derived from literature or experiments to reach these objectives. The methods then have to be translated into, the third concept; *functions* (e.g. adapt the robot system to the state of the child) of the robot. The functions are shaped by, fourth concept, *interaction design patterns* (e.g. use of prosody to express emotions by the robot). Fifth, *use cases* (e.g. a quiz between the robot and child in which they act as peers) are used to contextualize the methods and show which *effects* (sixth concept) we expect towards the objectives (e.g. children relate more to the emotional expressive robot). But also the effects in relation to the implemented functions and design patterns are described (e.g. an expressive robot supports emotional contagion - i.e. the child is more expressive, emotions are recognized). Seventh and last *instruments* are then used to measure these effects (e.g. arousal and valence observations by the child). In Figure 2.1 the seven generic concepts and their relations are shown. In the following paragraph we explain how the generic sDR is developed.

### 2.3.2. SITUATED DESIGN RATIONALE TEMPLATE

The situated Design Rationale is developed to support design of functionalities and evaluation before an experiment and reason about the effects and decisions afterwards. The explication from theory (objectives and methods) to functions and then to effects should thus be made clear. To make this possible we have to relate the concepts to each other, and as is said “a picture is worth a thousand words” we decided to use concept mapping [178] as a tool to describe the relations. In a concept map, relations between ideas, images, or words are linked with meaningful arrows. In our case the meaningful words are the concepts and the meaningful arrows the relations between the concepts[3].

The *objectives* come from the foundation of sCE, relevant theories (behavior change, education) are taken into account as well as knowledge on human factors (what are the capabilities of a child in the age group 7-10) and technology (what are the robot (in)capabilities) to come to a selection of *objectives*. Also based on the foundation of sCE *methods* are selected to **achieve** the *objectives* and which are supported by literature or derived from empirical experiments (e.g. provide variation, which supports competence and comes from educational theory). *Use cases* are then described to **contextualize** the *methods* and to show which *effects* are **brought about**. *Functions* are related to the *methods*. Only *functions* that **serve** a method are relevant here. In some cases, explicating the relation between *method* and *function* is quite straight forward. An example of this is a *method* that prescribes variation and a *function*“Provide multiple activities”. *Functions* are **shaped** into *interaction design patterns*. An example of this is the *interaction design pattern* “Recognizable emotion expression” that supports the higher level *functions* “Exhibit social behavior” and “Adapt robot to child state – within boundaries”. The *interaction design pattern* **shapes** the function and defines what is needed to reach, in this case, “Recognizable emotion expression”.

Then we specify the *effects* that the *interaction design patterns* and the *functions* **bring about**. This is a very important step. If a *function* cannot be related to an *effect* it should **bring about**, that *function* or *interaction design pattern* has to be reconsidered. The reason for this is that the relation between *functions, patterns* and *effects* is also the

[3] Using yEd `https://www.yworks.com/en/products/yfiles/yed/` we created a general concept map of sDR 2.1 in which the concepts and their relations are visualized

relation back towards the *objectives*. The *effects* **demonstrate** the result on *objectives*. An equally important relation is that from *effects* to *instruments* that **measure** the *effects*. When there is no *instrument* to **measure** an *effect*, the *effect* might be too specific or generic. The design is also guided by this step, because when there is one *instrument* that is used to **measure** many *effects* the results cannot be used to disambiguate between different *functions*. Therefore, either the *effects* have to be made more specific, or the *functions* need to be made more distinguishable from each other so that there is less ambiguity between the *effects*.

When there is a first complete version of the sDR, it has to be checked and decided on what will be the focus of an experiment. The sDR can support deciding where experiments are needed to get more information, but also review the *instruments* to see if they are specific enough to derive conclusions from the results. The results can then be used to reason about the decisions made and refine and extend the sDR. Figure 2.1 provides a generic sDR, which we will instantiate using an experiment performed within the ALIZ-e project in the next section.

It's interesting to see the similarities between Worth Mapping [53] and sDR. Both take into account the values of the end users; in Worth Mapping these are the objectives of the design while in sDR these are part of the methods to reach the objectives and used to enrich the use cases. To satisfy the values both identify needed elements or methods and functions to reach a worthwhile outcome. This means that Worth Mapping guides the interaction design by making relations between values, elements and attributes clear, while sDR makes the transition to context and effects. sDR uses the values and attributes to describe the use cases and contextualize the methods which in its turn constrains the functions and interaction design patterns. The measured effects then demonstrate the progress towards the objectives, but also if user values are met.

## 2.4. Instantiation of a sDR

We will now show how sDR can be used to describe the design and evaluation activities of the ALIZ-e project by instantiating the concepts with specific examples. We do this by going through the concepts, explaining decisions and showing parts of the sDR to exemplify the concepts. The complete sDR of the ALIZ-e project can be found here: `https://goo.gl/0HgUC8`.

In the complete sDR there are many intersecting lines, in a limited way this is also the case in the figures presented in this paper. As this problem can not be eliminated we used different arrows to make clear what the origin of lines are. In Figure 2.6 we added the outgoing arrow form to the text of the functions.

### 2.4.1. Objective

The overall objective of ALIZ-e is behavior change for self-management, with a focus on children with diabetes. The objective is thus behavior change and a decision needs to be made on which theory we will use to relate our progress to.

#### Choice for behavior change objective

Many theories for behavior exist, and the choice of one over the other guides the priority of objectives. We will briefly discuss Theory of Reasoned Action II [88], the Extended

Parallel Process Model [265] and the Self-Determination theory [74].

In the Theory of Reasoned Action II (TRA II) behavior is determined by intention, which is determined by attitude, perceived norm and perceived behavioral control (similar to self-efficacy). Actual control is determined by environmental factors and skills to deal with these.

The Extended Parallel Process Model (EPPM) argues that changing behavior, attitude and intention results from an attempt to control threat, while not changing behavior comes from fear. According to EPPM people deal with threats and fear in three different ways. First, a threat can be seen as insignificant so there is no motivation to change. Second, a threat can be perceived as so serious they feel not able to deal with it, because they don't have enough perceived self-efficacy and response efficacy. The third option is that the threat is perceived as serious and they feel empowered to do something about it because of high self-efficacy and response efficacy.

The Self-Determination Theory is a motivational theory that supports a continuum of motivation, from external regulation (completely extrinsic) towards more and more internally motivated to end in intrinsic motivation [213]. The motivation can be influenced by supporting three basic psychological needs: autonomy, competence and relatedness. Autonomy is about the willingness to do a task, competence is the need for challenge and feeling of effectance, and relatedness refers to the connection with others [73]. Long-term interaction is seen as a prerequisite for behavior change in the long run and several behavior change methods endorse the reasoning that for long-term interaction there is a need for a bond with the interaction partner (e.g. Motivational Interviewing [161]).

All three example theories show differences, but also similarities (e.g. self-efficacy is important in all three). Because of these similarities and the complexity of these theories, there is an ongoing effort to analyze behavior theories until the level of behavior change techniques and then evaluate those on effect [160]. As a decision had to be made we chose Self-Determination Theory as our starting point (see objectives in Figure 2.2), because this theory is used not only in behavioral change but also in education [175], for children in the relevant age group (7-11) [223] and in games where it showed to be a predictor of enjoyment and future game play [215].

### 2.4.2. METHODS

Another advantage of SDT is that there is an ongoing effort to connect the methodology of Motivational Interviewing (MI) to the theory of SDT [153]. Motivational Interviewing (MI) is a proven effective counseling style for promoting behavior change, but it is not grounded within a theoretical framework, SDT can provide this framework. MI techniques have also been used in persuasive technologies as the Health Buddy [27] and techniques from MI have been implemented in a social robot for adults with diabetes [147].

To reach the objectives we can thus draw upon methods of MI, we further draw upon (amongst others) educational, gaming and persuasive methods and methods used for rapport building in human-human interaction. These methods are overlapping; for instance Vygotsky's educational theory [256] and gaming theory [62] both endorse the importance of having challenging, relevant activities to support intrinsic (long-term) moti-

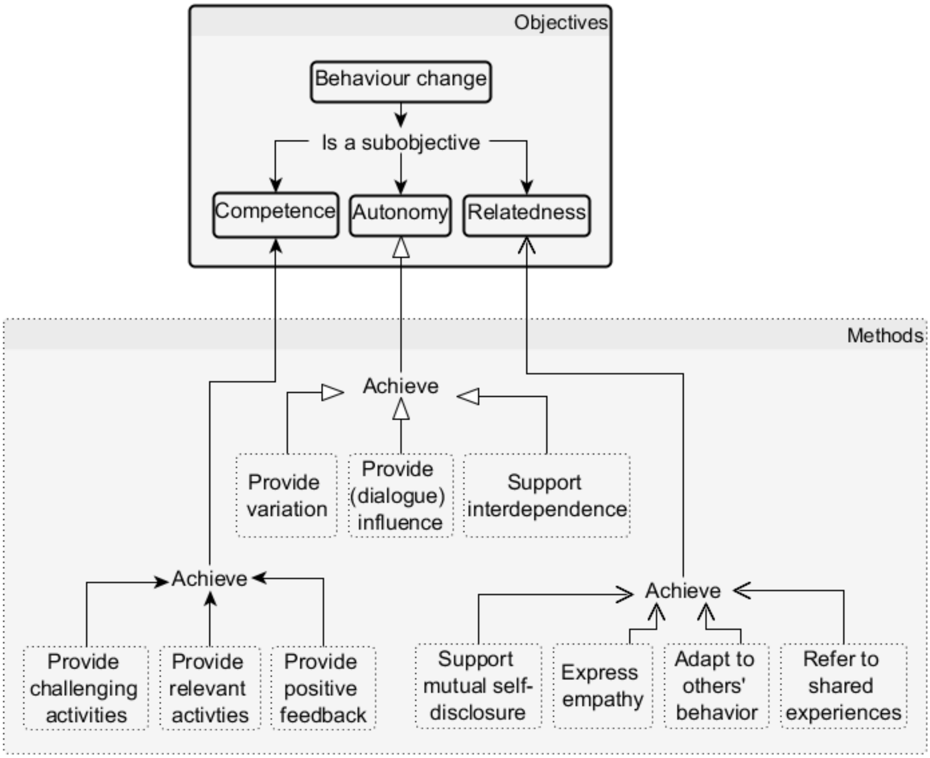



Figure 2.2: Objectives and methods that achieve them

vation. Literature supports the relation between challenging activities and self-determination theory [101]. Vygotsky and MI also state that the teacher/therapist should build up rapport with the student or client; in MI this is further elaborated in methods to build this up (e.g. express empathy). In Vygotsky the teacher can also be a peer in a collaborative learning sessions; the peers learn from each other and need each other's help. In such a setting the rapport building will have another dimension than with a teacher/therapist, e.g. the shared experiences and matching the personal norm will be differently implemented. In [270] an overview of methods to reach rapport is provided.

Figure 2.2 shows the methods used within the ALIZ-e project and their relations to the objectives. All methods come from literature; MI [153], educational [256, 4], gaming [62] and relation theories [270]. In some of this literature the methods are explicitly linked to SDT objectives (e.g. [153, 101]), other relations need to be derived.

As can be seen, there are three different objectives. These objectives are not completely unrelated, but all have their main focus which is depicted in the figure. The functions will connect the different methods to each other.

| ID | Use case | Description |
|---|---|---|
| 1 | Competitive quiz with robot peer | The robot and child play a competitive Trivial Pursuit based quiz where they alternate in answering questions. |
| 2 | Collaborative sorting game with robot peer | The robot and child play a collaborative game on a large touch screen on which they have to swipe images, that are on the screen, to the correct categories (most of the time 2, that are on the left and right side of the screen). |
| 3 | Imitation memory game with robot peer | The robot makes a movement (e.g. arms up) and then the child imitates this and adds another movement, which the robot has to imitate. The string of movements gets longer and longer, so its both a movement and memory game. Variations are: that the robot can only add movements, some movements are prohibited, and there are different levels of sequences (more complex) and movements (more difficult). |
| 4 | Watching educational video with robot peer | Robot and child watch a video together. |
| 5 | Providing a combination of activities | Provide multiple activities as described in Use Case 1-4 |
| 6 | Engaging in small talk with robot peer | Some interaction about hobbies, activities, friends, diabetes. |
| 7 | Support robot from one activity to another | The child has to help the robot from one activity to another, by walking with it (holding hands) or carry the robot. |
| 8 | Helping robot to stand up | When the robot falls over the child helps it in getting up. |

Table 2.1: Overview of the ALIZ-e use cases.

### 2.4.3. USE CASES

The objectives, methods and (later on) related effects and measures won't change a lot during the course of a project. A method can be added, but as the objectives are the starting point these will be relatively stable. The choice for a method also guides the expected effects and with these the measures. This is different for the other concepts we discuss, the use cases, functions and interaction design patterns. The instantiations of these concepts will be refined and added on during the whole project. Within the ALIZ-e project we focused on developing a social robot for long-term interaction with children and as the domain we chose behavior change for improving self-management of children with diabetes. To further specify this setting, taking into account the knowledge on the domain and users, we created eight use cases over the course of the project (see Table 2.1) describing the interaction in more detail. For more information on how these use cases were incorporated in experiments we refer to [149], in which an experiment is described containing most of the use cases.

Each use case contextualizes the methods and provides situational context of the

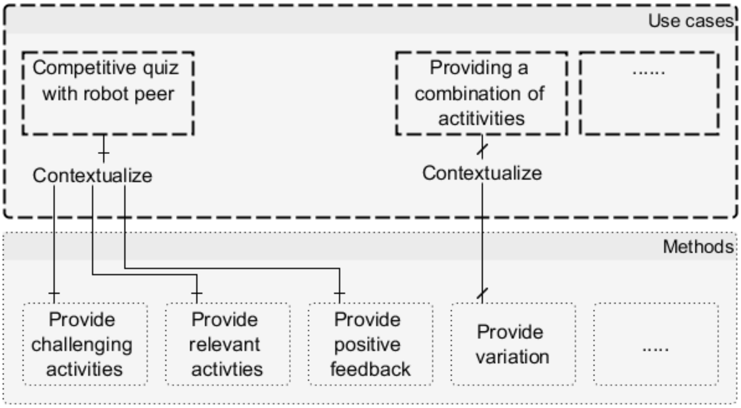



Figure 2.3: Use cases and how they contextualize methods.

effects that are measured. The competitive quiz for instance contextualizes methods which focus on competence, while providing a combination of activities is related to provide variation (see Figure 2.3).

## 2.4.4. FUNCTIONS

Based on the methods and use cases a selection of functions was implemented during the project. In Table 2.2 the functions used in the ALIZ-e experiments are named with a short exemplification next to it. We evaluated (parts of) these functions. Some of the more complex social behaviors like maintain social relationships are encompassed in for instance the function "personalize activities". Choosing the right level of function description is a bit of trial and error. We don't want the functions on implementation level, because this would complicate the picture sDR too much. The functions should be with enough detail to be able to relate them to specific methods and specific effects. You don't want the functions to encompass too little or too much, because the sum of the parts can be different than the sum of the whole. Some functions contribute to one method, others contribute to multiple methods. In Figure 2.4 this is shown, the functions related to the methods of Figure 2.3 are shown, but it is also shown that most functions are related to multiple methods and that these methods can be related to different objectives (see Figure 2.2).

## 2.4.5. INTERACTION DESIGN PATTERNS

There are many interaction design patterns possible for the use cases we looked at in ALIZ-e, but as social behavior and the emotions that come with this are very important. We looked at this in more depth. We looked for instance at the recognition of robot emotion expression for different robots (iCat and NAO) [127] and at the effect of embodiment (virtual or physical) on the effectiveness of social behavior [150]. Figure 2.5 shows how the different aspects of the voice and body influence the emotion expression and thereby the social behavior.

| Function | Exemplification |
|---|---|
| Personalize activities (based on personal info, performance, history etc.) | A game should be challenging and relevant, and small talk should be relevant |
| Provide multiple activities | The child should be able to switch between activities and the same objectives should be presented in different ways |
| Provide open questions | The child should have the opportunity to express him/herself |
| Disclose robot information | The robot should disclose "personal" information about itself, a background story |
| Adapt robot to child and activity – within boundaries | The robot should adapt its emotions to child and activity state. Be happy together with child, but also a bit sadder when losing. Recognizable emotion expression is necessary for this. |
| Provide acknowledgements | The robot should acknowledge what the user is doing |
| Provide compliments | The robot should provide compliments to the user on its actions |
| Exhibit social behavior | The robot should behave according to standard social norms; look behavior, turn taking, use of natural (non-verbal) cues (e.g. thinking behavior- uhmmm and gestures) |
| Show imperfection | The robot should not be all knowing and also need the help of the child sometimes |

Table 2.2: The different functions and an exemplification

### 2.4.6. EFFECTS

The expected effects are derived from literature about the objective and used techniques and from the functional design. Both the up- and downsides of an implementation should be defined so that in an experiment it can be validated if the upsides outweigh the downsides. We identified three levels on which these up- and downsides can be reported within human-robot interaction (leaving out pure technical evaluation):

1. The child perceives and comprehends the "intentions" of the robot
2. The robot perceives and comprehends the intentions of the child
3. The situated Human-Robot interaction

Within the ALIZ-e project we looked at "perceive and comprehend 'intentions' of robot" (1) and "situated human-robot interaction" (3) in the experiments. The experiments on recognizable emotion expression were on level 1, while the situations where there was interaction with the robot during an activity (quiz, sorting game, small talk etc.) were on level 3. On level 1 the interaction design patterns are evaluated and on

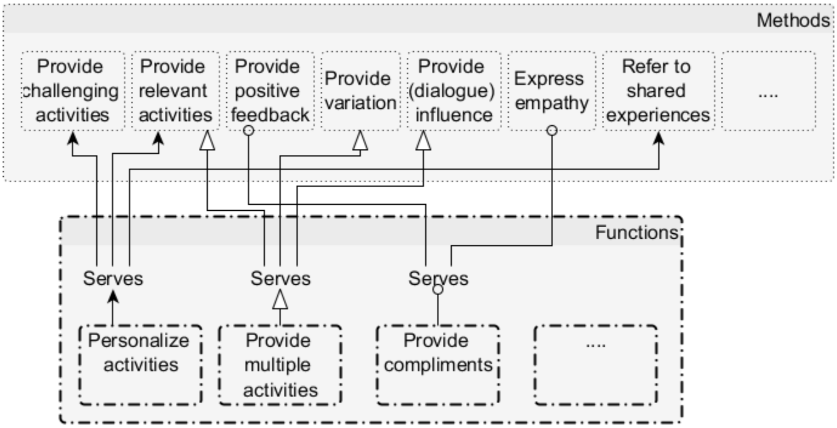



Figure 2.4: Methods and functions that serve them

level 3 the functions. The effects of the interaction design patterns are related to the functions they shape and of course the interaction design patterns. The effects of the functions are not only related to the functions, but also to the methods and objectives where the expected effects are derived from.

In Figure 2.6 a selection of the effects, and their related functions and instruments are shown. The effects show a direct relation with the objectives as effects on competence, autonomy and relatedness are expected. Next to this it can be seen that it is expected that most of the implemented functions, even all for this specific set of functions, contribute to the acceptance, trustworthiness, enjoyment and the robot being seen as empathetic. This set of expected effects is derived from the objective relatedness, from which this set is derived as being important. The relation back to the objectives is not drawn to make the sDR not unnecessarily complex, as these relations can also be found going back in the sDR. The interaction design patterns relate to their specific effects directly and indirectly via the function it shapes. The rules to adapt prosody for instance has a direct effect on understandability and an indirect, together with other patterns that shape the social behavior, on for instance trust.

### 2.4.7. INSTRUMENTS

After the expected effects are described there is a need to measure these. We prefer using objective instruments in combination with subjective instruments. Especially because it is known that children have the tendency to score extreme on questionnaires and there is thus a high chance on a ceiling effect. In Figure 2.7 it can be seen that although we would like to have objective measures, many are still subjective. Enjoyment is measured with a questionnaire and observations and emotional appearance and understandability both have questions for the child to check recognition of either emotion or spoken text of the robot. Having a forced choice question does eliminate some of the problems of a questionnaire, but it also means there is a need for a within subject design and this is not always feasible with specific user groups.

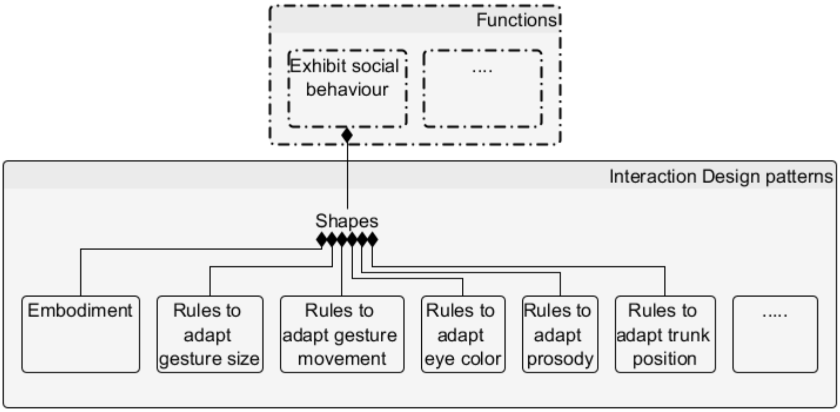



Figure 2.5: Interaction design patterns shape functions

## 2.5. EVALUATION OF THE SDR

The previous section described the sDR using the ALIZ-e project as an example. This section will show how a specific design and evaluation cycle can be supported by the creation of an sDR. In this cycle, a model for adaptive emotion expression for a NAO robot was developed. The robot's internal valence and arousal values were influenced by emotional state of the child and emotional occurrences in the activity (e.g. winning the game). This adaptation of internal values led to a change in voice, posture, whole body pose, eye color and gestures to express its emotional state. In an experiment 18 children (mean age 9) played a quiz with two NAOs consecutively (within subject design). One of the NAOs adapted its emotions according to the model and the other did not. A more detailed description of the method is provided in [242]. The objective this experiment focused on *relatedness* and the method *adapt to others' behavior*. The function to serve this method was *adapt robot to child and activity – within boundaries* in the use-case *quiz*. Effects were expected on *emotional contagion, preference, relatedness, empathy, acceptance, trust, fun and motivation*. Relatedness as effect is directly related to the objective of relatedness, the other expected effects are derived from literature on relatedness as being contributing factors to relatedness. The instruments were *arousal and valence observations, forced choice preference, specific questionnaires for relatedness, empathy, acceptance, trust, fun and motivation and open questions related to these aspects*. Figure 2.8 shows the sDR of this specific experiment, we limited the number of relations in comparison to Figure 2.1 by excluding the relation between effects and objectives and use cases and functions, both can be derived by following the other relations.

### 2.5.1. RESULTS

The objective results on arousal and valence observations showed that the children were significantly more expressive (smiling more) when interacting with the affective robot in comparison with the non-affective robot (M=33.59, SE=17.34) than for the non- af-

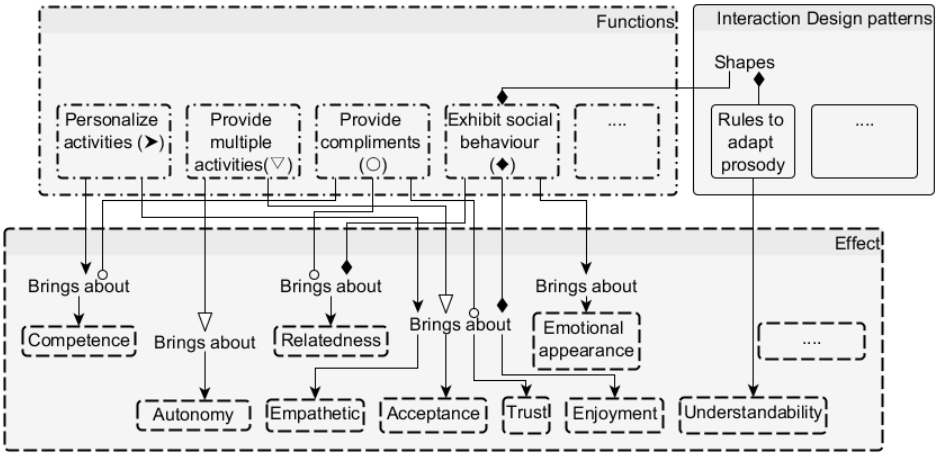



Figure 2.6: Functions and interaction design patterns bring about effects

fective robot (M=29.06, SE=13.53), (t)(16)=2.156, (p) <0.05, (r)= 0.47 (one-tailed). The answers on the questionnaire on robot-child interaction showed a ceiling effect. Both robots scored very high and the difference was not significant for any of the question topics. In the second questionnaire the children had to choose between one of the two robots on different aspects (e.g. fun, trust) and in the end prefer one of them. There were differences, but non were significantly different, although on trust there seemed to be a trend in favor of the non-affective robot. Finally they were also asked about their motivations to choose one or the other. The most noticeable motivations were clearly that the non-affective robot was more understandable, while the affective robot was preferred most often because it showed emotions.

### 2.5.2. EXPERIMENT AND sDR CONCLUSIONS

The expression results are quite clear and show a significant effect for the emotional contagion, but this positive effect is not supported by the questionnaires. These suffer from the ceiling effect; only with forced choice some differences can be seen, but still not large. Notwithstanding these ceiling effects we can conclude from the observations that adaptive emotional expressivity influences children to engage in more positive expressivity.

Another interesting result is "trust" where we see that the non-affective robot scores (non-significantly) higher than the affective one. Looking back at the sDR this means that a robot that adapts its state to the child is less trustworthy and might involve lower relatedness. Based on the results we are not ready to conclude this, because it could also be that the sDR is not complete. Reinvestigating literature we see that emotional voices can suffer from understandability issues [127]. This is also supported by the responses the children provide, where they indicate the non-affective robot is more understandable. Understandability is a known factor for trust in automation [107]; in addition, literature on trust of children in caretakers with an unfamiliar accent [132] indicates that understandability influences trust. We have to add understandability thus as a possible

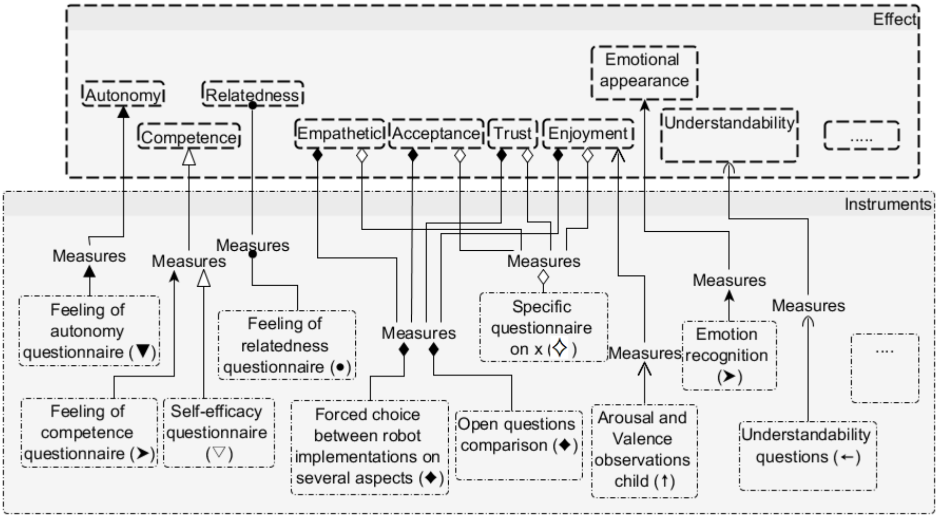



Figure 2.7: Effects are measured with specific instruments

downside for prosody which can be measured asking directly about understandability and in concurrence look at effects on trust and acceptance. Figure 2.9 shows the changed portion of the sDR.
The sDR shows the decisions made for the design of the experiment, this makes it possible to relate the negative result on trust back to the function that was implemented. It shows the sDR is not discriminatory enough on the effects and that this can be improved by adding a branch to indicate that an interaction design pattern could have influenced the trust. Finally, the experiment provides confirmation that adapting the emotion of the robot to the emotion of the child and activity has a (mainly) positive influence, which can be used for theory building on emotional adaptivity.

## 2.6. CONCLUSION AND DISCUSSION

The objective of this paper was to provide a formal template that supports the systematic design and evaluation of an experiment and reason about the effects and decisions afterwards. We reached this objective by formalizing the relations between theory, design specifications and evaluations and guidelines for creating it. The developed sDR supports the systematic, iterative and incremental design and evaluation of social robots for behavior change.

To come to this sDR we had to answer three questions. First, we had to specify the relevant concepts. We used the concepts as defined in the sCE method. Second, the relations had to be identified. For this identification we used knowledge on behavior change, social robotics and design specifications.

To make the decisions visible and to support reasoning about the effects and reusability we had a third question on representation of the concepts and their relations. We decided on using concept mapping to visually relate the concepts.

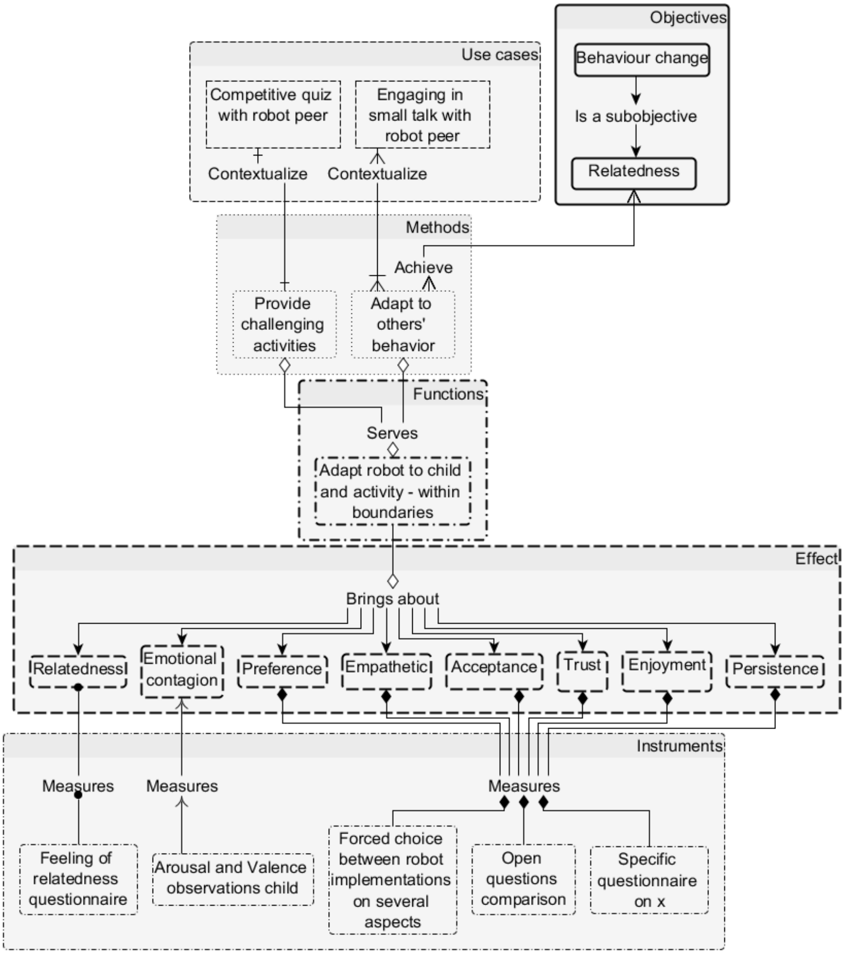



Figure 2.8: situated Design Rationale of emotional contagion experiment. The use cases "bring about" the effects, but for readability reasons we excluded this line from the overview as we did with the "demonstrate" lines from effects to objectives. [242]

After answering these three questions the sDR method was explained by instantiating the generic sDR template with the European ALIZ-e project. We walked through every concept and its relations to other concepts and also showed how the knowledge from theories and empirical evaluations are taken into account in this process. The complete sDR of the ALIZ-e project can be found here (https://bit.ly/2RXxWNd). It is interesting to see that, when multiple experiments are concatenated in one overall project sDR, the objectives, methods and their related effects and instruments stay stable over the course of the project. Use cases, functions and interaction patterns on the other hand are added, removed and refined according to the project's progress. This relatively stability of the sDR supports adapting and extending.

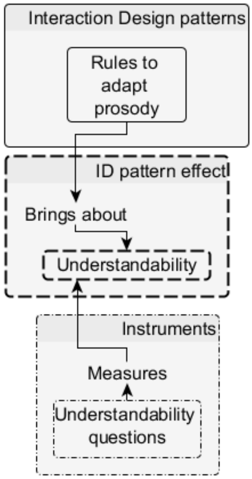



Figure 2.9: Refinement of sDR, based on emotional contagion experiment.

At this moment it is not hard to create an sDR for one experiment, as the decisions that are described in the sDR are decisions you take anyway. Which objective do you have with the project, what methods can be used, what functionalities do you want to address in this specific experiment and what effects do you expect and how do you measure these effects? By creating the sDR before performing the experiment shortcomings in the experimental setup can be found.

After the experiment is finished and you would like to do another experiment with the same objectives but other functionalities the sDR can be extended, the easy thing is that the sDR already shows decisions you don't have to think about anymore, the hard thing is to incorporate the new experiment in the old sDR. Sometimes this is easy, e.g. when the functions and expected effects are really different. Other times this is harder, when new interrelations between for instance functions and effects appear. When this happens it means you have to rethink the definitions and try to concatenate or split functions to make the relations less complex or ambiguous. This stipulates the importance of having an ontology in which the concepts are defined, so others also know what is meant by it and can reuse it.

The use of sDR was further exemplified with a specific experiment. In this experiment we could see how sDR supports design and evaluation, the sDR can be adapted after interpretation of the results of the evaluation. With sDR we can reason on why a certain effect occurred (e.g. why did the effect on trust differ from the other effects?). As can be seen Figure 2.6 there is quite some overlap in effects for different functions in the current sDR of the ALIZ-e project, showing the interactions but also resulting in ambiguous results. This could be improved by identifying claims that are specific for a function or by changing the level of function description, but it will never be perfect needs continuous improvement. By making this possible it also creates the opportunity to identify elements that need to be added to aid the design and evaluation (e.g. experimental support on the design pattern of prosody).

Finally, sDR supports iterative and incremental theory building by showing which elements are validated, which are invalidated and which need more research and/or validation, all within a specific context. Theory building is possible, because the reasoning of the whole chain, from theory to instrument is clarified in the sDR, making it also possible to transfer the ideas to other domains and evaluate it there for more generalizable theories.

Although this is all desirable, it asks for well thought over decisions of the chosen effects and instruments. A further complication that we will not solve is that there can be relations we did not foresee resulting in unexpected effects or incorrect attribution of effects to certain functions.

Notwithstanding these complications sDR provides a method to evaluate a complex system, such as a social robot for behavior change, meanwhile getting an idea of the interaction between functionalities. These interactions are important, because a complex system is never just a combination of its parts. The awareness of interrelations makes it possible to create theories on a level that is fitting to what is “really” known. Furthermore, we will be able to distinguish between groups of outcomes and combine this with user characteristics to develop user profiles which can be used for fast adaptation of the interaction. This will be further explored in the PAL project, a H2020 project on behavior change for self-management of children with diabetes. We foresee reuse of the objectives, most of the methods, effects and instruments with refinement and extension of functionalities more focused on behavior change from the ALIZ-e sDR.
Next to this, by putting relations and concepts in an ontology we further formalize the sDR and make it in this standard format available for people outside the projects. This way, the research community can make use of the knowledge progress on social robots and avatars for children. The complete overview and the experiment specific sDR provide an elaborate guidance in understanding the decisions and the possibility to replicate it. We believe this will open the way to generalizing the results and applying it in other domains.

### 2.6.1. FUTURE WORK

This paper focused on formalizing, reporting and sharing of the design rationale. It’s essential to share this rationale with the research and design community and for this we will need an easy to use, preferably interactive, tool. This tool should support the creation of sDRs so they are easier to create, extend and understand. The sDR is now lacking a tool for visualization, the structuring of lines is currently a (mostly) manual and labour intensive job. This is a drawback for creating, adapting and extending an sDR. We would therefore like to develop a tool like sCE has for relating use-cases, expected effects and functions to each other `www.scetool.nl`. This should be extended with a good visualization tool, like they exist for network analyses (e.g. cytoscape.js - `js.cytoscape.org`). With the addition of the related ontology, code and information on the experiment it should then be possible to reproduce the experiment. At the moment the experimental code for the PAL project is stored at a GitHub repository with version numbers for each experiment, and we have the relevant sDR. Sharing this to the research community in a more structured manner should be possible in the future.

Another addition could be to visualize the expected positive and negative effects, this would be similar to sCE where positive and negative claims are made explicit. This will make the sDR both more informative and more complex, so we should think about how to visualize this.

## ACKNOWLEDGEMENT

This work is (partially) funded by the FP7 ALIZ-e project (grant number 248116), www.aliz-e.org and the H2020 PAL project (grant number 643783), www.pal4u.eu.

# I

# Interaction patterns within the design rationale

This Part contains:

- J. Kessens, M. Neerincx, R. Looije, M. Kroes, and G. Bloothooft, *Facial and vocal emotion expression of a personal computer assistant to engage, educate and motivate children*, in Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009 [127]
- I. Cohen, R. Looije, and M. Neerincx, *Childs perception of robots emotions: effects of platform, context and experience*, International Journal of Social Robotics **6**, 507 (2014). [55]
- R. Looije, A. Van der Zalm, M. Neerincx, R.-J. Beun, et al., *Help, I need some body: the effect of embodiment on playful learning*, in *RO-MAN*, (IEEE, 2012) pp. 718-724. [150]

## Influential interaction design patterns

In this Part we will discuss three papers that show the importance of design choices when implementing functions. Figure I.1 shows the sDR of this Part. In chapter 2 the sDR as method for design and evaluation is described. All papers in this dissertation contribute to the design and evaluation of a social robot for children with diabetes that supports progressive self-management. This means the objective, behavior change, is the same in all parts. The same is true for the self determination theory related subobjectives: competence, autonomy and relatedness [74]. The interaction patterns we will discuss here all relate to expression of emotions by the social robot. Expression of emotions shapes the function "exhibit social behaviors". This specific function contributes to two methods from Motivational Interviewing [207] that are used to achieve relatedness [238]. The effects that are relevant when evaluating these design choices are thus related to the design choice itself (e.g. understandability). Next to this, the effects are also relevant for knowledge on the influence the design has on the effect of the function it shapes (e.g. trust). Finally, Figure I.1 shows the instruments that are used to measure the effects.

### Emotions in different embodiments

The chapters in this Part will answer the following three research questions:

1. chapter 3: Multimodal emotions of a facial expressive robot [127]
   1.1. Design question: How to model the four Ekman emotions of anger, fear, happy and sad in the face and speech of the iCat, so that they are recognizable for children?
   1.2. Hypothesis: Children will show better understanding, acceptance, trust, fun, empathy and performance when interacting with an iCat that expresses multimodal emotions (i.e., increasing in the following order: no emotions, facial, facial-and-vocal).
2. chapter 4: Bodily expressive robot versus facial expressive robot [55]
   2.1. Design Question: How to model the five Ekman emotions of anger, fear, happy, sad and surprise in the postures and LEDs of the NAO, so that they are recognizable for children?
   2.2. Hypothesis: Three factors influence the recognition rate of robot's emotions: (1) the recognition rate differs between robot embodiments, (2) the rate is higher when the emotions are expressed in a congruent context (compared to no context), and (3) the recognition improves over time.
3. chapter 5: Physical versus virtual embodiment of a robot [150]
   3.1. Hypothesis: Children's performance, attention, trust, enjoyment and preference in quiz task are higher, when interacting with a physical NAO compared to a virtual NAO.

The overarching research question of these three chapters is: What is the effect of expressing emotions in different contexts, using different embodiments and having single or multiple interactions on emotion recognition, engagement, motivation and performance?

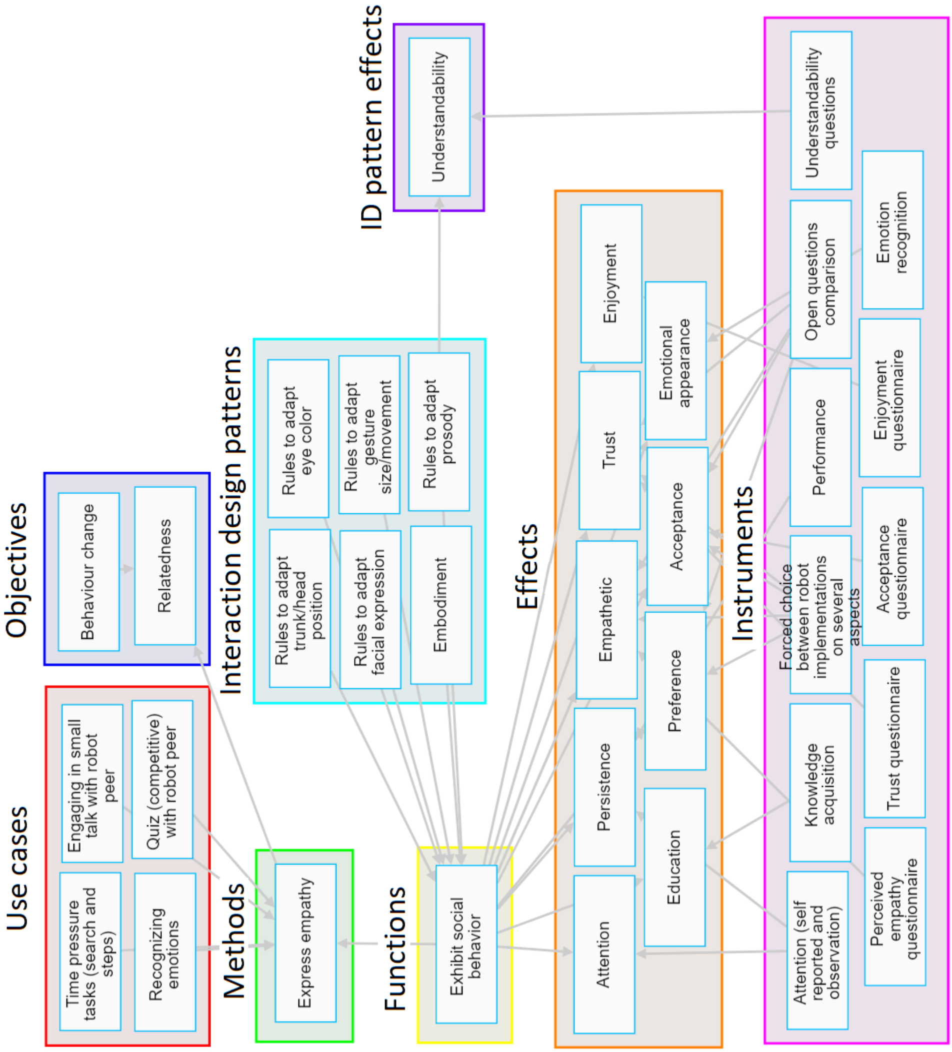



Figure I.1: situated Design Rationale of Part I sDR supports answering the overarching question of this part with support of the figure created: What is the effect of expressing emotions in different contexts, using different embodiments and having single or multiple interactions on emotion recognition, engagement, motivation and performance? As it supports incremental development sDR can be used to incorporate the effects of all three experiments. Figure I.1 shows in one view what aspects are taken into account during the design and evaluation and what the rationale was behind these decisions. We can conclude that the emotions expressed by a robot can be recognized well [127, 55], especially when there are multiple interactions and context is provided [55]. It should be noted that the understandability of the emotional voice had a detrimental effect on trustworthiness of the robot [127]. A physical robot was shown to have a positive effect on attention without a detrimental effect on performance, compared to a virtual robot [150].

With the support of sDR we can reason about unexpected effects. Kessens [127] shows that the understandability of the voice influences the child experience of the robot. The difficulty of understanding the voice, has a detrimental effect on the trust infested in the robot. This disambiguation is possible by making the relations between effects, functions and design decisions clear. It also shows the influence design choices have of on the effectivity of a function.

Finally, the sDR shows that more experiments in the interaction design patterns for emotion expression are possible, but it also shows that the objectives (recognizable emotion expressions) are reached.

# 3 | Facial and vocal emotion expression of a personal computer assistant to engage, educate and motivate children

**Abstract**

*The general goal of our research is to develop a personal computer assistant that persuades children to adhere to a healthy lifestyle during daily activities at home. The assistant will be used in three different roles: as companion, educator and motivator. This study investigates whether the effectiveness of the computer assistant with an iCat robot embodiment, can be improved when it expresses emotions (tested for each of the three roles). It shows that emotion expressions can improve the effectiveness of the robot to achieve its role objectives. The improvements that we found are small, however, probably due to a ceiling effect: All subjective measures are rated very positively in the neutral condition, thus leaving little room for improvement. It also showed that the emotional speech was less intelligible, which may limit the robots' effectiveness.*

## 3.1. Introduction

Only 50% of the chronically ill adheres to their physician's advice [216]. These persons experience difficulties to perform the desired daily activities if the frequency of visits to the physiotherapist or dietitian decreases. A personal computer assistant at home, which explains why chronically ill should perform the daily activities related to their therapy and supports them in adhering, could help. For example, the Healthbuddy®is a commercial product that supports people with a chronic disease, such as COPD and diabetes, to adhere to their therapy. In the Healthbuddy, Motivational Interviewing (a behavioral change therapy) is used, which relies on the willingness of the users to change [161]. To realize a change in behavior, people need to know why they should adhere to their therapy, and they should receive support in their efforts.

Not only adults may suffer from chronic diseases, but also an increasing number of children suffers from obesity and diabetes [75]. These children are often afraid to get excluded from their classmates when they adhere to their therapy (e.g. decline sweets). Feeling different than the others is for children a bigger problem than for adults. Therefore, children could possibly benefit even more from a non-judgmental personal assistant to engage, educate and motivate them in therapy adherence.

A personal assistant should be able to understand and express emotions to maintain an engaging relationship with the user. This is supported by research on persuasive technology [90], affective computing [192] and the media equation [201]. This research shows that the more human-like the interaction with a computer is, the more persuasive, engaging and fun the interaction is.

Our main goal is to develop a robot prototype that can act as a personal assistant and has the educating and motivating skills to persuade children to perform desired behavior. In a study by Blanson Henkemans et al. [30], a personal computer assistant was developed for persons with overweight. The computer assistant was represented by an animated virtual robot - the Philips 'iCat'- which could show different emotional expressions in the face (e.g., to express empathy). It was shown that the iCat reduced the 'standard' decline in motivation to perform self-management, and it also helped to lower the body mass index. In two other studies ([145, 146]), it was shown that both adults and children enjoy working with the iCat as a personal assistant, both with the virtual and embodied version. This appeared to be particularly the case when it expressed the relevant facial emotion expressions. However, the various roles of the iCat were rather limited in these two experiments. Therefore, this paper aims at improving iCat's persuasion capabilities in two steps. The first step concerns the enrichment of the three roles as follows: as companion, educator and motivator. The *companion robot* gives emotional support and can act as a playmate. The *educational robot* teaches and explains why adherence is important. The *motivational robot* encourages adherence and a change in lifestyle. The second step focuses on the enhancement of the emotional expressiveness of the iCat by using emotional speech.

The current study investigates the effectiveness of the iCat's emotion expressions on children through both objective and subjective measures. In the experiment, three different emotion conditions are compared: a neutral condition, a condition with only facial emotional expressions, and a condition with both facial and vocal emotional expressions. Our general hypothesis is that the iCat which uses facial and vocal emotions scores better on all measures than a neutral iCat or an iCat that uses only facial emotion expressions.

## 3.2. Method

### 3.2.1. Research questions

Various body parts can be used for emotion expression, e.g. the face, the voice, the body posture and hand movements [82]. Several studies have been performed that investigate the effectiveness of facial emotion expressions of an electronic assistant (e.g. [30, 145, 146]) and of emotional synthetic speech (e.g. [82, 43, 137, 167, 219, 220]). However, to combine facial emotion expressions with emotional speech is quite new. We chose to model four emotions (anger, fear, happiness and sadness), as these emotions are denoted as 'basic emotions' [184] and could be modeled in both the face as speech of our robot. The research question is whether the effectiveness of the human-computer interaction can be improved by adding facial and vocal emotion expressions. Effectiveness is the degree of success with which the roles' objectives are achieved, consisting of measures for companionship, education and motivation.

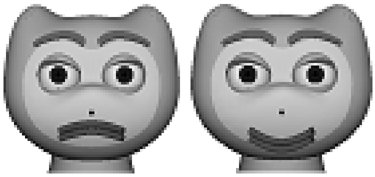


Figure 3.1: Most extreme lip positions of the iCat



### 3.2.2. THE ICAT ROBOT

The experiment was conducted with the iCat robot. This robot was designed as a human-robot interaction research platform by Philips Research [191]. The iCat robot is a social interface which means that it can express itself emotionally through its facial and vocal expressions. There are two versions of the iCat robot, namely an embodied one and a virtual one. In this study, the embodied robot was used, as previous research has shown that it is considered more attractive by users [69].

### 3.2.3. FACIAL EMOTION EXPRESSION

The way in which humans express emotions with their face has been studied extensively in the past (i.e. [82, 229]). In [229], a literature review is given on how emotions can be classified by the use of facial muscles. Obviously, this overview is not exhaustive as a specific emotion can be expressed in different ways [82]. For instance with the emotion anger, the lips could be pressed together firmly or the mouth could be opened.

In order to express emotions in its face, the iCat is capable of moving its eyebrows, eyelids, lips, eyes, and head. In comparison with humans, the iCat has limitations in its degrees of freedom to express emotions within its face, e.g. the lips are connected to four servo motors, thus restricting the possible ways the lips can be moved. As an example, Figure 3.1 shows the most extreme lip positions of the iCat robot. When the robot speaks, the lips open and close. These lip movements are identical for all speech sounds.

In a pilot study [137], the original emotion settings developed by Philips ('original settings') and modified emotion expressions were investigated ('modified settings'). The emotion expressions were modified in such a way that they are more similar to the results of Ekman's study [82]. We conducted a pilot study, in which 19 children of between eight and nine years old (mean=8.36, sd=0.5) viewed sixteen videos with six facial emotions of the iCat (anger, fear, happiness, sadness, surprise and disgust), and whom had to make a forced choice on the emotion expressed. In the experiment described in this paper, we used the settings from the pilot study with the highest recognition rates. Table 3.1 gives an overview of the emotions and settings that we used in the experiment and the corresponding recognition rates from the pilot study. It can be seen that the mean recognition rates are quite high, except for the emotion fear.

Furthermore, in the experiment, the emotional iCat follows the participant with its head and eyes, whereas the neutral iCat is placed with its face in the direction of the participant. In all conditions, the iCat blinks and nodds now and then.

| Emotion | Recognition | Settings |
|---|---|---|
| Anger | 83.3 | modified |
| Fear | 38.9 | original |
| Happy | 66.7 | modified |
| Sad | 66.7 | original |

Table 3.1: Recognition rates (%) of facial emotion expressions.



### 3.2.4. VOCAL EMOTIONAL EXPRESSION

As with emotional expressions in the face, also emotional speech has been analyzed extensively [43, 167, 219]. Four categories of speech parameters are commonly used to classify the parameters that affect the vocal emotion expressions; pitch, timing, voice quality and articulation. The type of speech parameters that can be manipulated depends on the speech synthesis technique that is used. As noted in [220], there exists a trade-off between the flexibility of acoustic modeling and the perceived naturalness. For formant synthesis, also known as rule-based synthesis, the artificial speech is entirely created by rules, without using pre-recorded human speech. This type of synthesis is therefore very flexible, but it is isn't perceived as sounding natural. As opposed to formant synthesis, unit selection sounds very natural. During synthesis, the most appropriate units are selected from a large database with speech recordings. Manipulation of speech parameters in order to express emotions, is less straightforward, however. In our application, the speech is used as a voice for a robot, and therefore the naturalness of unit selection voices are not needed. We choose to use diphone synthesis (and not formants synthesis) as for this type of synthesis there exists emotion manipulation software. In diphone synthesis, speech is generated by concatenating diphones (transitions of two speech sounds), which are extracted from human speech recordings. Pitch (F0) and timing (duration) parameters can be easily manipulated, but manipulation of voice quality and articulation is very complicated.

In a previous study [137], we compared different methods to generate emotional diphone speech synthesis. The general method consists of three steps. In the first step, a speech synthesis system is used to automatically generate from text: the speech sounds, their duration and intonation pattern (pho-files). In the second step, the pho-files are manipulated corresponding to the intended emotion, or left unchanged for the neutral emotion. In the third step, the speech is generated with MBROLA diphones [156]. In the pilot study, we used the MARY [156] system, to generate the English pho-files and the English female MBROLA diphones to generate the speech files. In the current study, we used the Dutch, female voice ‘Diana’ of Fluency [156] to generate the pho- and speech files. In the pilot study, we compared emotion settings from two emotion editors - Emofilt [40] and EmoSpeak [220] - and copy synthesis. From the pilot study it appeared that the highest recognition results are obtained with EmoFilt. For this reason, we decided to utilize Emofilt in this study as well. In this study, we focus on emotion expression of sentences with matching semantics, as these are most relevant for the application we have in mind. Table 3.2 shows that the recognition rates are higher if the semantics of the sentences match the intended emotion[1].

[1] differences are not significant

| Emotion | Neutral semantics | Matching semantics |
|---|---|---|
| Anger | 52.5 | 72.5 |
| Fear | 42.5 | 65.0 |
| Happy | 33.3 | 45.0 |
| Sad | 52.5 | 62.5 |

Table 3.2: Recognition rates (%) of vocal emotion expressions.



### 3.2.5. EXPERIMENT

#### PARTICIPANTS

In the experiment, children of eight and nine years old were tested (mean = 8.5, sd = 0.5). Eighteen children (10 male, 8 female) participated in the experiment. The children were recruited from a primary school in Soesterberg. For participation, the children received a gift voucher and a photograph of themselves with the iCat robot.

#### THE ICAT'S ROLES

Three different tasks were performed, for each task the iCat played a different role:

Companion: The companion robot gives emotional support and can act as a playmate. This is implemented by letting the iCat play the role of an electronic pet. To this end, the iCat asked the participants to imagine that it was their pet (a cat). Next, a story was told by a professional male voice. The story consists of emotional events in the life of the cat. Each event ends with an emotional expression by the iCat, for instance: "I'm certainly not going outside! There is a big and scary dog" (fear). Participants were asked to classify the emotional state of the iCat by performing a forced choice between five possible emotions (anger, fear, happy, sad, neutral).

Educator: The educational robot teaches and explains why adherence is important. For this, the iCat plays the role of a quizmaster. The subject of the quiz is 'korfbal', a typical Dutch team sport that has similarities to basketball. First, the iCat robot explained the rules of the sport and afterwards questions are asked about those rules. The sport korfball was selected as it is not a popular sport among children, and thus, most of the participants doesn't have foreknowledge. The rules told by the iCat come from an instruction manual of the Royal Dutch Korfbal Union [134], especially written for eight to ten years old children. After each question, participants had to choose the correct multiple choice answer (three alternatives). Feedback about the answers was provided to the participants by the iCat robot.

Motivator: The motivational robot encourages to adherence and to healthy lifestyle. In the experiment, the iCat played the role of coach. The iCat tried to motivate participants to perform two different tasks. The first task was a search task. The children had two minutes to collect as many blue marbles from a large basket filled with richly colored marbles. The second task consists of a physical exercise. The children had to make as many steps as possible in two minutes time. These steps were registered by a step counter. The iCat tried to motivate the participants with statements like "keep on going".

Figure 3.2: Experimental setup

### EXPERIMENTAL CONDITIONS

As children have a limited attention span, we had to minimize the number of test conditions. For this reason, we decided to test only three possible emotion conditions: no emotions (neutral), only facial emotions (face only), and both facial and vocal emotions (face & speech). We left out the condition with only vocal emotions, as we expect that the effect of facial emotion expressions is larger than that of vocal emotion expressions. We used a between subjects design for the emotion conditions, thus, each participants tested two emotion conditions. The order of the emotion conditions (neutral, face only, face & speech) and the order of the iCat role (companion, educator, motivator) were balanced across participants. For each task, two versions were available (two stories, two quizzes and two types of tasks), which were always presented in the same order.

### EXPERIMENTAL SETUP

During the experiment, the participant and experimenter were seated in different rooms. The participant was placed in a room that was decorated as a living room. Two iCats were placed on a table in front of the participant, see Figure 9.3. A webcam, placed between the iCat and the participant, was used to make recordings of the participant's face. Each participant worked with two different emotion conditions which were coupled with the iCat. In this way, it was clear for the participants that the two iCats have distinct personalities. Behind the seat of the participant, two video cameras were placed on the wall. The video information and the information from the webcam was sent to the experimenter in the other room. A Wizard-of-Oz setting was used, which means that the experimenter emulated the speech recognition.

In the experimental setup, it was taken into account that the attention span of children is limited. Therefore, the experiment as a whole was restricted to an hour. The experiment is introduced by the two iCats and the experimenter. Next, the three tasks are performed. Finally, the experiment is finalized by filling in the general questionnaire. An overview of the structure of the experiment is given in Table 3.3.

| Part of experiment | Questionnaire | | Time |
|---|---|---|---|
| introduction | A | | 12 min |
| task X version 1 | B | C | 14 min |
| task X version 2 | B | | |
| task Y version 1 | B | C | 14 min |
| task Y version 2 | B | | |
| task Z version 1 | B | C | 14 min |
| task Z version 2 | B | | |
| finalization | D | | 5 min |

Table 3.3: Structure of experiment; "X/Y/Z" refers to the tasks in which the iCat plays the role of companion, educator or motivator.

| list | nr | Measure | |
|---|---|---|---|
| A | 1 to 4 | | none |
| | 5 | experiment | expected fun |
| | 6 | participant | experience |
| B | 1 | robot | fun |
| | 2 | robot | intelligibility |
| | 3 | task | fun |
| | 4 | task | difficulty |
| | 5 | task | acceptance |
| C | 1+2 | robot | difference |
| | 3 | robot | preference |
| D | 1 | experiment | fun |
| | 2 + 5 | robot | acceptance |
| | 3 + 6 | robot | empathy |
| | 4 + 7 | robot | trust |
| | 8 | robot | preference |

Table 3.4: Overview of questionnaires.

| Emotion | neutral | face only | speech & face |
|---|---|---|---|
| Anger | 83.3 | 87.5 | 83.3 |
| Fear | 79.2 | 83.3 | 83.3 |
| Happy | 75.0 | 83.3 | 79.2 |
| Sad | 75.0 | 100.0 | 91.7 |

Table 3.5: Recognition rates (%) for the various emotions.



#### MEASURES

The effectiveness of the iCat was tested both subjectively and objectively. Objectively, performance and reaction times were measured. The performance measures were: the recognition rates of the emotions (companion), the percentage of correct answers to the quiz (educator), and the number of steps taken or marbles collected (motivator). Subjectively, lists with questions were asked. An overview of the subjective measures per questionnaire is given in Table 3.4. Results were analyzed with ANOVAs and Tukey HSD post hoc tests (if appropriate).

## 3.3. RESULTS

### 3.3.1. GENERAL QUESTIONS

On the question about the foreknowledge on korfball, nobody answered that they were actually playing korfbal. Half of the participants answered that they had played it once or twice and the other half that they hadn't. Participants who already knew something about korfbal did not answer more questions right than those who didn't ($F(1,178)=0.026$, $p=.87$).

### 3.3.2. OBJECTIVE MEASURES

In the task in which the iCat plays the companion role, the emotion recognition rates were determined. The recognition rates per emotion are presented in Table 3.5. Three important conclusions can be drawn. First of all, comparison with the recognition rates in Table 3.2 shows that the recognition rates in the neutral condition are higher than in the pilot study. This result indicates that the classification of emotions out-of-context is more difficult than in-context (short story). Secondly, despite these high recognition rates, facial emotion expressions further improve the recognition rates. Thirdly, the additional use of vocal emotion deteriorates (or doesn't improve) the recognition rates of the various emotions. None of the differences are significant.

Other objective measures include reaction times and performance. Performance measures where recognition rates (companion), percentage of correct answers to the quiz questions (educator) and z-scores (motivator). For the motivator task, z-scores were used in order to take the data for the two types task together in one analysis. The objective measures are summarized in 3.6.

The objective measures show that emotion expressions of the iCat tend to have a positive effect: The recognition rates of the emotion grow, the reaction times in the edu-

| iCat role | Measure | neutral | face only | speech & face |
|---|---|---|---|---|
| companion | % correct recognition | 78.1 | 88.4 | 84.4 |
| | reaction time (in sec) | 5.7 | 5.9 | 5.6 |
| educator | % correct answers | 75.0 | 66.7 | 66.7 |
| | reaction time (in sec)* | 5.9 | 5.9 | 4.7 |
| motivator | z-score | -0.28 | 0.03 | 0.25 |

Table 3.6: Objective measures (significant differences (p<0.05) are indicated with an asterix ('*').

| | measure | neutral | face only | speech & face |
|---|---|---|---|---|
| task | difficulty* | 3.64 | 3.69 | 3.75 |
| | acceptance | 4.08 | 4.33 | 4.28 |
| | fun | 4.39 | 4.67 | 4.56 |
| robot | fun | 4.53 | 4.67 | 4.53 |
| | intelligibility* | 4.42 | 4.22 | 3.72 |
| | acceptance | 4.67 | 4.50 | 4.75 |
| | empathy | 5.00 | 4.75 | 4.75 |
| | trust | 1.92 | 2.00 | 2.00 |

Table 3.7: Subjective measures (significant differences are indicated with an asterix ('*').

cator task[2] are faster when using both speech and face expressions of emotion, and also the performance of the participants seems to improve if they are motivated by an iCat in that condition. However, ANOVA and post-hoc analyses show only one significant difference: the reaction time of the educator-iCat is significantly shorter for the 'speech & face' condition compared to the other two conditions.

### 3.3.3. SUBJECTIVE MEASURES

In Table 3.7 the results of the subjective measures are given. The trends do not differ between the different iCat roles (companion, educator, motivator). Therefore, the results are presented together. From Table 7 two main conclusions can be drawn. Firstly, all subjective ratings (except the ratings for 'trust') have high values. Secondly, almost all differences are non-significant: Post-hoc analyses show that the educator task was judged as more difficult than the motivator ($p<.005$) and the companion task ($p<.05$). Also, the speech intelligibility of the 'speech & face' condition was rated lower than the conditions 'neutral' ($p<.01$) and 'face only' ($p<.001$).

After performing each subtask, participants were asked whether they perceived a difference between the two iCats and which one they preferred. They did not often perceive a difference between the two iCats, see Table 3.8, especially not during the motivator task. However, again none of the results shows significant differences. Furthermore, participants significantly preferred the iCat with only facial emotion expressions compared to an iCat that uses both vocal and facial emotion expressions (t-test, $p<.1$). This indicates that neutral speech was preferred over emotional speech, possibly because it was perceived as more intelligible (see Table 3.7).

[2] After removing outliers (data points that are more than three standard deviations removed from the mean)

| **Comparison** | **difference** | **preference** |
|---|---|---|
| *face only* vs. *neutral* | 44 | 44 |
| *speech & face* vs. *neutral* | 50 | 47 |
| *face only* vs. *speech & face* | 61 | 71* |

Table 3.8: Perceived difference and preference between emotion conditions, given as a percentage, and first preferred over second condition. Significant differences (p<0.1) are indicated with an asterix ('*').



## 3.4. DISCUSSION AND CONCLUSIONS

In this study we successfully designed and implemented three robots, which all proved to perform their roles very well i.e, they were engaging, motivating and educating. The results of this experiment indicate that adding emotion expressions to a companion robot might improve its effectiveness. Objective measures showed improvements. Firstly, the recognition rates of the iCat with facial emotion expressions are better than that of a neutral iCat. Secondly, the iCat with both vocal and facial emotion expressions seem to motivate the children more to perform their task (searching marbles and making steps) compared to a neutral iCat. Finally, the reaction times to quiz questions improved significantly for the condition when vocal and facial emotion expressions of the iCat were applied. However, the differences we found were very small, and in general non-significant. The effectiveness of the robot was also measured subjectively. The subjective measures show limited improvements, which might be explained by a ceiling effect: Almost all aspects of the iCat (e.g. empathy, fun, acceptance) were judged already very positively in the neutral condition, thus leaving ample room for improvement. It also showed that - although vocal emotions expression can be recognized correctly – the speech is less intelligible. This effect is stronger for the iCat in the educator task than in the motivator task. This is not surprising, as for the educator task the *content*of the speech is more important, and for the motivator task, the *expressivity* is more important. Actually, emotion in speech should be tailored to the specific role or task objectives that have different priorities for correct content recognition and perceived expressivity. In this experiment, reduced intelligibility of emotional speech explains why the emotion recognition rates do not further improve when in addition to facial emotions also vocal emotion expressions are used. It may also explain the result that the iCat that only uses facial emotion expressions is significantly preferred over a robot that uses both facial and vocal emotion expressions. Although the speech intelligibility may sound reduced, it does not necessarily mean that less words are understood correctly. Objective evaluation might reveal whether besides the speech quality, also the understanding of the speech is lowered. It could also be worthwhile to investigate whether speech intelligibility is not affected if less extreme vocal emotion expressions are used than the ones we used in our study.

The ceiling effect, that we found in this study might be prevented by testing participants for a longer period of time, e.g. in various sessions. The idea is that the enthusiasm of performing the task will decrease in time. Also, the ceiling effect might be reduced by choosing other tasks which the participants dislike more, e.g. for the motivator task children could be asked to stop watching TV.

In future research we would therefore like to investigate the effectiveness of emotion expressions in other less interesting tasks. Furthermore, we will investigate the speech intelligibility of the vocal emotions expressions both objectively and subjectively, for various levels of emotion arousal.

## 3.5. ACKNOWLEDGEMENTS

We would like to thank all the children of “groep 6” from the Postiljon and their parents for their participation. The “emotion in speech” approach has been developed and tested in the BSIK MultimediaN project (www.multimedian.nl). Furthermore, the SuperAssist project of the IOP-MMI SenterNovem provided the application.

# 4 | Child's perception of robot's emotions:effects of platform, context and experience

**Abstract** *Social robots may comfort and support children who have to cope with chronic diseases like diabetes. In social interactions, it is important to be able to express recognizable emotions. Studies show that the iCat robot, with its humanoid facial features, has this capability. In this paper we look if a Nao robot, without humanoid facial features, but with a body and colored eyes is also able to express recognizable emotions. We compare the recognition rates of the emotions between the Nao and the iCat. First a set of bodily expressions of the Nao for five basic emotions (angry, fear, happy, sad, surprise) was created and evaluated. With a signal detection task, the best recognizable bodily expression for each emotion was chosen for the final set. Then, fourteen children between 8 and 9 years old interacted both with the Nao and iCat to recognize the emotions within context, in a story-telling session, and without context. These interactions were repeated one week later to study the learning effect. For both robots, recognition rates for the expressions were relatively high (between 68% and 99% accuracy). Only for the emotional state of sadness, the recognition was significantly higher for the iCat (95%) than for the Nao (68%). The emotions shown within context had higher recognition rates than those without context and during the second interaction the emotion recognition was also significantly higher than during the first session for both robots. To conclude: we succeeded to design a set of well-recognized dynamic emotional expressions for a robot platform, the Nao, without facial features. These expressions were better recognized when placed in a context, and when shown a week later. This set provides useful ingredients of social robot dialogs with children.*

## 4.1. Introduction

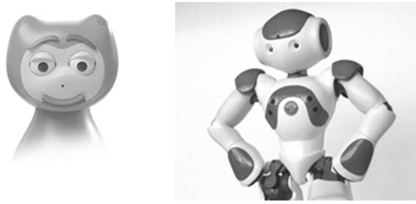
With computers and robots stepping out of their industrial environment and into the human society, they can be of surprising help in healthcare. Relatively "non-interactive" robots can provide substantial support for surgical, rehabilitation, and medication delivery purposes, whereas "highly interactive" robots can provide more cognitive, affective and social support [174]. Our research focuses on the latter: the development of socially assistive robots and investigating their potential to comfort and support children who have to cope with a chronic disease like diabetes [152]. This research is part of the ALIZ-e project that focuses on the development and evaluation of new social robot roles and personalized dialogs for such children (aliz-e.eu). In previous research emotion recognition of a robot with facial expression capabilities, the iCat (figure 9.3a), was evaluated while in the ALIZ-e project the Nao platform is used (figure 8.1), which does have minimal facial expression capabilities. Therefore there is a need to compare the emotional expressivity of the Nao's (mostly) bodily behaviors and the facial emotional expressions of the iCat. Furthermore, the Nao is a commercial available and widely used platform within the social robotics community, so knowledge on it is expressive behavior is of interest for a larger community.

In previous research, Looije, Neerincx and de Lange assessed the extent to which a robot with an expressive face, the iCat (figure 9.3a), could employ these cues to comfort and support children (i.e., as an educator, motivator and buddy) [146]. The iCat is a robotic research platform created by Philips[1]. With movable eyes, eyelids, eyebrows and lips, the iCat has the ability to show facial expressions and thus show emotions. The six basic emotions from Ekman and Friesen [82] were programmed into the iCat and validated by Kessens, Neerincx, Looije, Kroes & Bloothooft [127]. Several studies have been performed to test the recognition of emotions expressed by the iCat. In [11], for instance, the recognition rate was compared between a virtual and a physical iCat, using participants aged between 16 and 57 years old. For happiness, anger, disgust, and surprise, the recognition rate was above 70%, while it was around 30% for fear. No differences were found between embodied or virtual emotion expression. In [16] the virtual iCat was used to compare the recognition of the emotions between young (18-27) and older adults (65-75 year). The results show that, in accordance with research on human emotion expression, the ability of recognizing emotions decreases for older adults. Fear and anger especially decreased in the recognition performance, respectively from 60% to 30% and from 60% to 20%. In Kessens et al. [127] it was tested if children (8-9) were able to recognize emotions of the embodied iCat (movie clips). The results showed that the children had trouble, as the older adults, in recognizing certain emotions (most noticeable "happiness" (only around the 35%)). The recognition increased after adaptation of the emotions.

The Nao robot (version V3+) (figure 8.1) is a state-of-the-art mobile humanoid robot created by Aldebaran-Robotics. Contrary to the iCat, it does not have moveable facial features, but it has the ability to alter its body posture due to electric motors and actuators that create 25 degrees of freedom. The Nao also has tactical sensors, two build-in cameras, speakers and colored eye-leds. The moving feature and the speakers and colored leds were used for the experiments described in this paper[2]. Moving body parts create the ability to show affective bodily expressions. Before the Nao could be used for this research, a set of dynamic emotional expressions needed to be created. Static emotions of another Nao robot were previously assessed both for adults [15] and for children [14]. In these experiments the adults confused happiness and excitement most often while pride, fear and sadness were recognized very well. The children also had the lowest recognition rate for happiness, but higher than adults, at 83%. Fear and sadness were recognized 92% of the time, and pride even 100%. The Nao's ability to lighten up its eyes with different colors can help to support the emotions being expressed [6]. The colors that will be used to support emotional expressions in this research, are based on the investigation of Kaya and Epps [122].

This paper will assess and compare how well children can recognize the emotions of

[1]http://www.hitech-projects.com/icat/platform.php
[2]http://www.aldebaran-robotics.com/en



(a) (b)

Figure 4.1: Smiling iCat (a) and NAO (b)

the iCat's moving face and the Nao's moving body in order to establish a social human-robot interaction. The main question of this study is whether children can recognize robot emotions just as well in a robot that does not have facial features, as in a robot that does have facial features. If this is the case, it implies that facial features are not always required to establish affective robot dialogue acts.



Two experiments are described. In the first experiment, a set of emotional expressions for the Nao were created and validated based on affective human body postures from Bianchi-Berthouze and Kleinsmith [22], De Silva and Bianchi-Berthouze [71] and Coulson [59]. The second experiment assessed and compared children's emotion recognition for the iCat and Nao. For the iCat, validated facial expressions [127] were used and for the Nao, the bodily expressions created in the first experiment were used. The assessment consists of the following research questions:

1. Are emotions recognized more accurately for the iCat's already validated facial expressions, or for the newly created Nao bodily expressions? Because people mainly focus on facial expression while interacting with other humans [52], we hypothesized that the iCat will produce higher recognition rates.
2. Does the recognition rate improve when the emotional expression is presented in a corresponding context? We hypothesized that addition of a corresponding context to the robots emotional expression, will improve recognition rates [7].
3. Does recognition improve with repeated exposure? Emotional attachment seems to help asses emotional expressions [116, 233] and emotional attachment to a device is, according to Norman and Ortony[184], the sum of emotional experiences. In this study, the recognition rates of the second interaction are therefore hypothesized to be higher than those of the first interaction.

## 4.2. BACKGROUND

Socially assistive robots will ultimately provide engaging and motivating personalized therapy for participants and patients. For this purpose, social robots will need human-oriented interaction skills [237]. Reeves and Nass [201] noted that humans, regardless of their age and knowledge about technology, often project social qualities to the behavior of technology. People apply social interaction models when they interact or observe an autonomous robot, in order to improve the communication and to better understand the robot [39]. Conveying to people that a computed product has social presence, and thus has social influence, can be done by adding one of five social cues described by Fogg

[90]. These social cues are: *physical* (face, eyes, body movement), *psychological* (personality, feelings, empathy), *language* (speech, language recognition), *social dynamics* ((cultural) patterns used in social interaction) and *social role* cues (doctor, teammate, opponent, teacher, pet).

While in the past some theorists used to argue that emotions have no useful function, or even disrupt rationality, reason, and other cognitive processes (e.g. Hebb, 1949; Mandler, 1984, as cited by Keltner, and Gross [124]), more recently, theorists have come to believe that emotions optimize individual adjustments to social environments [124]. The theorists that do agree that emotions serve important functions, debate about the definition and bases of emotions [124]. In [125], theories of the functions of emotions are discussed. They divide the functions into the *individual*, *dyadic* (between two people) and *group* levels. The current article focuses on the dyadic level of interaction between a child and a robot. Expressing emotions at this level helps to understand the other person's emotions, beliefs and intentions [125]. It evokes complementary and reciprocal emotions in others and also serves as incentive or deterrent for the other party's social behavior.



### 4.2.1. VERBAL CUES AND NON-VERBAL CUES

In recognizing emotions in other human beings, the context of the emotion is important for correct perception. According to Barrett, Linquist and Gendron [7] emotional words that explain the context of the emotion, reduce the uncertainty that is inherent in most natural facial behaviors and allow quick and easy recognition and perception of the emotion. Apart from communicating emotions by means of verbal information, humans can express emotions in non-verbal ways to complement or emphasize a verbal message, or even substitute a spoken message. We identify several non-verbal cues to emotional content.

### 4.2.2. FACIAL EXPRESSION AND BODY POSTURE

One way to substitute a spoken message is by using *facial expressions* [52], like raising the eyebrows in disbelieve or surprise [3]. Facial expressions provide important information in social communication. It is also the manner of emotional expression that has been studied the most. Ekman and Friesen found six basic facial emotions that are recognized by people of all cultural backgrounds [82]. These emotions are happiness, sadness, anger, fear, surprise, and disgust.

In addition to the facial expression, Darwin (as sited in [68]) already identified *body posture* to be connected to emotions in his early work. This is thus a second important physiological cue to express the internal affective state of human beings [15]. Whether it is possible to recognize the six basic emotions from just body posture was investigated by [59]. Coulson showed participants different 3D models with different postures and concluded that anger, happiness and sadness were all recognized with high agreement rates. 95% for happiness and sadness, and 90% for anger. Fear and surprise were less frequently attributed to the same models with 71% agreement for surprise and 67% for fear. For disgust, the participants could not identify a posture and had only 43% agreement between the participants. Coulson [59] used wooden stick figures that lacked detail like fingers and foot position. Shaarani and Romano [224] used 3D models with more details, and paid more attention to the strength of the emotions being expressed. In general, the findings of their research showed that happiness had the highest percentage of recognition followed by anger. The most difficult emotion to recognize by body posture was again disgust. Bianchi-Berthouze and Kleinsmith [22] also created a sample of affective body postures. Twelve participants of different gender, race and age performed as actors. With a marked suit they recorded movements for anger, happiness and sadness. For each emotion, the frame with of the most extreme expression was used as stimulus. In research by De Silva and Bianchi-Berthouze [71], fear was also taken into account to expand the affective body posture database. The concordance between the actors and observers was highest for fear, followed by sadness, anger and happiness.

### 4.2.3. EXPERIENCE



Several studies indicate that a well-established emotional attachment, will benefit the recognition of emotional expression [233, 116]. To establish an emotional attachment, Norman [177] suggest that multiple exposures are needed. The emotional attachment is the sum of the emotional experiences. Experiment two is therefore repeated. Results will show if a learning effect is present, but might also indicate a possible emotional attachment.

### 4.2.4. DYNAMIC VERSUS STATIC EXPRESSIONS

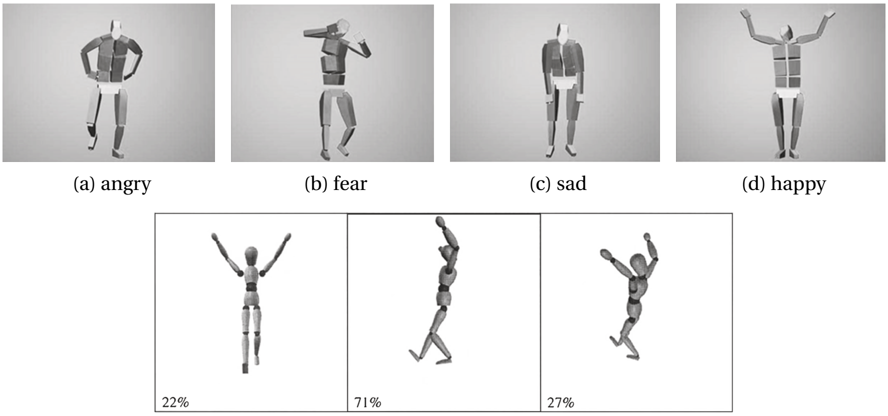
Most literature on facial expressions uses static images of different faces. When implementing facial expression into a robot, this will be shown in a dynamic state, considering that the robots are designed to move. The effects of static and dynamics on identifying emotions from faces were examined by Kätsyri and Sams [121]. For natural human faces there was no significant difference in identification between static or dynamic faces. Dynamics did increase identification in synthetic faces. Both results concern identification of identity and emotional expressions in faces.

### 4.2.5. COLOR

Kaya and Epps [122] investigated what colors evoke specific emotions among American college students. Green and blue were reported as relaxing, calming colors. Blue was also associated with sadness and depression. Red was either compared to love and romance or to negativity, evil, Satan and blood. White was considered a pure, simple and clean color. While most color-emotion relations are very personal, bright colors seem to be rated as positive and dark colors as more negative. Boyatzis and Varghese [37] confirm these findings for children, but found that besides personal experience, gender also plays a role in the association between color and mood. Boys are more likely than girls to respond positively to dark colors.

## 4.3. DESIGN OF BODILY EXPRESSIONS

The Nao has been used in emotional expression studies before [15, 14, 113]. These studies used key poses (static postures) of emotions different from those in this study. Therefore, new affective dynamic bodily expressions for the Nao had to be created and validated.




(a) angry (b) fear (c) sad (d) happy

(e) surprise (front, side and above)

Figure 4.2: Consulted affective postures; a, b, c and d from the database from [22] and [71], e is taken from [59]

### 4.3.1. CREATION OF AFFECTIVE DYNAMIC BODILY EXPRESSIONS

Anger, sadness, fear and happiness were based on the affective body posture database from Bianchi-Berthouze and Kleinsmith [22] and De Silva and Bianchi-Berthouze [71]. For expressions related to surprise, Coulson's research [59] was consulted. Because the literature did not provide an affective body posture for disgust, this emotion was left out of this study. Figure 4.2 shows the consulted database pictures.

First, these postures were used as key poses for the Nao. Due to Nao's limited degrees of freedom, not all movements could be precisely replicated from the pictures in Figure 4.2 and had to be slightly altered. For sad, 3 dynamic bodily expressions were composed and for surprise, only one dynamic bodily expression was programmed into the robot because it had limited degrees of freedom and could not show other postures that were found in the literature. For the other emotions, two expressions were created of which the best expression could be selected. The expressions were made dynamic by giving the robot a neutral starting point. From this neutral position the Nao would move to an affective pose and then back to the neutral position. The Nao has the ability to show different colors around the eyes and we wanted to include the feature in the bodily expressions for the Nao because ultimately, this feature will be used as well. Findings of Kaya and Epps and Boyatzis and Varghese showing relations between colors and human emotions were used in determining the colors for the different expressions [122, 37]. For angry, the leds around the eyes turned red. Sad was linked with dark blue eyes. Happy and surprise were both given light colored eyes, and for fear a yellow-orange color was used. Johnson, Cuijpers and van der Pol [113] also used the Nao's leds to imitate emotions and found similar results. Anger was associated with red, surprise and happiness with yellow, sad with blue, but for fear they found that black and gray were most.

### 4.3.2. EVALUATION OF AFFECTIVE DYNAMIC BODILY EXPRESSIONS

#### PARTICIPANTS

Eight participants, five men and three women, took part in this experiment. The expressions concerned a set of universal emotions that will ultimately be used for western European adults and children. This experiment therefore used western European participants (mean age was 24.6 and std = 2.87). Results from previous studies showed that recognition rates can differ between age groups[8, 9], but that all age groups are quite good in recognizing emotions.

#### PROCEDURE

To investigate which of the expressions was best recognized for an emotion, a signal-detection task [232] was created for this experiment. For every participant there were five trials that all focused on one of the created emotional expressions, which is the target for that trial. The Nao showed the target emotion among the other emotions and a neutral expression in every trial. If the robot showed the target emotion, participants had to score the movement as a 'signal/yes' and if the robot showed another emotion, the participants had to score the movement with 'noise/no'. Comments made by the participants were noted and used later to improve the expressions.

In every trial the signal that the participants needed to spot was one of the expressions. Within every trial, 12 expressions were shown; four targets (the two postures twice) and eight non-target expressions. For the sad condition, the first expression was shown as explained, but the second and third were only shown twice every other trial. These two expressions were the least convincing in the literature. For surprised, the one expression was shown four times per trial. In this way, all trials would contain the same amount and ratio of targets and non-targets.

Between trials, the order of the expressions was randomized. The order of the trials was different for half of the participant group. One half worked with the order; angry, fear, happy, sad and surprised, and the other half did a backward order; surprise, sad, happy, fear and angry.

### 4.3.3. RESULTS

Table 4.1 shows the results of the signal-detection task. H stands for the number of hits (were a signal was correctly seen as one), FA stands for false alarms (a noise was detected as a signal) and M stands for misses (a signal was not detected or seen as noise). The hit-rate (h) indicates the probability of responding yes to a signal stimuli. Hit-rates close to 1 indicate that participants make correct decisions. False alarm-rates (f) indicate the probability of responding yes to a noise stimuli. False alarm-rates close to 0 indicate few incorrect decisions. D' indicates the discriminability between the signals and the noise stimuli by calculating the distance between the mean of the noise and the mean of the signal in standard deviation units. The higher d', the better the discrimination between a target and a noise stimulus. D' could not be calculated for happy because there were no false-alarms. A' is another measurement for the discriminability between signal and noise and indicates the area under the 'receiver operating characteristic' (ROC) curve. A' always has a value between 0.50 (no difference between signal and noise) and 1 (perfect distinction between signal and noise) and is roughly calculated as follows; 1-0.25*[(1-h)/(1-f)+(f/h)].

| | H | FA | M | H | f | d' | A' |
|---|---|---|---|---|---|---|---|
| Angry1 | 12 | 1 | 4 | 0.75 | 0.01 | 2.92 | 0.93 |
| Angry2 | 14 | 1 | 2 | 0.87 | 0.01 | 3.39 | 0.96 |
| Fear1 | 14 | 3 | 2 | 0.87 | 0.04 | 2.93 | 0.96 |
| Fear2 | 11 | 3 | 5 | 0.69 | 0.04 | 2.27 | 0.91 |
| Happy1 | 14 | 0 | 2 | 0.87 | 0 | - | 0.97 |
| Happy2 | 12 | 0 | 4 | 0.72 | 0 | - | 0.94 |
| Sad1 | 11 | 1 | 5 | 0.69 | 0.01 | 2.73 | 0.92 |
| Sad2 | 6 | 1 | 2 | 0.75 | 0.01 | 2.98 | 0.93 |
| Sad3 | 4 | 1 | 4 | 0.50 | 0.01 | 2.31 | 0.87 |
| Surprise | 22 | 3 | 10 | 0.69 | 0.05 | 2.16 | 0.9 |



Table 4.1: Hits, false alarms, misses, hit rate, false alarm rate, d' and A' for the tested affective dynamic bodily expressions.

As can be seen in Table 4.1, the best expressions, with the highest d' or A' within one emotion, were chosen as the final expression for that emotion. For angry and sad, the second expression was chosen as the final expression. For fear and happiness the first version was chosen. For surprise, after the signal detection task, vocal feedback from the participants was used to alter and improve this expression.

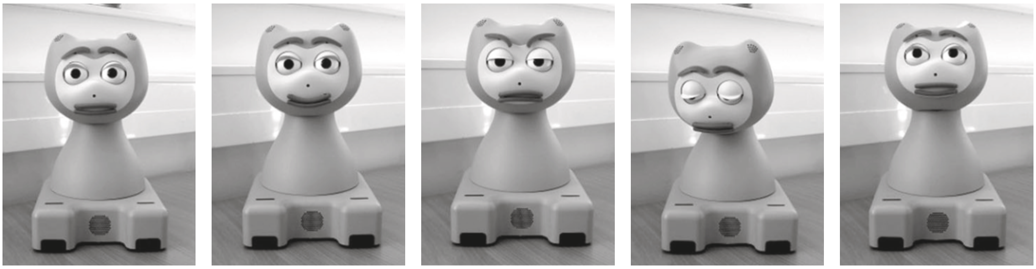
The vocal feedback pointed out that the upper part of the Nao had to be tilting slightly more backwards to improve the expression. Feedback from the participants might also explain why certain postures scored higher in recognition rates. The first angry expression was found to look more 'aggressive' than 'angry' and the second fear expression was found to be more 'repulsive' then scared.

By altering some of the movements of the Nao, five affective bodily expressions, that reflect five of the six basic emotions [82], were created. In the method section 4.4.1, the final expressions that came out of the first experiment are presented in photographs. The expressions all have one extreme moment, that could be described with the joint values of that moment. In the appendix, Table 4.4 is included that shows the parameters of the maximum joint values and the values for the RGB, hue and saturation of the final expressions. For more specific information about the expressions, one of the authors can be contacted or videos of the chosen movements can be seen on the internet[3]

## 4.4. COMPARING BODILY AND FACIAL EXPRESSIONS

In the second experiment the created and evaluated affective bodily expressions for the Nao were compared with the facial expressions of the iCat on recognition accuracy of children.

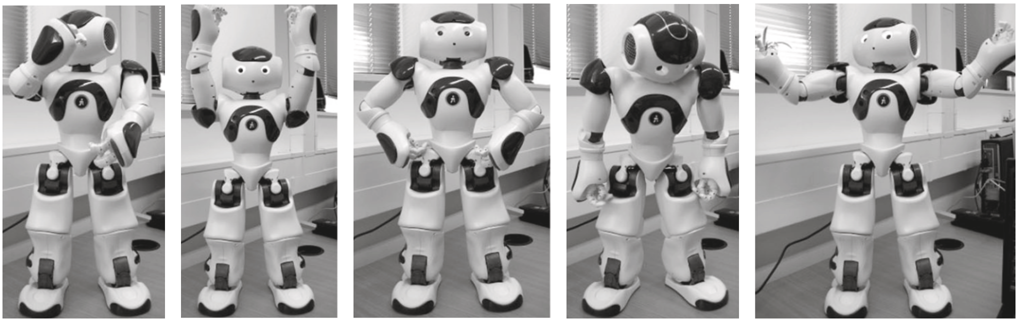
[3] http://mmi.tudelft.nl/SocioCognitiveRobotics/index.php/SocioCognitiveRobotics



(a) scared (b) happy (c) angry (d) sad (e) surprised

Figure 4.3: iCat emotions





(a) scared (b) happy (c) angry (d) sad (e) surprised

Figure 4.4: NAO emotions

### 4.4.1. METHODS

#### PARTICIPANTS

In this study, 14 children within the age of 8-9 participated (mean age=8.64 std=0.49). Five boys and nine girls were recruited from the fifth grade of the elementary school OBS de Watersnip in Zoetermeer, the Netherlands. 13 of them were native Dutch speakers. The non-native Dutch participant was asked if she understood everything that was mentioned and she did.

#### ICAT

The iCat research platform has an onboard processor that can be controlled by a laptop using a USB connection. The iCat was already programmed to show the six basic emotions for previous research (Figure 4.3). The voice used for the iCat consisted of wav files made in the speech synthesis program Fluency using the Dutch 'Diana' voice.

#### NAO

The Nao is a humanoid robot created by Aldebaran-Robotics. It can be connected with a laptop through WiFi or Ethernet. The Nao robot has speech synthesis in English and

French. In order to use a Dutch voice, wav files were created with Fluency. Because it might be confusing for the children to interact with two different robots with the same voice, 'Grietje' another Fluency voice was chosen for the Nao. Emotional perception can be affected by visual and audio presentation [165] which means that the emotions in the voice, will affect the overall perceived emotions of the robot. To keep this effect under control, the same program and level of affect was used to create both computer-simulated voices for the robots. The Nao's bodily expressions and eye colors were created and validated in experiment 1 and shown in Figure 4.4.

### Wizard of Oz



A Wizard of Oz setup was being used for this study so the participants believed that they were communicating with an autonomous system while in fact the experimenter operated the system. The structure of the dialog was scripted with XML and the robots were controlled by the experimenter.

### Experimental design

This research was conducted as a within-subject design. The dependent variable; the recognition rate of the emotional expressions, was tested within three different independent variables. The first one was expression. The expression of emotions will be done through facial expressions with the iCat and through bodily expression with the Nao robot.

The second independent variable, context, had two levels (within and without). In the 'within context' condition, a story was told by the computer, with a different voice than the voices used by the robots. After one or two sentences spoken by the computer, one of the robots said something in line with the story and then displayed the appropriate emotion. For example; the computer said that a big dog was standing in the yard. The iCat would then continue the story by saying: "I absolutely do not dare to go outside anymore: there is a big dog in the yard!", followed by expressing the appropriate emotion, in this case fear. In the 'without context' condition, the robots simply expressed an emotion after saying "the next emotion will be shown". In all conditions, the five emotions were shown twice. The order of the emotions in the different conditions did not differ between the participants. Without context had the same order of emotions for both robots but varied in the 'within context' condition because the stories told by the robots were different.

The third dependent variable was time. To investigate if emotion recognition can increase with multiple experiences, this experiment was conducted twice with a week in between. In the second session the stories told by the robots were swapped. Slight alterations in the stories had to be made, for example; were the iCat said "I do not feel so well. I hope I do not have to go to the vet.", the Nao said: "I do not feel so well. I hope I do not have to go to the robot-doctor."

### Procedure

The experiment started with an introduction of the robots in the classroom. After the children saw the robots, they filled in the introduction questionnaire, about their thoughts on the robots and how much they thought they would enjoy the experiment. Then, the children came to the experiment room (Figure 9.3) individually where the experiment

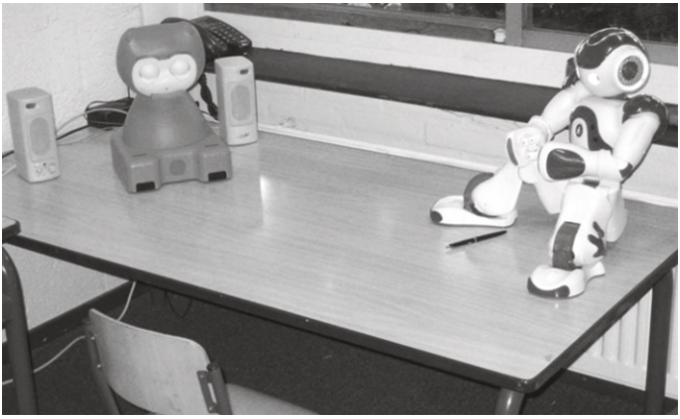


Figure 4.5: Experimental Setup

| Sequence nr. | Robot condition | | | |
|---|---|---|---|---|
| 1 | iCat – C | iCat - NC | Nao – C | Nao – NC |
| 2 | iCat – C | iCat – NC | Nao – NC | Nao – C |
| 3 | iCat – NC | iCat – C | Nao – C | Nao – NC |
| 4 | iCat – NC | iCat – C | Nao – NC | Nao – C |
| 5 | Nao – C | Nao – NC | iCat – C | iCat – NC |
| 6 | Nao – NC | Nao – C | iCat – C | iCat – NC |
| 7 | Nao – C | Nao – NC | iCat – NC | iCat – C |
| 8 | Nao – NC | Nao – C | iCat - NC | iCat – C |

Table 4.2: Sequences of robot use. C = context, NC = no context

started. To familiarize the children with the robots, a short personal introductory conversation was held, in which the robots asked their name and said that they looked forward to the experiment.

After the introduction, the child interacted twice with both robots as can be seen in Table 4.2. Participant 1 was assigned to sequence 1 (interaction with iCat within context, then without context, then interaction with the Nao within context and then without context), participant 2 was assigned to sequence 2, etcetera. Sixteen participants were anticipated but two dropped out before completing the experiment. With 14 participants, sequences 5 and 8 were only used once. The second week the children were assigned to the same sequence to investigate the learning effect of a second session.

On questionnaires the children filled in what emotions they thought the robots expressed during the different conditions. These questionnaires consisted of forced choice answers: 'happy', 'sad', 'anger', 'fear', 'surprise' and 'no emotion'. After the tests, another questionnaire about their impression and opinion about the robots was completed in the classroom. This questionnaire consisted of questions about whether or not they liked the robot or if they thought the robots could be trusted.

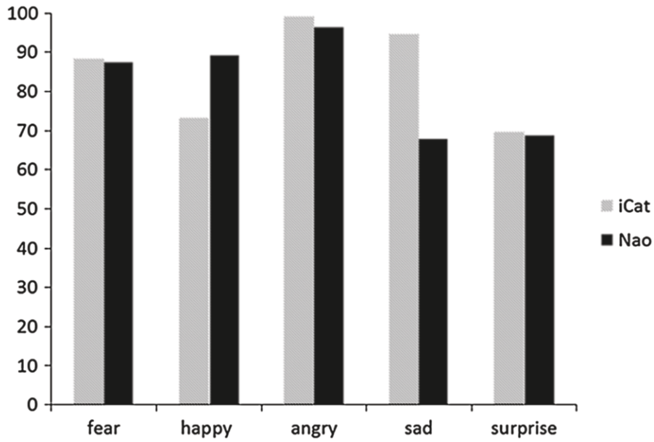



Figure 4.6: Percentages of correct recognitions for the iCat and for the Nao robot.

#### ANALYSES

This study used a within subject design. Emotional expression (facial vs. bodily expressions), presence of context and session number were the independent variables and the number of correct recognitions were the dependent variables. The data were analyzed with T-tests and ANOVA's.

### 4.4.2. RESULTS

First, the correct recognitions per emotion and per robot were compared. Figure 4.6 makes this easier by showing the correct recognition rates for all five emotions graphically. The light bars represent the iCat, and the striped, dark bars represent the Nao. Happy is the only emotion that has more correct recognitions for the Nao than for the iCat, however, an ANOVA shows this is not significant ($F(1, 1118) = 2{,}1259$, $p=0.15$). For sad, an ANOVA shows that the recognition rates for the iCat (0.95) are significantly higher than for the Nao (0.7), ($F(1,1118) = 5.93$, $p=0.02$). This is the only significant difference in recognition rates of emotional expressions between the two robot platforms.

The correct recognitions per emotion and per robot were also analyzed to see if this number differentiated significantly from the number of incorrect recognitions. For both robots, there were significantly more correct recognitions than incorrect recognitions. This means that the Nao movements (in combination with certain eye colors) used in this experiment consisted of a good set of emotions that could be recognized by children.

Next, a confusion matrix was made to provide further information about the incorrect recognitions. In Table 4.3 this confusion matrix is shown, in which the rows indicate what emotion was shown by the robots and the columns indicate which emotion the participants recognized. Inside the cells the percentage of correct (bold) and false recognition is presented. This table also shows that certain emotions were confused with other emotions. Surprised, is often recognized as 'no emotion' for the iCat (17.86%) and for the Nao (19.64%) (bold). This expression is for both robots confused with happy. The sad expression is least recognized for the Nao, and the happy expression for the iCat might be

| | Fear | | Happy | | Angry | | Sad | | Surprised | | No emotion | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **iCat** | **Nao** | **iCat** | **Nao** | **iCat** | **Nao** | **iCat** | **Nao** | **iCat** | **Nao** | **iCat** | **Nao** |
| Fear | *88.39* | *87.5* | 0.89 | 0.89 | 0 | 0.89 | 1.79 | 6.25 | 4.46 | 1.78 | 4.46 | 2.68 |
| Happy | 6.25 | 0 | *73.21* | *89.28* | 0.89 | 0 | 5.36 | 0 | 5.36 | 6.25 | 8.04 | 4.46 |
| Angry | 0 | 1.78 | 0 | 0.89 | *99.11* | *96.43* | 0 | 0 | 0 | 0 | 0.89 | 0.89 |
| Sad | 0.89 | 3.57 | 0 | 5.36 | 0 | 8.04 | *94.64* | *67.86* | 2.68 | 5.36 | 1.79 | 9.82 |
| Surprised | 1.78 | 0.89 | 7.14 | 5.36 | 0.89 | 0.89 | 1.78 | 4.46 | *69.64* | *68.75* | *17.86* | *19.64* |

Table 4.3: Percentage of emotion recognitions per emotion for each robot. Rows indicate the shown emotion and columns the recognized emotion.

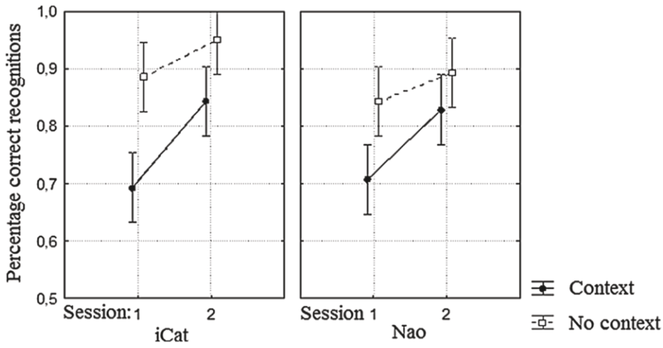



Figure 4.7: Factorial ANOVA with three variables: robot, session and context.

low in recognition because it is too similar with the neutral expression. For both robots the anger expression is the best recognizable.

In previous research by Kessens et al. [127], the iCat had the highest recognition for angry followed by sad and happy and fear with the lowest recognition rate. In this experiment the iCat also had the highest recognition for angry followed by sad, but fear was better recognized in this study followed by happy and surprise. The consulted literature for the bodily expressions did not agree among each other. In this experiment, the recognition rates for the dynamic expressions are, from high to low; angry, happy, fear, surprised and sad. For De Silva and Bianchi-Berthouze's [71] static emotional postures, sad had the highest recognition rate (88%) followed by fear (68%), happy (65%) and angry (55%). And for Coulson [59] the static posture for surprise had a recognition of 71% which comes close to the recognition rate of 68.75% for the Nao's dynamic expression.

The main research questions focused on (1) the difference in emotion recognition for the two robots, (2) the effects of context and (3) learning effects. First, an ANOVA showed that there was no overall significant difference of recognition accuracy between the iCat and Nao emotions $F(1, 1118)=1.24$, $p=0.27$. Second, emotions expressed in context were significantly better recognized than emotions expressed without context $F(1, 1118) = 29.79$, $p = .00$. Third, in the second session, emotions were better recognized than in the first session ($F(1, 1118) = 18.76$, $p = .00$). A factorial ANOVA showed no interaction effects found between the variables (Figure 4.7).

The children's opinions about the robots were collected with the last questionnaire.

The first question “how was working with this robot” could be answered with a 5-point scale, with 5 ‘great’ and 1 ‘not great at all’. After the first experiments the iCat scored a 4.5 and the Nao a very high 4.9. After the second experiment the robots both dropped to 4.4 which still is between ‘very nice’ and ‘great’. On the question what robot they would like to use again, the children answered either Nao or both robots. No one was specifically interested in the iCat. The questions about how friendly the robot was, and if the robot was trusted, the iCat first scored between 4.2 and 4.7 and the second time between 4.4 and 4.7. For the Nao these numbers were the first time between 4.1 and 5 and the second time between 4.4 and 4.7.

### 4.4.3. Conclusion



The first hypothesis stated that five of the six basic emotions [82] would be easier to recognize correctly by children, when expressed through facial expressions from the iCat, compared with bodily postures from the Nao, because the face is the main feature humans use for expressing emotions [52]. Only for the sad emotion this hypothesis was true. The facial expression from the iCat had significantly higher recognition rates than the bodily expressions from the Nao. But when looking at the entire set of emotional expressions, no difference was found between facial and bodily expression. This finding shows that both robot platforms are able to express the basic emotions [82] in a way that children can recognize them, and that facial features are not crucial to express emotions in a robot. It might be that humans are able to recognize emotions just as well in bodily expressions as in facial expression, but for now this research only shows that there is no difference in recognition rates of emotional expressions shown in the robot platforms used.

Barrett, Lindquist and Gendron [7] stated that emotional context will strengthen the certainty of the emotion recognition. Therefore it was expected that emotions in this experiment will be better recognized when they are being expressed in the context condition. The conducted ANOVA shows that this is indeed the case. Thus, when one of the robots used in this study expressed an emotion when it is telling a story, this emotion is more likely to be recognized correctly.

The last and final hypothesis stated that in the second session the correct recognition rates would be higher than in the first session. The results show that this is indeed the case for both platforms, and for both conditions. The second session, all conditions show higher recognition rates than in the first session.

## 4.5. General Discussion

In this study two experiments were conducted. First, five affective bodily expressions based on five of the six basic emotions, were successfully created for a humanoid robot that has no facial features, the Nao. In the second experiment, 8 and 9 year old children interacted twice with two different social robots, the iCat that can show facial expressions and the Nao that can show bodily expressions. By comparing the emotional recognition rates between these robots, we could asses if facial features are necessary for a robot to show recognizable emotional expressions. It turns out that this is not the case. The bodily expressions were recognized just as well as the facial expressions. Effects of context and repeated interactions were also tested. Furthermore, the children's opinions about the robots were assessed. Almost all questions were answered very positively for both robots, but if the children could choose between the two robots they would go for the Nao, or both robots but not for the iCat. Where Kessens et al. [127] discovered that children were very excited about the iCat, this experiment showed that the children were even more excited about interacting with the Nao.

Complementary analysis methods were applied to design and test emotional behaviors, aiming at a sound empirical-founded set of discriminable robot expressions. In the first experiment, the signal detection theory was used to select the best alternative of a limited design set [232]. This method is rather common practice in the medical field (e.g., [158]), but less apparent in the field of Human-Robot Interaction. The second experiment used more common methods for this research field to compare robot behaviors for two platforms, i.e., a confusion matrix and an ANOVA [100].



One thing that could have contributed to the high recognition rates for the Nao's expressions, is the use of eye color. Appropriate colors were implemented in the emotional expressions for the Nao, because the project will use both features (movement and eye color) in future studies. Adding color to the Nao's posture will make the postures less generic for similar robots but the main concern for this paper is to make recognizable emotional expressions for the Nao using its features; movement and colored led lights. There were no conditions in which the Nao had no eye color. Therefore, it could not be tested if the colored leds made the emotions more easy to recognize. A similar study [83] also created emotional postures for the Nao without using the led colors, and found more confusion between some emotions than revealed in this study, but this could also be an effect of using different age groups as participants. It would be interesting to investigate if the emotional expressions created in the current study are less recognizable by children, without the use of colors or with the use of incongruent colors.

This also applies for the use of the voices. Different voices were used for the robots to exclude a possible confusion of the participants when both robots would speak to them in the same voice. Although the voices used for the robots were closely related to each other (high pitch, female voices), they were not the same. The emotional expression in the voices was kept as equal as possible, but in this research we did not study the effect of emotional voices on the recognition rates of the non-verbal emotions. In further research we would have to test the effects of using different voices, and using different levels of emotional speech for the same robots and same emotional expressions to test if the emotions were recognized based on the expressions or based on affect in speech.

This study focused on some basic emotions that were distinguished by Ekman and Friesen [82]. Such emotions can be projected on a dimensional emotion space, for example with the dimensions of arousal (level of activity) and valence (level of (dis)pleasure) (e.g., Truong, van Leeuwen and Neerincx [245]). In the future, we will enrich Nao's expressiveness by incorporating such a dimensional model which can display emotions on different intensity levels (Beck, Cañamero and Bard [13]; Beck, Hiolle, Mazel, and Cañamero [15]).

Our future end-users, children with diabetes, will interact with the robot repeatedly. Therefore, it is interesting to investigate how the attachment between a child and the Nao can be established, such as the attachment to a robot dog [261]. This research by

Weis, Wurhofer and Tscheligi [261] used the view from Norman and Ortony [177] that emotional attachment is the sum of emotional episodes of user experiences with a device. The children that interacted with a robot dog rated positive on all three levels of user experiences [177], which shows that it is easy for them to create an emotional attachment with a robot dog. Tanaka, Cicourel and Movellan did a 5 months longitudinal study to the interaction between toddlers and a humanoid robot [235]. At the end of the study the toddlers interacted with the robot not like a toy, but as if it was a peer. Attachment can help to assess emotions correctly, as a core feature of emotional intelligence [233, 116]. When recognition rates increase over time, not only a learning effect might be present, but possibly also an attachment between the observer and the robot expressing the emotions. It should be noted that presence of attachment is not investigated in this experiment, but this issue might be beneficial for further long-term human-robot interaction studies.



This study aimed at creating a set of emotional bodily expressions for the Nao, in order to use the Nao in the ALIZ-e project. The, already validated, emotional facial expression of the iCat could not be embedded in the Nao, because the Nao has no moving facial features. This study succeeded at creating a set of emotional expressions for the Nao, that are well recognized by children of the age 8-9, showing that a robot platform does not necessarily need facial features to express recognizable emotions.

## 4.6. ACKNOWLEDGEMENTS

This work is (partially) funded by the EU FP7 ALIZ-E project (grant number 248116). Thanks to Stella Donker and Linda van Ooijen from Utrecht University. Thanks to OBS de Watersnip from Zoetermeer, The Netherlands, and the children and parents for participating. Thanks to Andrea Kleinsmith and Nadia Bianchi-Berthouze for access to their database of affective postures. A. Kleinsmith, R. De Silva, N. Bianchi-Berthouze, "Cross-Cultural Differences in Recognizing Affect from Body Posture", Interacting with Computers, 18 (6), (2006) 1371-1389. Thanks to Mark Coulson for his pictures of affective body postures from his research. And thanks to Bert Bierman, Stella Donker and Linda van Ooijen for their contributions.

## 4.7. APPENDIX

| Joint | Happy | Angry | Scared | Sad | Surprised |
|---|---|---|---|---|---|
| LshoulderRoll | 9,226 | 35,330 | 1,668 | 13,885 | 37,176 |
| LShoulderPitch | -62,054 | 79,100 | 37,879 | 91,229 | 8,523 |
| RShoulderRoll | -8,176 | -45,882 | -1,672 | -8,704 | -47,903 |
| RShoulderPitch | -63,983 | 84,642 | 6,594 | 84,203 | 25,931 |
| LElbowYaw | -71,195 | -7,737 | -24,964 | -58,978 | -119,496 |
| LElbowRoll | -36,736 | 81,649 | -39,197 | -24,959 | -55,369 |
| RElbowYaw | 55,809 | 6,589 | 50,535 | 52,733 | 119,496 |
| RElbowRoll | 29,007 | 86,048 | 88,334 | 5,979 | 49,310 |
| Lhand | 0,272 | 0,265 | 0,284 | 0,268 | 0,272 |
| LWristYaw | -43,685 | -42,986 | -40,521 | -42,718 | -41,751 |
| RHand | 0,232 | 0,216 | 0,243 | 0,225 | 0,232 |
| RWristYaw | 60,379 | 61,522 | 60,291 | 61,873 | 58,973 |
| HeadYaw | - | 21,355 | -38,059 | 24,871 | -24,261 |
| HeadPitch | - | -15,384 | 1,843 | 28,035 | -15,823 |
| LHipYawPitch | - | - | - | -37,051 | -10,544 |
| LHipRoll | - | - | - | - | 12,809 |
| RHipRoll | - | - | - | - | -4,459 |
| R | 255 | 209 | 255 | 6 | 255 |
| G | 170 | 0 | 170 | 45 | 255 |
| B | 255 | 3 | 0 | 117 | 127 |
| Hue | 300 | 359 | 40 | 218 | 60 |
| Sat | 85 | 255 | 255 | 242 | 128 |
| Val | 255 | 209 | 255 | 117 | 255 |

Table 4.4: Parameters of the joint values during maximum emotional expression of the Nao's emotional movements. The movements start in a neutral position, go to the maximum expression and go back to the neutral position . The bottom 6 rows indicate the parameters used to light up and color the eye leds for every emotional bodily expression. Video's for the full robots movements can be viewed at the following link: http://mmi.tudelft.nl/SocioCognitiveRobotics/index.php/SocioCognitiveRobotics

# 5 | Help, I need some body: The effect of embodiment on playful learning

**Abstract**

*Children with a chronic disease like diabetes need to learn how to self manage their disease. Knowledge about their condition is indispensable to reach this goal. Within the European project ALIZ-E a robot companion is being developed that should, among others attributes, have the capability to educate children. In this paper, a virtual agent on a screen is compared with a physical robot on the aspects of performance (learning), attention and motivation. The experiment consisted of two sessions in which children played a quiz consisting of health related questions with both the robot and the virtual agent, there was a week between the two sessions. It was found that performance and motivation were not affected by the embodiment, but the robot did attract more attention and, when forced to choose, the children had a preference for the robot.*

## 5.1. Introduction

As the number of people with a chronic disease grows, an important subject in their medical care is the promotion of self-efficacy and self-management. Effective selfmanagement allows patients to spend less time in the hospital and enables them to have a higher quality of life. To be able to manage their disease, patients need to learn about their disease and the skills required to manage it. This is true for both adults and children. Children with a chronic, lifestyle related disease can benefit from extra support in their daily activities. To provide this support, the ALIZ-E project (www.aliz-e.org) aims to develop a social robot for these children, suitable for longterm social interaction. One of the goals of the project is to research how to use a robot to educate children on both their disease and a general healthy lifestyle, thereby increasing their self-efficiency. The overall scenario of the project is based on a medical setting, where children diagnosed with diabetes spend a complete week in the hospital after diagnosis, so that they can learn about managing their illness and its implications. The robot will not only function as an educator, but also as a friend and motivator. This paper describes an experiment which focuses on the role of the robot as an educative companion and compares a virtual embodied agent (from here on: agent) and a robot (see Figure 5.1) In order to make this comparison , we created a scenario in which a quiz, based on Trivial Pursuit®, with health related questions is played. Since we mainly want to study the effects of interaction on learning in children, we need not burden sick children, so we performed this experiment with healthy children, under the presumption that the results are generalizable to children with diabetes.

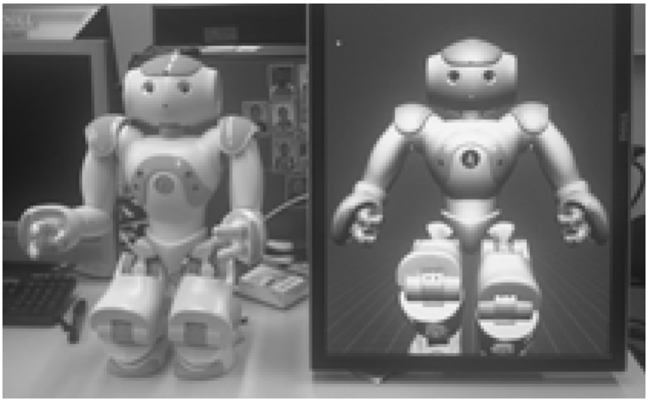


Figure 5.1: On the left the robot and on the right the lifesize agent



## 5.2. LEARNING, ROBOTS AND AGENTS

### 5.2.1. LEARNING

Learning can be defined as "persisting change in performance or performance potential that results from experience and interaction with the world" [79]. To achieve change there is a need for interaction so that information can be processed and stored. This interaction can be achieved in multiple ways, for example face-to-face with a human teacher, but also via a computer interface, virtual agent or robot. When there is a need for many repetitions, this is hard to ask from a human educator, but this is not a problem for an electronic one. Another advantage of electronic educators is their capability to keep track of a user's progress and adapting to it. Given these advantages, an electronic educator can be especially beneficial in situations where it is difficult for a human educator to tailor education to one individual (in a classroom setting), when a human teacher simply isn't available or has little time.

To achieve the above-mentioned change in performance, several means to enhance the interaction are used. Two vital aspects in these means are motivation and attention; both of them have been shown to improve the retention of educated materials [123, 164]. To improve learning, both human and electronic educators strive to hold attention and increase motivation. Games have been shown to have an added value over standard e-learning methods, because of the fun they provide. Balancing fun and educational value ensures a high motivation and longer interaction (attention) [196, 105].

### 5.2.2. AGENTS AND ROBOTS

Next to games, virtual agents are used to attract attention and to keep motivation in an e-learning environment. Studies show that the presence of an agent enhances memory performance [21], self-efficacy and enjoyment [112]. The available studies in which a robot has the role of an educative companion [236, 162], show that the use of a robot in role can be effective, but that there are still many questions on how and when a robot has an added value.

### 5.2.3. EMBODIMENT

Several studies have looked into the differences between agent and robot (e.g. [146, 147, 11, 257, 194]). Embodiment (here: having a body in the real world) is the main difference between an agent and a robot. Literature names several factors that are influenced by embodiment, either virtual or physical, that could cause an effect in learning (e.g. attention [76], attitude [133], trust [168]). Results from studies comparing agents and robots are inconclusive. They are often performed with agents and robots having a different appearance [194], [226] and the evaluation is usually done with adults as subjects, for one session only. We will use a robot and an agent with the same appearance and use our target group (children) to evaluate the embodiment effect on learning and two of its contributing factors, motivation and attention, during two sessions. Given the embodiment of the robot in the real world, we hypothesize that the robot will get more attention and motivates children more, resulting in better learning. This leads to the following individual hypotheses:

1. Learning: Children learn more while interacting with a robot than while interacting with a virtual agent
2. Attention: A robot keeps children's attention longer/better than a virtual agent
3. Motivation: Children are more motivated to learn when interacting with a robot than with a virtual agent In order to test these hypotheses a field experiment with a within-subject design of two sessions was conducted.

5

## 5.3. IMPLEMENTATION

### 5.3.1. QUIZ SCENARIO

Literature on educating children [254] provided requirements for the scenario for an educative companion role. The main requirements are active involvement, capability to address developmental and individual differences and for it to be entertaining for children, thereby making it intrinsically motivating. We decided to use a quiz scenario because a quiz adheres to all of these requirements and this scenario is feasible to implement.

We based the quiz that is used in this study on a two player Trivial Pursuit®: four topics were included *1 ("schijf van vijf"/food pyramid),2 (energy balance),3 (eating healthy),4 (the heart)*. We used multiple choice questions with four answer options (see Appendix for question examples). The child reads the questions of cards and the agent/robot reads them from "memory". Questions were taken and adapted from existing material or newly created. A teacher checked all the questions to make sure they were of the right level. We also included a scoring element, for this allows children to monitor their progress, and creates an element of friendly competition between the agent/robot (who was at the same knowledge level as the child) and the child, both of which are known to enhance motivation.We created a visual scoreboard and made it such that, if a child answered all questions fast (within 10 minutes) and correct it could complete the scoreboard completely. A topic is completed when three questions are answered correct (see Figure 5.2).

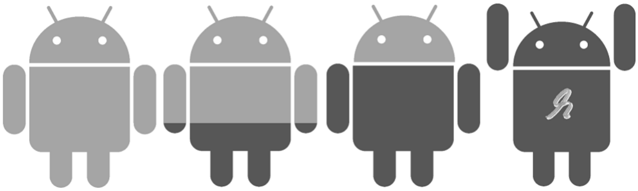


Figure 5.2: Scoring robot

### 5.3.2. GAME PLAY

The game-play is as follows: Child's turn:

1. Robot/agent asks the child to pick a topic.
2. If that topic is played for the first time, a video is shown to introduce the topic.
3. Robot/agent asks the a question on that topic and provides the four multiple choice options. These options are also displayed on a separate screen.
4. Two chances are given to answer the question correct.
5. If the correct answer isn't given, the robot explains the correct answer.
6. Scoreboard (displayed on separate screen) is updated .
7. The robot/agent makes a comment, praising the child or encouraging it to do better.

Robot's turn:

1. Robot/agent picks a topic.
2. Child reads the question and answer option from a card.
3. Robot/agent gets two chances to answer correctly.
4. If it fails, the robot/agent asks the child to explain the answer (explanation is also given on the card).
5. Scoreboard is updated, robot/agent comments on it.

### 5.3.3. ROBOT/AGENT INTERACTION IMPLEMENTATION

We developed a script containing both the dialogue options and gestures. The dialogue script contains the whole interaction of playing the quiz and takes into account the most common aspects that can go wrong (e.g. the child doesn't understand the robot/agent or the other way around). Gestures are both incorporated in the dialogue script (e.g. pointing in the right direction when in dialogue contains the phrase 'look there') and in idle movements. Looking at the user at the right moments was also implemented, combined with the gestures this should make the whole interaction more natural. The robot we used was the NAO[1]. To ensure that the agent (a virtual representation of the NAO) looked as similar as possible it was displayed on a 30-inch monitor. The speech of the robot and the agent was produced by the same text-to-speech editor.

[1] www.aldebaran-robotics.com

### 5.3.4. WIZARD OF OZ

The whole experiment was scripted and automated, except for the speech recognition, which was not yet available. We used a Wizard of Oz setting to replace this.

## 5.4. EXPERIMENT

### 5.4.1. PARTICIPANTS

A fifth grade of a Dutch primary school participated in the experiment. It was a class for children with a lag in, for example, the Dutch language. Therefore, the class was small (11 children). The class consisted of three girls and eight boys. One child was excluded afterwards because he changed schools during the experiment. The average age was 11.1. None of the children suffered from diabetes.

### 5.4.2. MEASURES

Below we present the metrics used for each of the three hypotheses.



**Hypothesis 1: Performance** The answers on the quiz questions provide an indication of performance, but because the participants might receive a different number of questions or use different learning strategies, we did not use this as a measure. Instead, we created a knowledge test consisting of random questions from the quiz, two from each subject. Per subject, one question was taken from the questions asked to the child, the other was picked from the questions asked to the robot/agent.

**Hypothesis 2: Attention** To measure attention we used subjective self-reported attention (self-reported by the child) and objective video analysis. In the video analysis we looked at where the child looked (robot/agent, quiz screen or elsewhere), how many times and how long (in total and per gaze) the child looked there.

**Hypothesis 3: Motivation** For measuring motivation we asked questions on sub-factors that influence motivation [103, 51, 66]; enjoyment, trust and preference. The questionnaires were made suitable for children, following guidelines that were derived from [154] and [17].

### 5.4.3. EXPERIMENT DESIGN AND PROCEDURE

Since the number of participants is small, we decided to do a within-subject design; this entails that all children interacted with both the robot and the agent. We randomized the order in which they interacted with the robot or the agent in both of the session. Each session consists of two phases of 10 minutes, with a short break between them. A phase consists of one full quiz-game against either the robot or the agent (including introduction and goodbyes). In the first phase, the child receives and asks questions on topic 1 and 2 and in the second phase on topics 3 and 4. This was done so that a possible primacy or recency effect would not bias the results. A schematic overview of this is provided in Figure 5.3. Figure 5.3 also shows the experimental procedure. All children started with a knowledge test, which serves as the baseline of their knowledge on the four topics in the quiz. It was administered on the same day they were introduced to

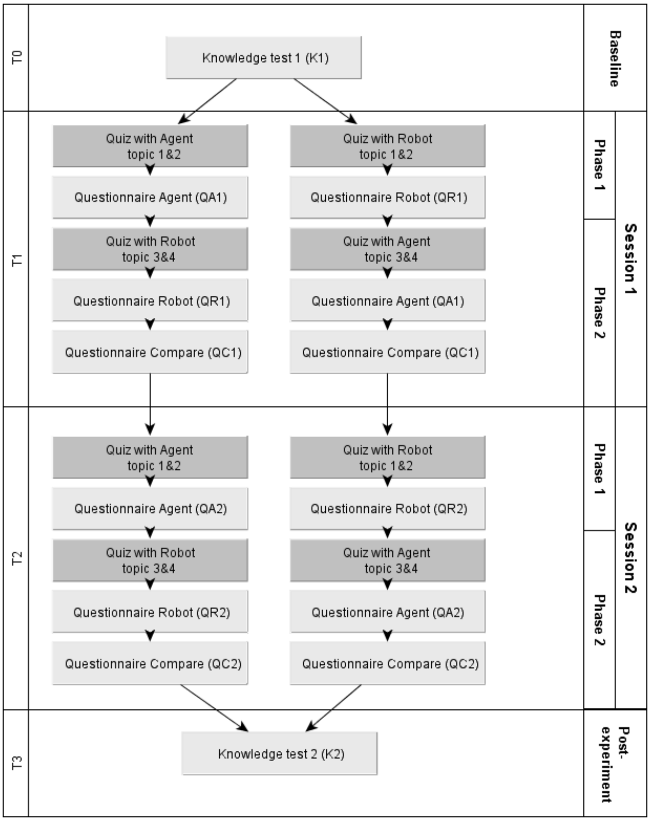



Figure 5.3: Experiment design

the agent and the robot. In the days after this all children interacted with the robot and agent for the first time. They played the quiz with either the robot or the agent and then answered questions about that interaction. They proceeded with a quiz with the other opponent. After this interaction they answered questions specific for the interaction and questions comparing the robot and agent. One week later the procedure was repeated and another week later the knowledge test from the start was repeated to measure their end level (see Appendix 5.8 for the knowledge test).

### SETUP

Figure 5.4 provides a schematic overview of the experimental setup. The children were recorded with two cameras (number 1 and 2 in Figure 5.4), camera 1 records the face of the child, camera 2 gives an overview of child, robot/agent and quiz screen. The cameras were explained to the children as being there 'to see whether the robot makes mistakes and if so, when, so that we can fix it'. The experimenter was located in another room, close by. To enable the experimenter to see what was going on in the other room and to act as a wizard of oz, a microphone and network camera (number 3) with an overview

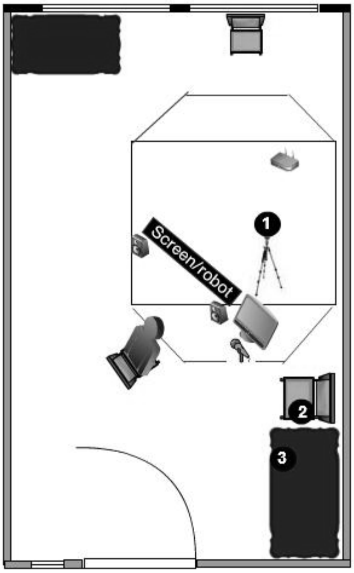



Figure 5.4: Experimental Setup



look were placed in the same room as the child. The children were told they could call the researcher using the microphone if they were scared or when something wasn't working. This explained the presence of the microphone and was reassuring for the children.

## 5.5. RESULTS

Eleven children participated in the baseline knowledge test. One child left school, so he did not participate in the interactions or the second knowledge test, he was excluded from all analyses. One child was sick during the second knowledge test, but his interaction results are part of the analysis.

### 5.5.1. HYPOTHESIS 1: PERFORMANCE

We excluded an outlier (severely distracted during second test) and a participant with missing data (not present during the second knowledge test) from the performance data, leaving us with an N of eight. On the first knowledge test, the participants answered respectively 5 questions out of 8 correctly, in the second knowledge test this was 5.6. This difference is not significant, p=0.25. We compared the progress on subjects that were played against the robot with the subjects that were played the agent, but no difference between the conditions was found. What could be concluded is that good performance on the quiz correlates with a good performance on the second knowledge test (r=0.855 and p <0.01). The two knowledge tests were not correlated (p=0.56), which shows that good performance on the quiz is a better indicator of their final score then the baseline test.

| | On robot/agent | On quiz-screen | Elsewhere |
|---|---|---|---|
| Agent (S1) | 39.4 | 36.8 | 29.5 |
| Robot (S1) | 45.9 | 39.0 | 29.3 |
| Agent (S2) | 38.3 | 41.0 | 28.1 |
| Robot (S2) | 42.5 | 33.8 | 29.3 |

Table 5.1: Number of times looking at measuring point (during 10 minutes).

| | On robot/agent | On quiz-screen | Elsewhere |
|---|---|---|---|
| Agent (S1) | 79.0 | 328.0 | 195.1 |
| Robot (S1) | 123.6 | 290.8 | 185.6 |
| Agent (S2) | 65.3 | 327.8 | 209.4 |
| Robot (S2) | 120.0 | 296.3 | 183.6 |

Table 5.2: Average duration of look (in seconds).



Furthermore, questions that were asked during the quiz were more likely to be answered correctly in the knowledge test. In the baseline test 65% of the questions that were not in the quiz were answered correctly and this was 67.5% in the second knowledge test. The questions that were in the quiz were answered correctly in 58% of the time before the quiz and 75% after the quiz.

### 5.5.2. HYPOTHESIS 2: ATTENTION

There were no differences in user reported attention; most children reported they found it easy to keep their attention directed to the robot/agent. Objectively, there were differences between robot and agent. Table 5.1 provides an overview of mean number of times per 10 minutes that a child looked at one of the measuring points (robot/agent, quiz screen, elsewhere) and Table 5.2 shows the average duration of looking at the measuring points. A two-way repeated measures MANOVA with as independent variables session and condition was performed. The effect of condition was significant ($p < 0.01$). Looking further into the differences between robot and agent condition it was found that the duration of looking at the opponent was significantly higher in the robot condition ($p < .001$). Furthermore, the children looked more often towards the robot, the difference with the agent condition was significant ($p = 0.05$). Since the total time of one phase is always 10 minutes, the time children looked at either the quiz-screen or elsewhere in the agent condition must be different. The results show that they look longer at the screen. Next to this we looked at correlations between attention and performance, we did find a negative correlation between number of times looking elsewhere and performance ($r = -0.828$ and $p < 0.05$).

### 5.5.3. HYPOTHESIS 3: MOTIVATION

No significant differences were found in the measures we defined as factors for motivation (enjoyment, trust and preference). Enjoyment and trust both scored almost 3 on a 3-point scale in both the first and the second session and for both the robot and the agent. Only in the forced choice preference there was a difference between robot and agent; 8 children chose the robot and 2 the agent.

## 5.6. Conclusion and Discussion

### 5.6.1. Hypothesis 1: Performance

We did not find a significant learning effect overall, and no significant effect in either condition separately. The hypotheses on performance can thus not be confirmed for this particular scenario.

### 5.6.2. Hypothesis 2: Attention

The children did look significantly more often at the robot. Next to this they did look significantly longer at the robot than at the agent. The robot was thus better in retaining attention and did not draw attention from the quiz screen. They did look a shorter amount of time at the quiz screen and at other things in the room when interacting with the robot, but because the number of switches between items was not significantly different we conclude that the robot was not distracting. It is interesting to note that the number of times looking elsewhere is correlated to the performance, which suggests our measure is indeed effective for measuring attention.



### 5.6.3. Hypothesis 3: Motivation

The children were highly motivated to work with both the agent and the robot. The differences between the two were not significant. There was a small preference for the robot when they were forced make a choice between the two and this might influence the motivation in the long run.

### 5.6.4. Observations

During the experiment, the experimenters made a lot of observations that might help other researchers design their experiments or lead to new research questions. The novelty of the robot proved to be distracting, which is why we only introduced the robot after explaining the experiment. We recommend having such an introduction before an experiment, both to take away the fear and questions children might have and to reduce the possible bias of children being more interested in something just because it is a novelty to them.

During the introduction of the experiment, a week before the actual start of the experiment, children came with many questions regarding the robot, i.e.: 'will the robot hit us', 'can it dance and play football', 'can we please touch it'. They did not ask anything about the quiz or the agent. During the experiment it was noticed that children who took charge in the game during the second session (e.g. saying whose turn it was, choosing the topic), were also the children who were outgoing towards the experiment leaders, which could be interesting in relation to previous research on personality [203]. The robot/agent as peer with which the children were in competition had a positive effect for their motivation. Many children informed the experimenter about how well they did and when the robot/agent commented on the score they became extra competitive and excited when they did better than the robot/agent. Although the questionnaires showed no difference in anthropomorphisation, behavior like offering help or asking non-game related questions only occurred in sessions with the robot. By giving the robot and the agent a (different) name and constantly using this to refer to them, children seemed more ready to see them as different characters and assign them certain character traits.

### 5.6.5. DISCUSSION EXPERIMENT

The fact that we found significant differences using a participant group of only 10 children is promising. The small number of participants could also be the reason that only in attention significant results were found. Other possible reasons that we did not find learning effects is that we had a large number of children with a language deficiency; the fact that the material was on a 5th grade level, but the language proficiency of some of the children up to two and a half year lower, might have prevented these children from fully absorbing the material. Apart from that, the children already scored quite good on the baseline test (thus making it harder to see a learning effect). Last, some errors in the used text-to- speech engine may also have contributed.

Furthermore, there were some problems with the game flow. The robot/agent sometimes asked a question without announcing it, this caused many children to not quite catch the question. The biggest problem however was confusion of the answer possibilities 'b' and 'd'. The children had a hard time distinguishing the b and d from the speech of the robot/agent and might misjudge its answer. However, worse was the fact that the Wizard also had trouble distinguishing the b and d from the children's speech, which sometimes led to misjudging the children, which they found very frustrating. The setting of the experiment might also have provided a distraction: with a window looking out to the schoolyard and a window to the hallway, the room was not aural and visual distraction proof. Although one could argue this is a natural setting, it might obscure the results.

The measurements could also be improved; for instance, for knowledge we only looked at long-term retention using multiple choice, but short term and free recall are also interesting and could be added in a next experiment. For attention we now measured eye gaze directed to the opponent (robot or agent), quiz-screen and 'somewhere else'. Looking 'somewhere else' included looking at the quiz cards or obviously thinking about the answer, which is of course not procrastination. In the future, it would be interesting to see if we could distinguish game related eye gaze from procrastination or distraction.

### 5.6.6. DISCUSSION

Robots are still quite expensive and should therefore only be used in settings where they have a distinct advantage. In the current experiment the robot did not take advantage of the extra attention it received, so an agent should have the preference in the current scenario. On the other hand, a learning companion should be nice to work with for a longer period of time and the forced choice and attention that were in favor of the robot point towards a possible higher motivation in the long term for an embodied companion. Two changes in the experiment could be exploited to look if a robot has benefits as a learning companion over an agent in certain situations. The first is that the experiment could take place over a longer period of time and with more than two interactions; [111] shows that only in the third session there evolved a difference between the motivation for an adaptive or non-adaptive robot. Second, the robot could take advantage of the attention it receives to direct attention to objects in the physical world. This is in agreement with results from [226] who looked at the difference between a virtual and robot bunny that gave advice on objects in the real world versus the virtual world.

### 5.6.7. Overall conclusion

No difference between the agent and the robot was found on the knowledge tests, our measure of performance. The learning potential, of which attention and motivation are two factors, did differ between robot and agent. For motivation we did not find significant differences, but we did for attention. So one of the determinants of performance was influenced positively in this context. Overall, we can conclude that the robot has potential to be a more effective learning companion than an agent. Attention is kept longer, even in the second session; children also prefer the robot and think it is nice, which could increase the long-term motivation. However, there is no reason to prefer the robot over the agent as an educational companion in the current setting.

## 5.7. Acknowledgements

We would like to thank the Egelantier school in Soest (the Netherlands), especially the children and teacher from the 5th grade, for their cooperation.

## 5.8. APPENDIX

This appendix contains the selection of questions asked in the post knowledge test. It consists of two questions from each topic, one posed by the robot to the child and one posed by the child to the robot during the quiz. Within brackets and in italics there is some extra explanation for some of the questions, when direct translation was not possible



1. What is meant by 'energy balance'?
   - That you exercise a lot and eat little
   - That eating and exercise are in balance
   - That you eat a lot and exercise little
   - That you eat for the same time as that you exercise
2. What does the F in BOFT stand for? (BOFT is a mnemonic for children to remember four important topics of a health lifestyle (excercise, eat breakfast, drink water in stead of sweet drinks and not to much tv and videogames).
   - Biking (in Dutch Fietsen) instead of taking the car
   - Drinking fresh (in Dutch Fris) water from the tap, instead of soda
   - Soda (in Dutch Frisdrank) is tasty and healthy
   - Eat fruit (in Dutch Fruit) every day
3. Can you take a vitamin pill instead of eating fruit and vegetables?
   - No, only adults can do that
   - Yes, if you take a blue one
   - No, a vitamin pill doesn't contain everything you need
   - Yes, a vitamin pill contains everything you need
4. Why is it important that blood is pumped through your body?
   - Otherwise you wouldn't hear ticking in your chest
   - So that you can fall in love
   - It makes sure all substances in your body go to the right place
   - Otherwise your veins get a weird colour
5. Why do you need meat and dairy?
   - It protects you against diseases
   - It contains a lot of vitamins
   - It's very good
   - It helps you building bones and muscles
6. If you prefer fruit to vegetables, is it ok to just eat fruit?
   - Yes, as long as you eat a lot of fruit
   - No, you can't
   - Yes, you can.
   - Yes, as long as you eat enough apple sauce
7. What is transferred from your bowels to you blood?
   - Carbon dioxide

- Vitamins
- Fuel
- Food

8. When do you eat healthy?

- When you eat varied from all parts of the 'schijf van vijf' (Dutch equivalent of the food pyramid - 'Disc of five'), and not too much
- When you eat everything from the schijf van vijf
- When you eat something from at least three of the sections
- When only eat things from the vegetable and fruit section.

# II

# SITUATED DESIGN RATIONALE FOR ONE USE CASE

This chapter contains:

- J. B. Janssen, C. C. van der Wal, M. A. Neerincx & R. Looije. *Motivating children to learn arithmetic with an adaptive robot game*,in ICSR conference (pp. 153-162) 2011. [111]
- M. Tielman, M. A. Neerincx, J. J. Meyer & R. Looije, *Adaptive emotional expression in robot-child interaction*, in Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction (pp. 407-414). ACM 2014 [242]
- E. J. Van Der Drift, R. J. Beun, R. Looije, O. A. Blanson Henkemans & M. A. Neerincx, *A remote social robot to motivate and support diabetic children in keeping a diary*, in Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction (pp. 463-470). ACM 2014. [250]

## Adapting and relating to the user

This Part contains three papers which evaluate three functions that contribute to feeling of competence and relatedness. Figure II.1 shows the sDR of this part. Where in Part I we looked at the design and evaluation of emotional expressions, we will look at more complex interactions in this part. The emotions are incorporated into an adaptive and/or more expressive robot. This is necessary to support several methods that support the three factors of Self Determination Theory. Personalizing the activity by adapting the difficulty of exercises [111] contributes for instance to the method "Provide challenging activities" which supports the feeling of competence. The math game was also relevant for the children as it supported them in a class activity. Adapting the emotions to the child [242] is part of the function to adapt robot to child and the activity and contributes to the methods "adapt to other's behavior" which is directly connected to supporting relatedness. And the third and last paper [250], shows how the function "disclose robot information" supports the method "support mutual self-disclosure" and contributes to relatedness.

In the previous part use cases were discussed as contexts in which emotions are expressed, in this part the use cases are an integral part of the setting. This should be taken into account when transferring results to another context, e.g. self-disclosure by the robot during a game that is played for fun can have a different effect than self-disclosure during a diary task.

## Develop a set of behaviors for a robot that invoke feelings of competence and relatedness of the children interacting with the robot.

The chapters in this Part will answer the following three research questions:

1. chapter 6: Increasing motivation by adapting difficulty [111]
   1.1. Design Question: How to challenge children, aged 9-10, within their dynamic individual capabilities (c.f. Zone of Proximal Development [256] and Optimal Challenge [62]) in a math and memory game with a robot?
   1.2. Hypothesis: Child's motivation to play a math game with a robot is higher when the game is adapted to his or her dynamic individual capabilities.
2. chapter 7: Reciprocal emotion elicitation [242]
   2.1. Design question: How to model robot's emotional expressions that represent: robot's current performance, match child's intro-extroversion trait, and adapt to child's performance and emotional state?
   2.2. Hypothesis: A robot with adaptive emotional expressions will "score higher" on relatedness factors in both behaviors (emotional expressivity of the child) and opinion (fun, acceptance, empathy, trust, preference and recognized emotional expressivity) in comparison to a robot without adaptive emotional expressions.

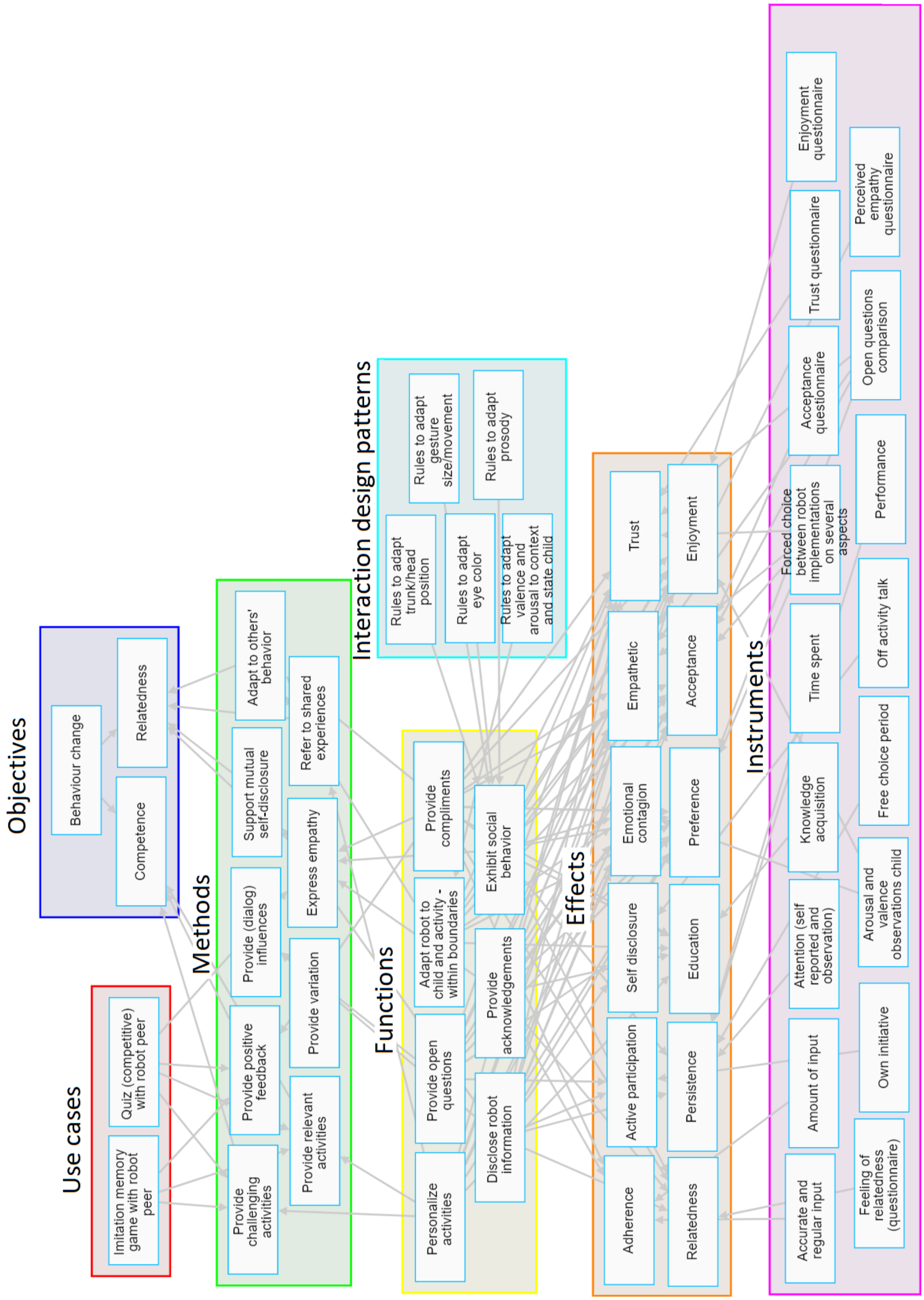



Figure II.1: situated Design Rationale of Part II

3. chapter 8: Stimulating mutual self-disclosure [150]
    3.1. Design question: How to design, within the context of a diabetes diary, self-disclosure and empathetic behavior by a robot based on mutual self-disclosure (e.g. [202]) and empathy theories [67]?
    3.2. Hypothesis: Empathetic behaviors and self-disclosure of a video-conferencing robot improve children's adherence to fill out their diabetes diary.

The overarching research question of these three chapters is: What is the effect of adaptivity and self-disclosure on liking, motivation, emotional expressivity and self-disclosure by the children?

The three studies performed in the respective chapters all yielded positive results. Adaptivity of difficulty [111], so increasing or decreasing the level according to the amount of errors increased the motivation of both slow and fast learners. Furthermore it heightened the level of assignments for the fast learners, as they were able to progress beyond the class level. The adaptive emotional expressions [242] were not so much noted, but children did react to it, by being more expressive themselves and liking the adaptive expressive robot more. And finally, the results of a robot that discloses information about itself [150] were very promising, although the number of participants in this study was severally limited, it showed that children disclosed significantly more personal experiences in their diaries when they were interacting with the robot and came to see it as a supportive friend.

Figure II.1 shows the sDR figure for these three studies. As in Part I, the concatenated sDR supports the reasoning over three chapters. By looking at the effects of the three separate studies we can make new inferences on what the effects will be when implementing a robot that adapts the difficulty of assignments, adapts its emotions to the child and discloses information about itself. In Part III we will discuss two studies in which a robot, that had multiple functionalities and was used in multiple use cases, is evaluated.

# 6 | Motivating children to learn arithmetic with an adaptive robot game

**Abstract** *Based on a 'learning by playing' concept, a basic arithmetic learning task was extended with an engaging game to achieve long-term educational interaction for children. Personalization was added to this learning task, to further support the child's motivation and success in learning. In an experiment, twenty children (aged 9-10) interacted three times, spread over days, with a robot using the combined imitation and arithmetic game to test this support. Two versions of the robot were implemented. In one implementation the complexity of the arithmetic progressed towards a predefined group target. In the other version the assignments increased in complexity until a personal end level was reached. A subsequent free-choice period showed that children's motivation to play (and learn) was high, particularly when the game progressed to a personal target. Furthermore results show that most children in the last condition reach higher levels compared to the predefined group level.*

## 6.1. Introduction

Children with a chronic lifestyle related disease have to take care of more aspects in daily life compared to their healthy peers. Support for these children in daily activities might therefore be beneficiary to them. The ALIZ-E project is aiming at a social robot for long-term interaction with these children. This robot should be able to perform three different roles over a relatively prolonged period of time: a *buddy* that provides a personalized and engaging interaction, an *educator* that teaches relevant knowledge and skills to 'empower' the child, and a *motivator* that persuades the child to adhere to a healthy lifestyle (e.g. the therapy, diet, medication) [146]. Several robot functions that support these roles are being developed incrementally, in an iterative process.

The overall scenario is based on a medical setting in Italy, where children diagnosed with diabetes will spend a complete week in the hospital after diagnosis to learn about the illness and the implications of it. For these (young) children one week away from home is a long time. The ALIZ-E robot intents to make the time in the hospital more pleasant, while supporting the education of the child's illness. Basic arithmetic skills help children with diabetes to count the carbohydrate intake for their nutrients. This ability can therefore contribute to young diabetics' self-efficacy.

In the project we use 'learning by playing' as a concept for the interaction. We combine the basic arithmetic learning task with an engaging game to achieve the project goal for long-term interaction. This last game is an imitation game, in which robot and child copy each other's sequence of arm movements and, subsequently, add a new movement to this sequence. The robot attunes the number and repertoire of moves to child's performance based on principles of motivational feedback, in such a way that the children like to continue playing until they achieve their target [203]. This ensures that the child keeps being challenged which is an important factor in both intelligent tutoring systems [267] and game theory [62]. The imitation game balances the perceived challenges with the perceived skills of the child and proves to be challenging for the children. By additional personalization of the arithmetic assignments, we expect to further improve the child's motivation and learning performance. These effects are studied in an experiment.

How to measure motivation of young children is a non-trivial question. In addition to questionnaires and observations during the game, we will evaluate the motivation for interaction with the robot by providing a free-choice period [200] [248]. Furthermore, we choose to perform the experiment with healthy children, since we want to burden sick children as little as possible [60]. By ensuring that the general characteristics of the children in the experiment are similar to the diabetic target group (e.g. age), we expect to find principles that apply for both groups. In future experiments we plan to test this in a group of children with a chronic disease.

## 6.2. ASPECTS OF MOTIVATION



Literature distinguishes two types of motivation: intrinsic and extrinsic motivation. Our research objective is to establish long-term motivation, ultimately to make a change in behaviour possible. Extrinsic motivation, though effective for short-term task compliance, has been shown to be less effective than intrinsic motivation for long-term task compliance and behaviour change' [85]. We will therefore focus on *intrinsic* motivation. Fasola and Mataric [85] indicate several factors that contribute to intrinsic motivation, including praise, competition, real-time feedback of performance, optimal challenge, self-efficacy and self-determination. Vallerand et al. [248] describe several variables that decrease the intrinsic motivation and should therefore be avoided. These variables include: material rewards, surveillance, deadlines, lack of self-determination and negative performance feedback.

For this study, we used the imitation game. In this game, the robot and the child build sequences of arm movements together. Turn by turn the players repeat the existing sequence and add a new movement to the sequence. During the game, the robot gives motivational verbal feedback to maximize the performance of the child. Most motivational aspects were already incorporated into the imitation game: positive robot feedback (praise, real-time feedback on performance), no material reward for the child and the absence of deadlines or negative performance feedback. Other aspects were difficult to manipulate: competition, self-efficacy, self-determination. Optimal challenge is a aspect we have a closer look into: when a game is too easy, the player will become bored as opposed to the game being to difficult, which will result in the user becoming frustrated or anxious [85, 62]. The optimal challenge is thus when there is a balance between perceived difficulty and perceived skills by the user. In the study presented here we implemented one version of arithmetic that should approximate the optimal challenge and one that does not.

## 6.3. IMPLEMENTATION

For this study we implemented and extended an imitation game with arithmetic assignments to mix fun and education. The new game is composed of two components: making arm movement sequences and solving arithmetic assignments. In line with Cohen et al. [54] emotional feedback in the game is attuned to match children's expectations. The game is presented to the children as a secret agent game, where arm movements are a secret code and to crack the code of the bad guys, the children have to solve arithmetic assignments. The performance on the two components of the game are not linked with each other.

Several worlds exist in the imitation game. The arm movements increase in difficulty depending on the current world (starting using one arm: 'left arm up' and extending towards both hands 'Both arms down'). Within a 'world', each player (the robot and the child) repeats the entire sequence and makes up a new movement, which is added to the sequence. The sequence ends when the length of the current world is reached or when the child makes a mistake. To prevent deception of the child, the robot does not make mistakes. The progress in 'worlds' is attuned to the performance of the child.

For the arithmetic implementation, 29 levels with 10 assignments are constructed. The levels have an increasing arithmetic difficulty (e.g. level 1: '$6+1$', level 10: '$41-10$', level 20: '$4\times 42$', level 29: '$1005\div 67$'). The assignments are selected randomly within a level and displayed on a screen next to the robot (see Figure 6.1). The robot provides motivational verbal feedback after each answer.

Two versions of the robot game are implemented: one that has a predefined arithmetic group level as end goal and one in which children could reach the boundary of their arithmetic capabilities. Next to this distinction there is a difference in the learning algorithm between the two implementations. As long as no mistakes are made both versions have an increase of three levels each step, for fast convergence to an appropriate level. When a mistake is made in the group level version the level is increased by one from that moment on, resulting in a more moderate learning curve. For the personalized level implementation, a simple form of sensitivity analysis is used [211]. In the case of a mistake, the level is decreased by one, increasing the self-perception of arithmetic skills. Next to this, the levels are also increased by one from that moment on. Other motivational aspects including the arm movement part are not manipulated, in order to get a fair comparison between group and personalized level implementation.

The group goal is set at level 20, which was considered the appropriate level for children half way through year six (fourth grade in U.S.) based on information from Goffree [97] and Borghouts [35]. For the personalized level 29 is the maximum level, because the chances are slim that the children reach this level.

The robot contained a user model, which kept track of the movement progress and arithmetic level of each participant. The participant continues with the movements and assignments in the level they ended last interaction time.

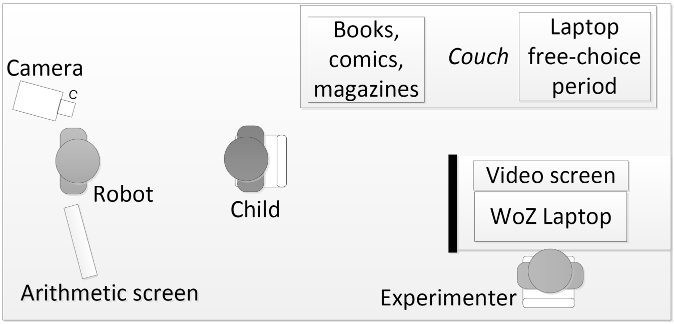



Figure 6.1: The experimental setup

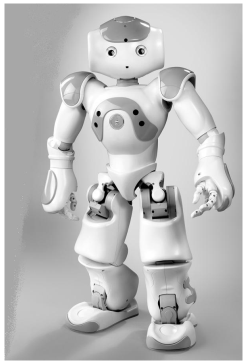


Figure 6.2: The NAO robot

## 6.4. EXPERIMENTAL METHOD

### 6.4.1. PARTICIPANTS

Participants were 20 Dutch children (11 F and 9 M, age 9 - 10 years) from elementary school 'Het Spoor' (Zeist, The Netherlands). This is a Jenaplan school, meaning that each child follows their own learning curve but still has to reach goals within a time frame. Balancing for their gender, the participants were randomly divided in two groups of 10 participants. All the parents/caregivers signed an informed consent.



### 6.4.2. MATERIALS

#### NAO ROBOT

The robot used in this experiment was the NAO (Aldebaran Robotics, see Figure 8.1). The NAO was provided with a unisex name: Charlie. Charlie, and names with similar pronunciation, is an uncommon name in the Netherlands both for boys (494 in 2006) and girls (363 in 2006). We provided no clues about the gender of the robot, since we wanted to prevent the children being prejudiced to liking the robot because of its gender. Fluency TTS (v4.0, using neutral voice 'Diana') was used to generate the wav-files the NAO used. The software for executing the imitation game involves: a Wizard of Oz interface, a dialogue model, a user model, the arithmetic assignments database. The control software ran on various computers.

#### WIZARD OF OZ

In order to test the feasibility of components before complete implementation, we use a Wizard of Oz set-up. In this interface, the experimenter does the sensing (e.g. the wizard interprets the movements and speech of the children), initiates the dialog and controls the laptop that displays the assignments and the progress. The movements of the robot are also preprogrammed and the experimenter just has to press the right order of buttons to make the robot imitate the children. To the children it looks like the robot actually recognizes and remembers the set. At a later stage fully autonomous robot behaviours will be tested within the project.

### EXPERIMENTAL SET-UP

The experiment was conducted at the school in an office space. Unfortunately, there was no possibility for the experimenter to occupy a different room nearby, so the experimenter was in the same room as the child and the robot. The effects of the presence of the experimenter were minimized by placing a covering screen. The interaction was recorded on video. Figure 6.1 shows the experimental setup.

## 6.4.3. EXPERIMENTAL DESIGN

The experiment performed had a between-subject design. The independent variable was the goal that could be reached in the arithmetic assignments, either the group or the personalized level. Between the two groups, one group interacted with a robot that adapted the level of the assignments to the child's performance and could proceed beyond the group level. The other group interacted with a group goal robot that followed a standard learning curve where the group level was the highest level that could be reached. The dependent variables were arithmetic performance and intrinsic motivation.

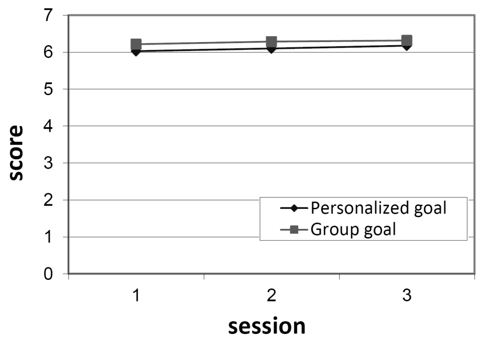
### MEASURES

Two measures were used to measure the intrinsic motivation of the children to play the game with Charlie. A **questionnaire (subjective)** was constructed based on the Intrinsic Motivation Inventory (IMI)[1] (see other research [57][193][72]). Two of the seven subscales of the IMI were included: Interest/enjoyment (intrinsic motivation for playing game with the robot, 7 questions) and Relatedness (bond with the robot, 8 questions). The answers were measured using 7-point Likert scales (1 being negative and 7 positive). The original questionnaire in total and the separate subscales individually, were all validated. The questionnaire was translated into Dutch focused on children. The layout was altered for every session to keep children motivated to complete the questionnaire.
As an objective measure, the **free-choice period** [200] [248] was used. The free choice period was a period of five minutes in which the child could choose what to do: keep playing with the robot, read children's comics or do interactive Internet learning games on a computer. The time spend interacting with the robot was measured and functioned as an objective measure for the intrinsic motivation of the child to interact with the robot.

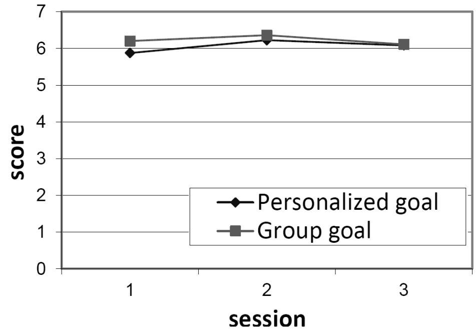
### PROCEDURE

During a short introduction in class, the children were able to see the robot beforehand. For the individual interaction moments, the experimenter introduced the child to the robot and explained the course of the experiment. Each interaction session lasted about 20 minutes, based on the average attention span for children of this age [58]. The child played the game with the robot for about 15 minutes. The game was ended after the 13th minute at a natural moment when the level was completed, resulting in a 13 to 17 minute interaction time. Afterwards the free-choice period was started by the experimenter. The researcher stated that the experiment had ended and that the child had 5 minutes to do as it pleased, choosing between the mentioned options as long as it stayed inside

[1] http://www.psych.rochester.edu/SDT/measures/IMI\relax$\@@underline{\hbox{}}\mathsurround\z@$\relaxdescription.php




(a) Interest/Enjoyment




(b) Relatedness

Figure 6.3: Questionnaire results

the room (options were presented in a random order). Finally, the child completed the questionnaire. The experiment was performed three times for each child over the course of two weeks. The rationale behind this was to experiment with the constructed user model and to overcome the initial enthusiastic response displayed by children when first meeting the robot. To reward the children, they received a picture of themselves with the robot. The school received technical Lego and was given a robotics lesson for the class after all sessions were completed.



## 6.5. RESULTS

### 6.5.1. MOTIVATION

First the quantitative results will be discussed, based in the two different motivation measures used in the experiment.

### 6.5.2. QUESTIONNAIRE

The results of the intrinsic motivation questionnaire are analyzed for each session. The answers represent the motivation to play the game with the robot and the bond with the robot. From the analysis the participant that did not start the third session is excluded and missing data is filled with a random participant from this condition to make ANOVAs possible.

We expected that the children in the personalized goal robot condition would score higher on the motivation scale than the children in the group goal condition. Results show that both scales are rated high (see Figure 6.3). The standard deviations are small, they ranged for Interest/Enjoyment from 0.39 to 0.70 and for Relatedness from 0.37 to 0.66. For Interest/Enjoyment the repeated measures ANOVA over the runs has as result: $F(2, 48) = 0.01$, $p=0.99$. The result for the ANOVA for the Relatedness questionnaire is $F(2, 48) = 0.16$, $p=0.85$. Thus both questionnaires do not provide significant differences between the two conditions.

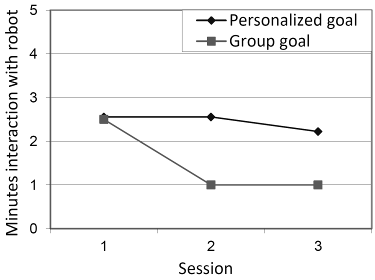



Figure 6.4: Amount of time spend with robot during free-choice period.

| Personalized goal robot | | | | Group goal robot | | | |
|---|---|---|---|---|---|---|---|
| Child | Run1 | Run2 | Run3 | Child | Run1 | Run2 | Run3 |
| 1 | * | * | 0:00 | 2 | 0:00 | 0:00 | 0:00 |
| 3 | 2:20 | 3:13 | 5:00 | 4 | 5:00 | 5:00 | 5:00 |
| 5 | 5:00 | 0:00 | 0:00 | 6 | 5:00 | 0:00 | 0:00 |
| 7 | 0:00 | 5:00 | 5:00 | 8 | 0:00 | 0:00 | 0:00 |
| 9 | 0:00 | 0:00 | 0:00 | 10 | 5:00 | 0:00 | 0:00 |
| 11 | 5:00 | 5:00 | ** | 12 | 5:00 | 0:00 | 0:00 |
| 13 | 5:00 | 5:00 | 5:00 | 14 | 0:00 | 0:00 | 0:00 |
| 15 | 0:00 | 0:00 | 0:00 | 16 | 5:00 | 5:00 | 5:00 |
| 17 | 5:00 | 5:00 | 4:30 | 18 | 0:00 | 0:00 | 0:00 |
| 19 | 0:00 | 0:00 | 0:00 | 20 | 0:00 | 0:00 | 0:00 |
| AVG | 2.6 | 2.6 | 2.5 | AVG | 2.5 | 1.0 | 1.0 |

Figure 6.5: Results of the free-choice period. Time interacting with the robot (in mm:ss). * entails the child stopped the interaction before the free choice period started, ** entails the child was absent

### 6.5.3. FREE-CHOICE PERIOD

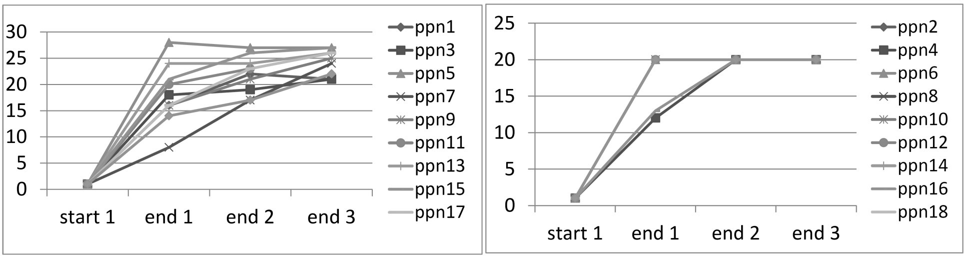
Table 6.5 shows the amount of time spend with the robot per participant for each session during the free-choice period and Figure 6.4 shows the means graphically. In the free-choice period following the first interaction the time spend with the robot is about the same (mean personalized = 2.6min, mean goal = 2.5min). This was expected beforehand, due to the new experience of the interaction with the robot. Video footage shows that the children were in general very excited to play with the robot. After the first session, the results started to differentiate between the two conditions. Most children that interacted with the personalized goal robot continued to play with the robot during the free-choice period, whereas the children that interacted with the group goal robot displayed, on average, a decline in the amount of time spend with the robot during the free-choice period. Child 1 stopped the experiment before the free-choice period and child 11 was absent during the third session resulting in missing data points.

Because the results are not evenly distributed, we ran a nonparametric Mann-Whitney U-test to establish whether the differences between the two conditions are significant. A one-tailed Mann-Whitney U-test showed that the difference was significant ($p < 0.05$) in favour of the personalized goal robot.

### 6.5.4. RESULTS ARITHMETIC ASPECTS

We expected the children to reach arithmetic level 20, which corresponded with half way through 6th grade. However figures 6.6a and 6.6b show that the children that interacted with the personalized goal robot, performed above the expected norm on the arithmetic assignments (average 24.7). Especially child 5 stands out in arithmetic skills. The graphs show that most children already reached level 20 after the second interaction. From these results, we can derive that most children participating in the experiment are ahead in their arithmetic education and that playing with a personalized goal robot makes sense, since the individual levels differ from each other.

We looked into the interaction between the free-choice period and the performance on the arithmetic assignments. When Figure 6.6a and Figure 6.6b are linked with Table




(a) Personalized goal

(b) Group goal

Figure 6.6: Performance for arithmetic assignments

6.5, it shows that the two children who played with the group goal robot during the free-choice period after session 2 and 3, were actually the children that did not reach level 20 after the first interaction. It appears as though the continuing increase in level motivated the children to play with the robot during the free-choice period. When looking at the children that played with the personalized goal robot, we see a similar trend. Child 5 performed very well on the assignments and reached his personal level during the first session. During the free-choice period the child chose to read instead of playing with the robot. However, Child 13 who also reached his personal level at the first session, did continue playing with the robot during the free-choice period. Hence, some children who perform at top level still like to play with the robot.



## 6.6. Conclusion & Discussion

In this paper we present a study that builds upon the principles of learning by playing. By combining a basic arithmetic task with an engaging game, we create a robot game for children. In an experiment we look whether personalization of the learning task has an effect on children's motivation and learning. In general we found that the children are very motivated to play the game with the robot. The motivation stays at a high level for all three interaction moments. The objective motivation, free-choice period, stays high when they interact with a robot offering a personalized learning goal. Most children who play with the personalized goal robot keep interacting with the robot the full five minutes of the free-choice period and the two children who are a bit slower to reach level 20 in the group goal session keep interacting with the robot during the free-choice period.

The personalized goal version shows that the group goal is not high enough for most of the children to reach their maximum capabilities. The group goal is thus not challenging. In sum, this robot game provides a promising approach to support long-term interaction even when the interaction is not all about fun. This is promising for the use of a social robot for long-term interaction with diabetic children. In a next study, diabetic children will participate in the study to see if the results can be reproduced with this specific population.

From a methodological perspective, the free-choice period proves to be very useful to study motivation effects with children. It appears that children answer the questions socially desirable. Despite several urges of the experimenter to rate how they really feel about the game, children seem to stay away from the 'negative' answers even though some children seem sometimes a little bored during the game (based on video footage). In future, we plan to include more detailed observations on communication behaviour, like eye-contact (gaze wondering off). In addition, we will improve the questionnaires. For example, research on Likert scales for children suggests to use a 3-point scale [154].

## 6.7. ACKNOWLEDGMENTS

This work is (partially) funded by the European Union FP7 ALIZ-E project (grant number 248116). Furthermore the authors would like to thank the teachers and the children of 'Het Spoor' (the school) for their participation in this study.

# 7 | Adaptive Emotional Expression in Robot-Child Interaction

**Abstract** *Expressive behaviour is a vital aspect of human interaction. A model for adaptive emotion expression was developed for the Nao robot. The robot has an internal arousal and valence value, which are influenced by the emotional state of its interaction partner and emotional occurrences such as winning a game. It expresses these emotions through its voice, posture, whole body poses, eye colour and gestures. An experiment with 18 children (mean age 9) and two Nao robots was conducted to study the influence of adaptive emotion expression on the interaction behaviour and opinions of children. In a within-subjects design the children played a quiz with both an affective robot using the model for adaptive emotion expression and a non-affective robot without this model. The affective robot reacted to the emotions of the child using the implementation of the model, the emotions of the child were interpreted by a Wizard of Oz. The dependent variables, namely the behaviour and opinions of the children, were measured through video analysis and questionnaires. The results show that children react more expressively and more positively to a robot which adaptively expresses itself than to a robot which does not. The feedback of the children in the questionnaires further suggests that showing emotion through movement is considered a very positive trait for a robot. From their positive reactions we can conclude that children enjoy interacting with a robot which adaptively expresses itself through emotion and gesture more than with a robot which does not do this.*

## 7.1. Introduction

One of the most promising fields in human-robot interaction is robot-child interaction. Children like robots, are more forgiving than adults when robots make mistakes and are quicker to ascribe human characteristics to robots[20]. In many applications of robot-child interaction, such as a robot as teacher or motivator, this interaction will take place for a longer period of time. Research has shown that for persistent interaction between robot and child, the child has to establish a social bond with the robot [118]. Forming a social bond is a complicated process, in which several aspects play a role. One aspect is expressive behaviour, which is important in showing internal states to an interaction partner. Two concrete examples of expressive behaviour are showing emotion and gesturing. Humans show their emotions in various ways and gesture while they speak to clarify their meaning. These expressions are very important in interactions and the forming of social relationships [42, 159].

Although much work has been done on expressive behaviour, few studies have integrated both emotional behaviour and gesturing. Moreover, research which takes into account the emotional state of the interaction partner is sparse, especially when taking into account the important role contagion plays in human interaction [157]. This paper presents a study of the role of the adaptive expression of gestures and emotion in robot-child interaction, based on the emotions of the interaction partner and relevant occurrences to the robot. In order to study this issue, the expressions first needed to be developed for the Nao robot. Based on previous research, a model of the adaptive expression of emotion and gestures for the humanoid Nao was designed and implemented. In this paper, we present the results of an experiment using a Wizard of Oz design which was done using this model, where we studied the influence of the expressions on the interaction behaviour and opinions of children.

## 7.2. MODEL

The Nao is a 57 cm tall humanoid robot, developed by Aldebaran [1]. It has 25 degrees of freedom in its body, but does not have movable facial features. The Nao is very suitable for robot-child interaction because of its size and appearance. Moreover, because of the degrees of freedom it is capable of many different bodily expressions, which makes it a good platform for expressive behaviour.

### 7.2.1. PREVIOUS RESEARCH

In emotion research, two approaches exist. The first approach distinguishes several basic universal emotions, such as happiness and sadness [81]. The other approach considers each emotion to be a specific combination of arousal, or how exciting the emotion is, and valence, how positive the emotion is [210]. Happiness is thus an emotion with a high arousal and high valence, sadness an emotion with low valence and low arousal. This is also the approach taken in this paper, as the combined use of arousal and valence allows us to design complex emotional states and smooth transitions between the basic emotions.

In order to design human-like expressive behaviour for a robot, it is first important to know how humans express themselves in interactions. We can distinguish three ways in which people express their emotion, namely through facial features, through body movement and through voice. As the Nao robot cannot display facial features, only body movement and voice were considered. People can recognize emotions from body pose alone, especially the emotions happiness, anger and sadness [59]. These poses have also been implemented in the Nao robot, and were well recognized by both children and adults [13, 14]. When considering body movement, trunk position especially is related to the valence of the emotion felt [70], while head position has a strong influence on both perceived valence and arousal [13]. Both adults and children can also recognize emotion from vocal cues alone [127]. The fundamental frequency of voice, speech rate and speech volume of the voice all seem to be related to the arousal of the emotion felt [5].

Aside from considering how people show emotion, it is also important to look at when they show them. This is crucial as people show emotion tailored to the context and reacting to their interaction partner. They mimic the emotions of others, smile when

[1] http://www.aldebaran-robotics.com/

7

they smile, frown when they frown [102]. People are also influenced by the emotions of others, emotions are contagious [157]. Although the exact link between mimicking and emotional contagion is not quite clear, both processes clearly exist in human interaction.

Most gestures used by people in interactions are classified as spontaneous gestures, as they are made without conscious thought. Four different types of gestures can be distinguished. Iconic gestures refer to concrete events and are closely related to the semantic content of the utterance. Metaphoric gestures are pictorial like iconic gestures, but represent abstract ideas. Beat gestures are related to the rhythm of speech and consist of moving the hand up and down in short movements. Finally, deictic gestures are pointing gestures. Iconic and metaphoric gestures have a preparation phase, a stroke and a retraction phase, beat and deictic gestures only have a stroke and retract phase. In all cases, the stroke of the gesture coincides with the related part of speech. Not all types of gestures occur equally often, beat and iconic gestures being the most common. Which kind of gesture occurs is also related to the type of clause they occur with. Narratives are subject to sequential constraints, extranaratives are not. Iconic and deictic gestures occur most with narrative clauses, metaphoric gestures most with extranarrative clauses [159].

Gestures serve several purposes, including that they show us what the speaker finds relevant [47]. Related to this is that gestures can tell something about the speaker, for instance pointing towards or from yourself can tell if you feel close to something. In robotics, gestures also serve to make a robot more life-like, as people almost always use gestures when speaking. Most systems which generate gestures for robots rely on a textural analysis to generate their gestures [48]. The effects of a robot gesturing on the opinions of their human interaction partners is not quite clear. Although some studies find that gesturing is always positive, no matter if it is semantically congruent [217], other results indicate that a robot gesturing might create a cognitive overload [131].



Aside from emotions felt and text spoken, personality also influences how we express ourselves. Extroversion in particular is a trait which influences speech and movement. Extrovert people have a stronger voice, smile more, move quicker and move more than introvert people [36]. Personality is very relevant for robotics, a study using extrovert and introvert robots has shown that people tend to like robots with a personality comparable to their own [115].

There are several ways to implement emotions and gestures in a robot. The most important design choice here is to have either a functional or a biological inspired robot. If the robot reacts directly to input from the environment it is functional, if it reacts according to its internal state, which in turn is influenced by input from the environment the robot is biological inspired. Some studies use a functional approach, such as [221], who employ a state-machine. Most current systems are biologically inspired, however, as the existence of an internal state makes for a more insightful model and can help a robot with long-term interaction and decision making. [45], for instance, use a stimulation model strongly inspired by emotion theory. [106] have developed an architecture based on motives, emotional state, habits of interaction and percepts of interaction.

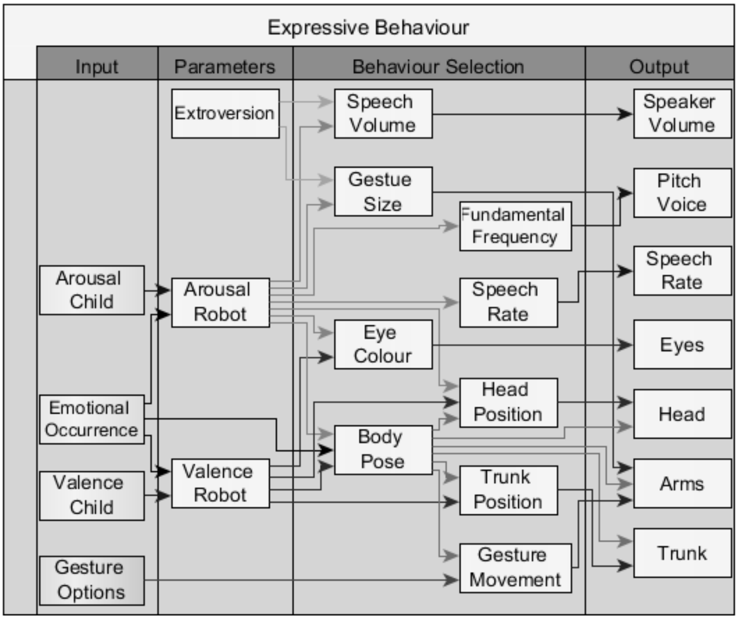



Figure 7.1: Model for expressive behaviour of the Nao robot

### 7.2.2. ADAPTIVE EMOTION EXPRESSION

Based on the knowledge from previous studies, it is possible to design a model for the adaptive expression of emotion and gestures for the Nao robot. The full model is presented in Figure 7.1. It consists of four phases, an input phase, adapting the internal parameters based on this input, reasoning about the correct behaviour and the output to the Nao robot. In this section, all phases will be discussed in more detail.

The model needs *input* from its environment in order to decide on the correct emotional and gesture behaviour. The first kind of input is information about emotional occurrences. The emotions of people are influenced by their environment, so the same should be the case for the robot. For this reason, it is important for the model to know when things take place which influence the emotions. An example is the robot winning a game, which is an occurrence with a high arousal and valence. The second kind of input is the emotion of the child. As seen in the previous section, people are influenced by the emotions of our interaction partners, so the robot needs to have information about the arousal and valence of the child. The third kind of input are the possible gestures. Based on the text which the robot will speak, several possible gestures can be derived. This will happen outside of this model and can either be automated, or in the case of this study, hardcoded. The system will be presented with a list of the possible suited gestures and their type.

The model has three *internal parameters*, namely its extroversion, its arousal and its valence. Extroversion of the robot will be based on the extroversion of the child as we have evidence that people like a robot similar in personality. This parameter is constant.

The arousal and valence of the robot are represented on a scale from -1 to 1 and will be influenced by both emotional occurrences and the emotions of the child. Whenever an emotional occurrence takes place, the emotions of the robot will move in the direction of the occurrence. For instance, if a happy occurrence takes place with arousal 0.8 and valence 0.9, the arousal of the robot will move halfway to 0.8 and the valence halfway to 0.9. When no such occurrence takes place, the robot is influenced by the emotions of the child. In this way emotional contagion is considered. The emotions of the child will influence the robot in the same way as emotional occurrences, with the exception of the situation where the emotion of the child becomes too extreme. Whenever the valence or arousal of the child drops too low, or the arousal rises too high, the robot will compensate. This should exclude situations such as the child being very sad and becoming even sadder because the robot is very sad.

Based on the literature, several aspects of *behaviour* have been incorporated in this model. Emotions will be shown by the robot through full body poses as developed and validated by [13, 54] and Aldebaran[2]. These poses will only be executed when emotional occurrences take place, as it is impossible for the robot to constantly use them. The happy pose, for instance, has raised arms, which would make playing a game with a child very difficult. The head position of the robot will be influenced by both arousal and valence, the higher these values the higher the head position. The trunk position will be similarly influenced, but only by valence. The robot also has the possibility of changing its eye colours. Red colours will be associated with high arousal emotions, blue colours with low arousal emotions [122]. The voice of the robot will be influenced by its arousal. The higher arousal, the louder the robot will speak, the higher pitched its voice will be and the higher the speech rate. Speech volume is also influenced by extroversion, the higher the extroversion, the louder the voice. Finally, gesture movement will be chosen based on the type of narrative the gesture relates to and the type of gesture. Knowing how often people use specific kinds of gestures, the model will choose between the options reflecting this. The size of the gesture movement will be influenced by both arousal and extroversion [268].

Finally, the model has an *output* module in which the behaviours will be translated into specific voice characteristics and joint values for the Nao robot.

### 7.2.3. IMPLEMENTATION

The model was implemented for the Nao robot in the Prolog-based BDI-Agent language GOAL[3]. A GOAL program consists of a knowledge base with static facts, a belief base with changeable beliefs, a goal base with changeable goals, an action base specifying the actions to the environment, a program module specifying which actions to perform and which beliefs to change in which circumstance and an event base which processes the input from the environment. In the implementation of this model, the knowledge base was used to represent the dependencies between specific behaviours (such as head position) and the internal parameters. These parameters, along with information about the environment, were stored in the belief base. The program module specified when to adapt behaviours. Due to technical constraints, it was not practically possible to make

[2] http://www.aldebaran-robotics.com/
[3] http://mmi.tudelft.nl/trac/goal

gesture and speech perfectly synchronized in the Nao robot. We chose to still work with these imperfect gestures, as research has revealed that incongruent gesturing might still be perceived as more positive than no gesturing at all [217].

## 7.3. Experiment

In order to test the effect of the adaptive expression of emotion and gestures in robot-child interaction, an experiment was done. In this experiment, children played a quiz with a robot that shows the model-based adaptive expressive behaviour and a robot without such a model. We wished to know what the influence of the adaptive expression of emotion and gestures was on the opinions of the children about the robot and on the expressiveness of the children.

### 7.3.1. Experimental Method

#### Experimental Design

We applied a within-subjects design with a two-level independent variable: the adaptive expressive behaviour of the robot. One robot displayed adaptive expressions of emotion, the other did not. This means that two robots were used, only one using the model for adaptive expressive behaviour. From this point, we will call these robots the *affective robot* and the *non-affective robot*. The affective robot adapted its emotions and showed these through voice, body movement, body pose and gesture. The non-affective robot only showed small randomized body movements not related to emotion, such as swaying in the hips and slightly moving the arms. The two dependent variables in this experiment are the opinions of the children and the expressive behaviour of the children when interacting with the robot. During the experiment we also looked at the interpersonal synchrony between the emotions of child and robot, but as these results were of secondary importance, we have chosen to leave them out of this paper. Full results can be found in [241]

7

#### Participants and robot settings

All participants for this experiment were children from the primary school Dalton Lange Voren in Barneveld (group 5 and 6). 18 children participated, mean age was 8.89, SD 0.81. 9 boys and 9 girls participated. The mean extroversion of the children was 69, SD 10. During the interaction, the robot will adopt the extroversion of the child as its own. In order to determine the extroversion of the child, the corresponding questions from the BFQ-C questionnaire were used. This questionnaire is validated for children [166] and will give an insight to the extroversion of the children in the form of a score between 0 and 100.

#### Task

In order to test the effect of the adaptive expression of emotion in robot-child interaction, the child and robot need to interact in a meaningful way. In this experiment the children were told to play a quiz with the robot. In this activity the child and robot are seated across from each other, with a tablet on a seesaw between them projecting the quiz questions as seen in Figure 7.2. The game starts with the robot asking the child a question and then showing the child the questions on the tablet. The child then has to

Figure 7.2: A child playing the quiz with the robot using the tablet and seesaw

answer the question, getting two tries. When the question has been answered the turn goes to the robot. A new question will appear on the tablet, including the possible answers, which the child reads to the robot. The robot will then try to answer the question. This procedure is repeated until the quiz stops after 12 questions, 6 posed by each player. All questions are multiple choice with four possible answers. The robot has a 75% chance of answering the question correctly. The quiz questions were either trivia or on health subjects.

Before and after playing the quiz, the robot will have a short conversation with the child. It will first introduce itself, ask the child about its interests, such as hobbies and tell something about itself in this conversation. At the end of the quiz, the robot will tell the child who has won the quiz, express that it liked playing and say goodbye. The entire experiment was conducted in Dutch.



### MEASURES

A common problem with experiments testing children‘s opinions on robots is a ceiling effect. Children like all robots so much that it becomes impossible to distinguish between conditions. For this reason, two kind of measures were used in this experiment, video analysis to study the expressive behaviour of the children and questionnaires to get to know their opinions. We have added the video analysis in the hope of getting a better understanding of the unconscious opinions of the children, as conveyed by their behaviour. In order to study this behaviour, all interactions were filmed and the behaviour was analysed. The videos were annotated on several specific behaviours, such as smiles and frowns. A full list of the behaviours can be found in Table 7.1. From these behaviours, we can calculate two measures. The first is the weighed frequency of the expressions of the children, which is calculated by taking the frequency scores of the behaviours and adding them up, counting smiles, frowns and startles once and laughter, bouncing, positive vocalization, shrugging, sighing and negative vocalization double. The second measure is the valence of the expressions, which is calculated by taking the frequency of positive expressions (counting stronger expressions twice) and subtracting the frequency of negative expressions. The corresponding formula is as follows:

**Valence expressions** = Smiles + 2x (Laughter + Bouncing + PosVocalization) - (Startle + NegVocalization) - 2x (Shrugging & Sighing)

| **Expression** | **Properties** |
|---|---|
| Smiles | All instances where the mouth of the child angles upwards. As we only count instances and not duration, this was only counted when there was a change. So only when the mouth angles rose upwards. |
| Laughter | All cases in which the child laughed. Laughter is here classified as those smiles which are accompanied by sound or movement of the chest related to the happy feelings. |
| Excited bouncing | All cases in which the child either bounced up and down out of obvious excitement, or in which the child made a large excited gesture. An example of the latter is raising both arms, and other such gestures of success. |
| Positive vocalization | Every positive exclamation not directly related to the dialogue. Common words are *yay* or *yes*. |
| Frowns | All facial expressions obviously related to thinking, concentrating or misunderstanding. Also all facial expressions where the eyebrows are lowered. |
| Shrugging & Sighing | Raising the shoulders and dropping them again, or audibly letting out air. These two expressions are seen as signs of boredom |
| Startle | All signs of involuntary fright from the child, such as it being startled by sudden movement. |
| Negative vocalization | All negative exclamations not directly related to the dialogue, such as *nou zeg* or *jammer*. |

Table 7.1: Expressions and their definitions

| Subject | Nr. of questions individual robot | Nr. of questions forced choice |
|---|---|---|
| Fun | 9 | 1 |
| Acceptance | 3 | 1 |
| Empathy | 3 | 1 |
| Trust | 3 | 1 |
| Emotions | 3 | 1 |
| Preference | 0 | 1 |

Table 7.2: Topics of questions in questionnaires

Aside from the behaviour of the children, we also measured their subjective opinions through questionnaires. Although previous work has shown a ceiling effect with questionnaires, we still added them in hopes of being able to compare results between different studies. Two types of questionnaires were used, one questionnaire about an individual robot and one forced-choice questionnaire in which children had to choose between the two robots. Both questionnaires had questions on the same subjects. Table 7.2 shows the topics of the questions and the number of questions per questionnaire.

### WIZARD OF OZ

As described in the implementation section 7.2.3, the GOAL language was used to implement the model for the adaptive expression of emotion and gestures. For this experiment, however, a final step was necessary as the model relies on input, such as the emotions of the child. In the current experiment, an experimenter provided this information via a Wizard of Oz (WoOz) program. This interface allowed the experimenter to provide the valence and arousal of the child, giving guidelines in the form of specific emotions. Figure 7.3 shows the WoOz interface for the emotions of the child. The experimenter also performed the dialogue selection for the robot, all pieces of dialogue were scripted. The emotional occurrences were scripted into the dialogue, as the robot will always say something reacting to these occurrences. For instance, when the robot wins a game it will say *Yay! I‘ve won!*. With selecting this dialogue, the experimenter also sends the corresponding input to the model, which will automatically adapt the emotions of the robot accordingly and send a happy pose to the robot. The gesture input was scripted in a similar manner. Whenever a piece of dialogue was selected by the experimenter, the model received input on the possible gestures to display. The model automatically chooses which gesture is actually displayed. During the experiment, the experimenter operating the WoOz was sitting in the same room as the children, as it was necessary to see the child‘s face to interpret the emotions and the location did not allow for a video set-up.

### MATERIALS

The list of materials for this experiment can be divided into two categories, the technical devices and the computer programs. When it comes to the technical devices, two Nao robots were used, a video camera, a Dell laptop, a TP-Link router and one Samsung Galaxy tablet on a seesaw. The laptop was used by the WoOz and ran the WoOz interface

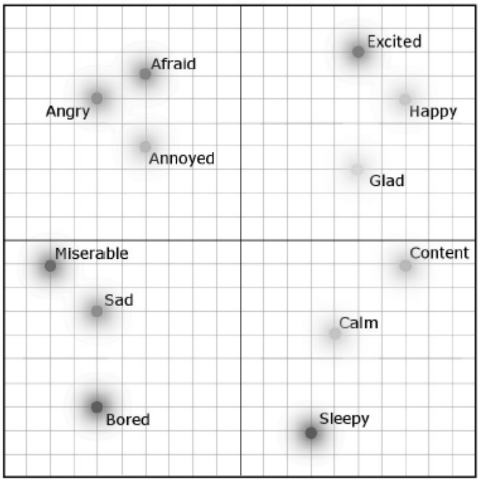



Figure 7.3: The interface via which the experimenter provided information about the arousal and valence of the child. The horizontal axis represents the valence of the child, the horizontal axis the arousal. The coloured points are references to specific emotions as context.

program through which the dialog was managed, the quiz operated and the emotional state of the child communicated to the robot. It also ran the GOAL program which made the decisions on which behaviour to display in the way described in section 7.2.2. Because the two robots used are identical in appearance, both wore a different little shirt. One robot had a plain orange shirt, the other a striped white and orange shirt. These shirts were used to make sure that the children understood that there were two different robots and help them to keep the robots apart. In addition to keeping the robots apart, it was important that the children remember the names of the robots, as the questionnaires refer to them by the names *Charlie* and *Robin*.

### PROCEDURE

The experiment was conducted in two sessions, an introduction session and an experimental session. The introduction session was the same for all participants and took the form of a short classical lesson with the robots. In this lesson, one robot was introduced to the children in order to make them more familiar with robots and to hopefully lessen the ceiling effect where robots are considered so cool that there would be no discrimination between conditions. The robot used in the introduction did not wear a shirt and was given a different name than the robots used in the experimental sessions. After the introductions, all children filled in the BFQ-C questionnaire. In the experimental session, the first robot was always named Charlie and always used the same dialogue and questions, while the second robot was always named Robin and also always used the same dialogue and questions (different from the first robot, of course). Which robot displayed the adaptive expressions of emotion and gestures was counterbalanced, half of the children played the first quiz with the affective robot, half with the non-affective robot. The children were shown into the room and the experimenter first explained the

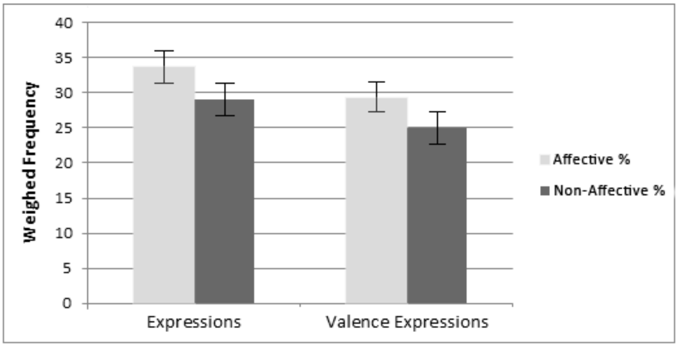



Figure 7.4: The weighed frequency of the expressions of the children, as well as the valence of their expressions.

quiz. In all sessions the first robot started with introducing itself to the child. After a short conversation about their interests, the robot asked if the child still understood the quiz and explained again when necessary. Next, the child and the robot played the quiz. After 12 questions (about 10 minutes), the robot ended the quiz and the interaction. The children were then presented with the questionnaire about the first robot. The first robot was then taken away, but kept in sight, and the second robot was brought to the child. The reason both robots were kept in sight is to ensure that the child viewed the robots as two different entities. The procedure described was repeated, the second robot introduced itself and had a short conversation with the child. The quiz was played for 10 minutes after which the robot ended the interaction and the same questionnaire as before was presented. After this, one more questionnaire about the differences between the robots was presented. The session ended with the possibility for the child to take a picture with one of the robots.



### 7.3.2. RESULTS

#### EXPRESSIONS

The first set of results are those representing the expressions of the children during the interaction. In one session there was a technical problem with the camera, meaning that for one subject no video was available for analysis. The expressions were scored as described in the Measures section (7.3.1), by the experimenter. In order to ensure the objectivity of this scoring method, two children were also scored by a second experimenter. These results show that the differences between conditions are comparable. For instance, experimenter 1 counted 30 smiles with the affective and 14 with the non-affective robot, while the second experimenter counted 24 versus 11 smiles. All other expressions also showed only minor deviations or were identical. Figure 7.4 shows the weighed frequency scores of expressions of the children when interacting with the affective and the non-affective robot. The results for the expressions of the child for the affective robot (M=33.59, SE=17.34) are significantly higher than for the non-affective robot (M=29.06, SE=13.53), $(t)(16)=2.156$, $p<0.05$, $(r)= 0.47$ (one-tailed). Of course it is

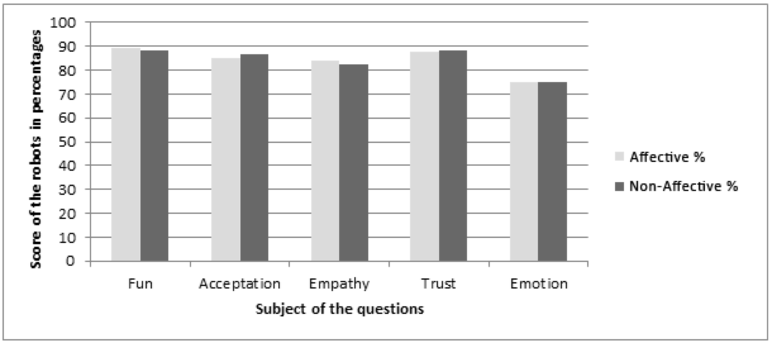



Figure 7.5: Opinions of the children on both the affective and the non-affective robot on several subjects.

important to also consider the valence of the expressions of the children. We would like to know if children react more positively or negatively to the affective robot. These results show that the children had a significantly higher valence in their expressions with the affective robot (M=29.24, SD=16.75) than with the non-affective robot (M=24.94, SD=13.89) $(t)(16) = 2.251$, $p<0.05$, $(r) = 0.54$(one-tailed).

### QUESTIONNAIRES

Figure 7.5 shows the results from the first questionnaire, about the individual robots. The questions were asked on a scale from 1 to 5, meaning that a score of 100% corresponds to the most positive answer given to every question and a score of 0% to the most negative answer given to every question. Both robots scored very high, the difference between the affective and the non-affective robot is not significant for any of the question topics.

7

Figure 7.6 shows the results from the second questionnaire, comparing the robots. Some data were excluded from this dataset, based on the motivations of the answers given. A preference for a robot motivated clearly by reasons which can be contributed to un-planned circumstances was not taken into account. One example is a child disliking one robot because it was slow to answer questions, which was caused by a crash of the program. When considering Figure 7.6, note that the children had to choose between the robots, meaning that the scores for any subject will add up to 100%. All these scores are based on a single question. Although some differences can be seen, none are statistically significant.

Aside from asking the children about their preference, the final questionnaire also asked for motivations. These motivations can be classified into different categories. Figure 7.7 shows the number of times that each kind of motivation was given for each robot. This figure also shows how often a child who gave a certain motivation eventually chose the affective robot or the non-affective robot in the final question of the forced-choice questionnaire. This question was which robot they preferred most, so the coordinate gives an indication of the influence of the motivation for the final preference. The most noticeable motivations are clearly that the non-affective robot was more understandable, while the affective robot was preferred most often because it showed emotions.

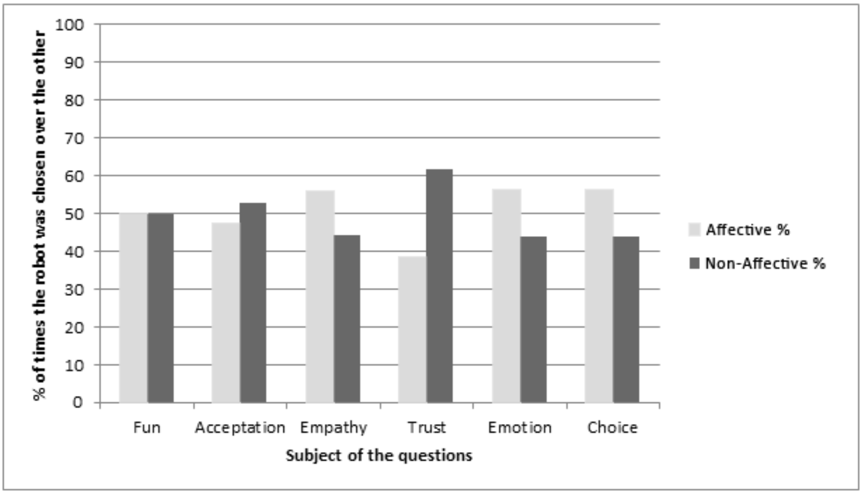



Figure 7.6: Forced choice between the two robots on several subjects.

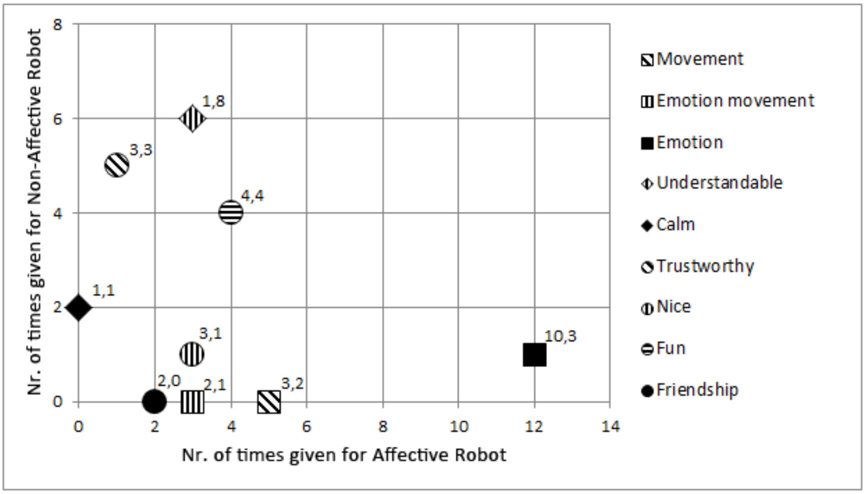



Figure 7.7: The number of times certain arguments were given as reason to choose one of the robots over the other. The coordinates represent the number of times an argument was given by a child who eventually chose the affective robot (X), or the non-affective robot (Y) as overall preferred.

## 7.4. Discussion and Conclusion

### 7.4.1. Behaviour of the children

When looking at expressiveness scores for children in Figure 7.4 we can clearly see that children show more expressions when interacting with an affective robot than with a non-affective robot. Moreover, we also see that children behave more positively in their expressions with an affective robot than with a non-affective robot. We can therefore state that when a robot displays adaptive expressions of emotion and gesture, children will also show more, and more positive expressions. Although there are very large differences in weighed expression frequencies between children, the affective robot tends to incite more smiles, more laughs, etc. from children. We already know that expressions from one interaction partner elicit expressions from the other in human-human interaction [102, 157]. From the fact that children show more expressions with a robot showing adaptive emotions than with a non-affective robot, we can conclude that this is also the case in robot-child interaction. This is relevant as it suggests that children interpret robot emotion in the same way as human emotions. It also means that it is possible to influence the behaviour of children by adapting the behaviour of the robot they interact with. As children showed more positive expressions with an affective robot, we can also state that children enjoy themselves more with a robot which shows adaptive emotion expressions and gestures than with a robot which does not.

### 7.4.2. Opinions of the children



The second dependent variable tested were the subjective opinions of the children. Through questionnaires, we tested if a robot adaptively expressing emotions and gestures elicits different opinions from children than a robot which does not. Looking at the results, we first see that the children are very positive about both robots, they clearly enjoy playing with robots. When asking the children for their opinions of each robot, no significant differences can be found between the robot using the model for adaptive emotion and gesture expression and the robot which did not. One of the possible reasons for this result is that there was a ceiling effect, indicated by the high opinions the children had of both robots. It is possible to make some suggestions about preferences when combining the data from the final questionnaire with the motivations given to the answers. Interesting from Figure 7.6 is that although the affective robot scores higher on empathy, emotion and general preference, the non-affective scores higher in acceptation and trust. Figure 7.7 shows an overview of the motivations for choosing either the affective or the non-affective robot over the other, for any of the questions. Looking at these reasons, we see that children particularly like the fact that the affective robot showed its emotions and that it moved more. They also thought this robot was fun and nice and they felt friendship. These reasons are given most often for the questions about fun, empathy and emotion. For the non-affective robot, the strongest argument for choosing it was that it was easier to understand. This can be explained by the fact that this robot had not fluctuations in the pitch of its voice. The fact that this robot moved less might also have contributed, as this leads to less signals to be processed by the child. The children also noted that they found this robot more trustworthy. Additionally, they liked that it was calm, and thought it was fun. All these reasons were given most to the questions about fun, acceptation and trust. We can take these motivations as evidence that the fact that the affective robot scored higher on empathy and emotion and the non-affective robot higher on acceptance and trust is not entirely due to chance. It seems there is some reason to believe that an affective robot increases empathy, but decreases acceptance and trust. Looking at the coordinates for the motivation of emotion, we see that 10 out of 13 children who gave the emotion argument also preferred the affective robot in the end. When asked which robot they thought nicer, one girl motivated her choice for the affective robot with *She showed her feelings and because of this I felt a stronger friendship*. This motivation gives a very clear statement of the positive effect showing emotion can have on robot-child interaction. There is also a downside to the expressive behaviour, however. Especially the voice of the affective robot has proven to make the speech of the robot harder to understand. The questionnaires show that it is very important for children to have a robot which they can understand well. Considering the coordinates with the motivation of understandability, we see that 8 out of 9 children who gave *easier to understand* as reason to choose a robot preferred the non-affective robot in the end. We can conclude that intelligibility is more important to children than emotion when it comes to a robot‘s voice. Noticeable is that a recent study using the same voice adaptations found no effect on understandability. [249]. As the only difference with this study was that the voice of the robot was constant, we can conclude that the fluctuations in voice might be a bigger problem than that the voice was too high or fast.

### 7.4.3. CONCLUSION

In an experiment with children we have shown that children display more expressions when interacting with a robot which displays emotion and adapts its expressions to the child than with a robot which does not. From this, we can conclude that we can influence the expressive behaviour of children by adapting the expressive behaviour of their robotic interaction partner. Moreover, as children showed more positive expressions with an affective robot, we can also state that children enjoy themselves more with a robot which shows adaptive emotion expressions and gestures than with a robot which does not. Data also shows that children particularly like it if a robot shows emotion through movement, while showing emotion through voice has the negative effect of reducing intelligibility. Although much is still unclear, we believe this work provides a first insight into the relation between adaptive emotion expression and the bond between robot and child.

7

## 7.5. ACKNOWLEDGMENTS

This work is funded by the European Union FP7 ALIZ-E project (grant number 248116). Furthermore the authors would like to thank the teachers and the children of ‘Lange Voren‘ (the school) for their participation in this study.

# 8 | A Remote Social Robot to Motivate and Support Diabetic Children in Keeping a Diary

**Abstract** *Children with diabetes can benefit from keeping a diary, but seldom keep one. Within the European ALIZ-E project a robot companion is being developed that, among other things, will be able to support and motivate diabetic children to keep a diary. This paper discusses the study of a robot supporting the use of an online diary. Diabetic children kept an online diary for two weeks, both with and without remote support from the robot via webcam. The effect of the robot was studied on children's use of the diary and their relationship with the robot. Results show that children shared significantly more personal experiences in their diaries when they were interacting with the robot. Furthermore, they greatly enjoyed working with the robot and came to see it as a helpful and supportive friend.*

## 8.1. Introduction

It is estimated that worldwide more than 490 thousand children between the ages of 0-14 suffer from diabetes mellitus type 1, and this number continues to rise explosively [109]. Diabetes is a chronic illness that impacts a child's life in almost every aspect. Effective self-management of diabetes is complex and involves many varied activities related to dosing insulin, monitoring metabolic control and regulating diet and exercise to name just a few. This can be especially challenging for children that are diagnosed at a very young age. They frequently have trouble coping with their diabetes [263], and are at risk for developing depression, anxiety disorders or eating disorders [99, 135, 225]. Health care providers advise children to keep a diary to monitor their health and how they feel on a daily basis so that appropriate treatment adjustments can be made [209]. Keeping a diary can significantly contribute to the quality of life of children with diabetes. A diary provides insight into patterns between blood glucose values and daily activities so that the child can better manage his/her diabetes. But despite this advantage of keeping a diary, children rarely take the time to do it or they do not see the value of keeping a diary. Digital diaries have been found to yield better compliance and accuracy in diary recording compared to paper diaries [185], but the problem of motivating children to start using these diaries still exists.

The EU-funded ALIZ-E project www.aliz-e.org aims to develop a social robot to support chronically ill children in their self-management. The scenario in which the robot is being tested is based on a medical setting where children recently diagnosed with diabetes spend one or two weeks in the hospital. During this time, the child and its parents are intensively trained and educated to manage the child's diabetes. But of course self-

Figure 8.1: NAO robot by Aldebaran Robotics

management of diabetes does not end there. It is an active and lifelong process that involves shifting and sharing responsibility for care tasks and decision-making between parents and child. The ALIZ-E project focuses on the potential role of a robot (Figure 8.1) in this process as an educator, motivator and companion that remains interesting to the child on the long term [1].

This paper describes an experiment which aims to improve children's diary adherence by means of a social robot. We created a scenario in which diabetic children keep a diary from home for two weeks, intermittently interacting with the robot via video-conferencing software. We studied the effect of the robot on the children's diary adherence, their engagement in the activity and their relationship with the robot.



## 8.2. Related work

Social robots are increasingly designed to be our pets, assistants, teachers and even emotional companions [141, 182]. They can positively affect people's motivation and compliance in areas such as education, health and well-being [50, 30]. The effectiveness of a social robot as a motivator largely depends on its ability to persuade its user. Persuasion is defined as "an attempt to shape, reinforce, or change behaviors, feelings, or thoughts about an issue, object or action" [90]. In this study, the issue we are addressing is the low diary adherence of children with diabetes. The robot attempts to persuade the child to keep a diary by using a variety of means that have been proven to be effective in human-robot interactions. Directly coercing the child into better keeping a diary is not an option for two reasons: first, coercion often has the opposite effect and causes the child to rebel even more [139]; and second, from a practical standpoint it is impossible to uphold this approach because the parents will not always be around to check up on their child. Rather than making active attempts to change the child's attitude and behavior, the focus should be on positively reinforcing the child's actions and utilizing the potential bond between the robot and the child as an incentive for the child to keep a diary.

To cultivate long-term relationships with users, the robot needs to engage in social behavior and dialogue [12]. Social dialogue includes things like greetings, chatting about general topics like the weather, and exchanging personal preferences [104]. Self-disclosure and empathy are known to greatly contribute to the closeness between conversational participants [2, 202, 163]. This effect is stronger with a physically embodied robot than with a virtual agent. There are many other factors we could consider which have the potential to positively influence children's diary adherence and long-term relationship with the robot. However, we feel that a robot which exhibits self-disclosure, empathy and physical embodiment provides a good starting point for this formative evaluation. These aspects can easily be implemented into the robot's dialogue and behavior.

### 8.2.1. SELF-DISCLOSURE

Studies have shown that self-disclosure plays a central role in the development and maintenance of relationships as well as psychological well-being [2, 56, 114, 140]. Self-disclosure is defined as "sharing information with others that they would not normally know or discover" [34]. Once a person engages in self-disclosure, it is implicitly expected that the other conversational partner will also disclose information (norm of reciprocity). Self-disclosure has been linked to a person's likeability [56]. People who engage in intimate disclosures tend to be liked and trusted more than people who disclose at lower levels. People tend to also like robots better when they disclose affective rather than task-related information in collaborative tasks [228]. Mutual affective self-disclosure between the child and the robot can contribute to the depth and quality of their interaction, and ultimately their relationship. Self-disclosure also contributes to the diabetes self-management of the child. Sharing daily experiences in the diary and reflecting on this information can help the child gain insight into how they can better manage or cope with their diabetes and ultimately improve their quality of life. In order to encourage the child to disclose information, we propose to utilize the reciprocal nature of self-disclosure and have the robot frequently disclose information about itself in order to encourage the child to do the same.



### 8.2.2. EMPATHY

Empathy plays a key role in patient-centered therapy, because it implies the apprehension of another's inner world and a joint understanding of emotions [237]. One of the most comprehensive definitions of empathy is by Davis [67], who defined it as "the capacity to take the role of the other, to adopt alternative perspectives vis-à-vis oneself and to understand the other's emotional reactions in consort with the context to the point of executing bodily movements resembling the other's". Robots cannot feel empathy, but they can emulate it in their behavior, for example by:

1. Showing empathic concern for others;
2. Taking the perspective of another;
3. Emotionally identifying with fantasy characters in books, films, etc.;
4. Expressing negative feelings in response to the distress of others.

We propose to incorporate empathy in the robot's behavior by showing sympathy or concern when the child says to be feeling down, and by reacting positively when the child is in good spirits. The robot also expresses its concern in asking the child how he or she handles certain issues related to diabetes (i.e. fear of exercising or pricking insulin in public).

### 8.2.3. Embodiment

Social robots do not necessarily need a physical body to interact with their users. Their tasks can often be performed just as well by a virtual 3D representation or avatar of the same robot, which costs considerably less and is much more robust. But having a physical form does offer substantial benefits compared to virtual robots. Embodied robots (robots with a physical presence) are more appealing and perceptive of the world around them than non-embodied robots [258]. Participants' impression of the robot's watchfulness, helpfulness, and enjoyableness is significantly affected by embodiment. In a study on a socially assistive robot exercise coach for the elderly [84], participants strongly preferred a physically embodied assistive robot over the virtually simulated one. The interaction with the physical robot was rated as more enjoyable and useful. Physical embodiment has also been found to evoke a higher degree of user engagement and presence [77].

The positive effects of having a physical body were shown to still be prevalent when the robot is shown remotely via a camera. The social presence of a remote physical robot was almost the same as a robot that was located physically near the user [194]. A remote projected robot and a physically present robot were found to be equally engaging and elicited equal disclosure from the user.

The ALIZ-E project develops a robot for long-term interaction with chronically ill children undergoing treatment. It is important that the children interact with the same robot throughout this entire period to provide a consistent experience. They interact first with the physical robot in the hospital, and later continue with the same robot shown remotely from home.



### 8.2.4. Hypotheses

The bond between the robot and the child can serve as an incentive for the child to keep a diary. To this end, the robot engages in self-disclosing behavior and encourages the child to do the same. It also shows emulated empathy and concern for the child. And because people generally find interaction with a (remote) physical robot the most rewarding experience, we choose to study the effect of the robot on the child's self-disclosure in the diary, engagement in the activity, and bonding with the robot. This leads us to the following hypotheses:

1. Adherence: the robot encourages the child to self-disclose more information in their diaries.
2. Engagement: the robot has a positive effect on the child's engagement in the activity.
3. Bonding: the robot conveys a sense of trust and understanding of the child.

In order to test these hypotheses an experiment in a real-world setting was conducted.

## 8.3. IMPLEMENTATION

### 8.3.1. ROBOT

The NAO robot (developed by Aldebaran Robotics) was used in this experiment. NAO is well-suited for interaction with children largely due to its friendly childlike appearance. Although its face lacks the capability to display emotions, the robot is able to show a wide array of emotions through its body language. The expressions used in this experiment were idle behavior and emotional expressions that were pretested in [55]. The authors found that the recognition rates for these basic emotions (e.g. happiness, sadness and surprise) were relatively high, between 68% and 99% accuracy. Idle behavior consisted of small body movements such as moving the head and hand positions while speaking or waiting for the user to answer. The robot in the experiment was given the unisex name "Charlie" in order to appeal to both boys and girls. A background story was written for the robot to answer basic questions about its likes and preferences, and about its reason for participating in the experiment. In the story Charlie is a hospital care robot in training and it hopes the child can help it learn more about diabetes by keeping a diary together. When the robot engages in self-disclosure, it does so keeping this background story in mind. The robot has an inquisitive character, and regularly asks the child about what it is like having (and coping with) diabetes.

### 8.3.2. DIALOGUE MODEL

A dialogue model was developed that structured the robot dialogue in an orderly fashion. The model consisted of two parts: the diary-related dialogue, and the interpersonal dialogue.

1. Diary-related dialogue: any dialogue related to the task of keeping a diary. This includes the login process, explaining of diary sections and filling in the diary itself. This part of the dialogue was very structured, and did not allow for much flexibility other than the child choosing what section he/she wanted to start with. Essentially this part of the dialogue was the same for every child.
2. Interpersonal dialogue (or 'small talk'): any dialogue not directly related to the task. There were two different types of small talk.
    - Self-disclosure small talk was used for the robot to share information about itself and optionally ask the child to do the same. For example, the robot told the child about its favorite pets, and then asked the child if he/she has any pets.
    - Diabetes-related small talk was used for the robot to talk about what it learned in school about diabetes (e.g. doing sports when having diabetes) and ask the child some basic questions about this topic (e.g. "How do you handle your diabetes when you do sports?"), and if he/she has any fears related to his/her diabetes (e.g. fear to prick blood glucose, or fear to exercise).

The small talk was used to enrich the dialogue between the robot and the child, with the goal of creating a bond between them.

### 8.3.3. Wizard of Oz

The main deliverable of the ALIZ-E project is an integrated and autonomous system comprising different modules. But because this system is developed incrementally and not all features have fully matured, some of the robot's functionality is simulated. The participants interacted with the robot which they believed to be autonomous, but which was actually controlled by the experiment leader in a Wizard of Oz (WoOz) setup. The NAO robot was capable of performing some actions autonomously, such as movements while speaking, 'blinking' of the eyes by switching LEDs on and off, and speaking (text-to-speech). Other actions could not be performed autonomously and needed human intervention. The experiment leader interpreted the user input (speech, gestures, and actions) to respond to the user by choosing the relevant remarks from the dialogue model. Certain movements such as cheering or nodding were initiated by the experiment leader by clicking the corresponding button in the WoOz interface. Updating the user model was also done by hand when the user provided new information. The dialogue was scripted in the dialogue model. In rare incidents, the experiment leader could type a response and have the robot say it via the text-to-speech module. The timing of interpersonal dialogue was largely up to the experiment leader. She could choose from one of the pre-scripted small talk phrases whenever a related topic was mentioned, or when there was a moment of silence.

### 8.3.4. Diary

The study required a diary that allowed the children to not only record their measured values, but that also had room for them to express how they feel. This allows the robot to respond empathetically to the emotional content of the diary. To this end, we chose to adapt an online diary for use in our experiment. Mijn Zorgpagina (literally 'my care page') is an initiative from Diabetesvereniging Nederland (DVN), the Dutch Association for Diabetes. In cooperation with DVN there were extra sections added to the diary to record the child's emotional well-being. The diary consisted of three different sections (see Figure 8.2): the 'values' section which consisted of different 'lines' for the recording of blood glucose values, carbohydrate intake, insulin doses and exercise; the 'emotions' section in which the child could rate his/her appetite, energy and mood levels using three sliders; and finally the 'daily activities' section which was a text box in which the children could write anything they wanted to share about their day.



In Figure 8.2, the video conferencing software (TeamViewer) we used is shown on the right-hand side of the screenshot (face of the child is blurred for privacy reasons). We used TeamViewer to enable desktop-sharing and communication via webcam and audio. Although the children could log into the diary from anywhere, we chose to provide each of them with a laptop that was stripped down to the bare minimum. The main reason for this was that we wanted to make sure all diaries were accessed from the same platform so that there would be no technical difficulties on their end. The children's accounts were set up in a way so that they would automatically log in when there was an internet connection. The child could then invite Charlie for a session. The sessions were all scheduled in advance. If they encountered any issues while attempting to log in or connecting to Charlie, the children could refer back to the child-friendly user manual that we provided or contact the experiment leader.


Figure 8.2: Screen as seen by the child showing diary and video conferencing software

## 8.4. METHODS

### 8.4.1. PARTICIPANTS

Six children affiliated with Dutch hospital Rivierenland in Tiel participated in the formative evaluation. They were recruited with the help of a diabetes nurse who informed the children and their parents of the experiment. Interested parents were contacted and they received further information. They could then decide whether they wanted their child to participate. The group consisted of two girls and four boys, aged 9-12 (M=10.8, SD=1.3). All children were diagnosed with diabetes mellitus type 1. On average, they had diabetes for six years. None of the children had any prior experience in keeping a digital diary.



### 8.4.2. EXPERIMENT DESIGN AND PROCEDURE

One week before the start of the experiment, we visited each of the children individually at home to introduce them to the robot. We explained the goal of the research and there was an opportunity to ask questions. The robot then introduced itself and asked the child to do the same. After that, we scheduled the sessions with the robot for the following two weeks. Children who forgot their appointments with Charlie were called to remind them. We feared that without reminders, children would forget their study participation, at the cost of valuable data. The goal of this experiment was gathering knowledge on the target group and their use of an online diary with or without the robot. As such the experiment can be seen as a formative evaluation. The number of participants was small and there was only limited time available for interaction with the robot. Therefore we decided to do a within-subjects design; this entails that all children kept a diary both with and without the robot. They did this for a total of 11 days. This allowed us to clearly see the differences in diary use for the same children in both conditions. We made two groups of three participants to spread the workload. This means that half of the children started on day 1 and finished on day 11, whereas the other half started on day 2 and finished on day 12. The sessions with and without the robot alternated every day. This excluded the weekends, which were always without the robot because access to the facility where the robot was stored was not possible. The children participated while in the comfort of their own homes using video conferencing software to contact the robot.

### 8.4.3. Measures

Below we present the metrics used for each of the three hypotheses.

1. Adherence: In this context adherence is defined as the extent to which children keep an accurate account of their values, mood and activities in their diaries on time (meaning on the same day the values were measured). If the robot has a positive effect on diary adherence, we expect to see 1) more completed diary entries with the robot, and 2) richer diary content (experience sharing/self-disclosure). We determine whether this is true by logging children's use of the diary and comparing the contents of the diary between sessions with and without support from the robot. Furthermore, we measure likeability and trust, because literature found these to be contributing factors for self-disclosure.
2. Engagement: An engaged child typically spends more time and is focused on the activity, and takes an active role in the activity. Comparing the time spent on the diaries between sessions with and without the robot can tell us what the influence of the robot on engagement is. Active participation and attention are measured by the experiment leader by observing the child's interaction with the robot via the webcam. A child is thought to take an active role in the conversation whenever he/she asks or tells the robot something of his/her own accord instead of waiting for the robot to ask something first. We can determine attention by observing the child's gaze direction. When the child looks away from the diary and the robot, he/she is distracted.
3. Bonding: A bond between two conversational partners is characterized by emotions such as affection and trust. To investigate to what extent interaction with the robot created a bond, we used a post-condition questionnaire which focused on the degree of relatedness the children experienced with the robot. When a child trusts the robot, believes it has feelings, and thinks that it can truly understand him/her, the feeling of relatedness is strong.

### 8.4.4. Instruments

The children were asked to complete a total of 7 questionnaires. The questions were phrased in simple terms and emoticons were added to Likert-scale questions for clarification (cf. [198]). The pre-condition questionnaire was answered prior to the experiment, and inquired about demographic information, interests and expectations. We used this questionnaire to learn more about our participants and whether they understood the experiment correctly. After each interaction session (5x) with the robot, the children were again asked to fill in a questionnaire. This questionnaire only contained two questions which inquired about the children's ratings of the robot and the diary. This data can be used to determine whether there are any changes over time. Finally, the children answered a post-condition questionnaire after all the sessions were completed. In this questionnaire we asked them about their overall judgment of the robot and the diary, the degree of relatedness with the robot, and feelings towards the robot. These data were mostly used to determine the extent of the relationship between the child and the robot.

Furthermore, we also made use of three different types of logs. The diary logs were provided by MijnZorgpagina and provided insight into the children's use of the diary on days when the robot was not present. The WoOz logs were automatically generated whenever a command was sent to the robot. We used these to determine use of the diary with the robot. Lastly, the experimenter leader's observation logs were used to note any changes in experience sharing, active participation and attention of the children during the sessions. Miscellaneous observations were also noted in these logs.

Initially we had planned to record all interaction sessions on video to carefully note any changes in active participation and attention. However, due to a technical problem, the video conferencing software crashed when the recording software was started. We came to rely on the experiment leader's observations for data regarding the children's engagement.

### 8.4.5. SETUP

Figure 9.3 provides a schematic overview of the experimental setup. To mimic the natural sitting position of humans during video calls, the robot (1) was seated in a chair during the interaction sessions. The webcam (2) was aimed at the robot from an angle slightly higher than eye level because we wanted the children to believe it was operating its own laptop. During the interaction sessions, the experiment leader (3) was seated out of view. She was in control of the robot's actions using a laptop and an extra screen (4). She was also able to see the screen and webcam of the child using the video conferencing software (5) on a second laptop. The experiment leader made as little noise as possible so as not to alert the child to her presence. To this end, we used a silent mouse (6) that lacked the "click" sound when a button was pressed.



## 8.5. RESULTS

All six children completed the full experiment and the associated questionnaires. Diary use prior to the experiment was found to be very low; none of the children kept a diary on a daily basis. Five out of six children used an insulin pump and three of them indicated they did not keep a diabetes diary (anymore) because they were able to read out values from their pumps if needed. Only one child said that he currently kept a diary all by himself, but after enquiring about his diary use he admitted to only doing it once every three months when it was required for a hospital consultation. All of the children said to be at least moderately interested in technological advances such as robots or gadgets.

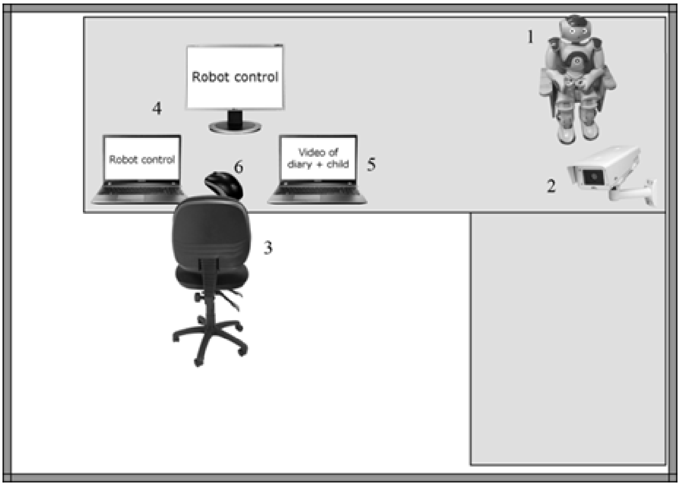



Figure 8.3: Experimental setup

### 8.5.1. ADHERENCE

The log data from the diary revealed that not all children completed their diaries on time (meaning on the same day the values were measured). When a diary entry was not completed on time, it was usually completed the day after. When they were not interacting with the robot, two of the children sometimes neglected to keep their diary. One of them missed all (6) entries, the other left only 2 entries uncompleted. Three children logged on multiple times (1, 2 and 8 times) to supplement values as new information became available. The differences between the conditions in filling out the diary were not significant due to the high interpersonal variation. Figure 8.4 shows the differences in diary completion between conditions in percentages of the total actions performed. For example, all children fully completed the diary in the robot condition, but some of them also supplemented the diary with extra information when it became available in this condition. The total of all these actions adds up to 100 percent.

We compared the amount of characters used in the daily activity logs between the days with the robot and without the robot as an indicator for experience sharing. It was found that children wrote significantly more in their logs when they were interacting with the robot (M=83.10, SD=43.96) than without the robot (M=36.8, SD=54.31); t(5)=4.13, p=.009; but there were clear differences between participants. In Table 8.1 it is evident that one participant (5) always filled in high quantities of data, whereas some others filled in less on days without the robot (2 and 3), or nothing at all when the robot was absent (1 and 4).

Overall, the children rated the diary positive throughout the experiment. The diary received an average rating of 3.9 on a scale from 1 to 5.

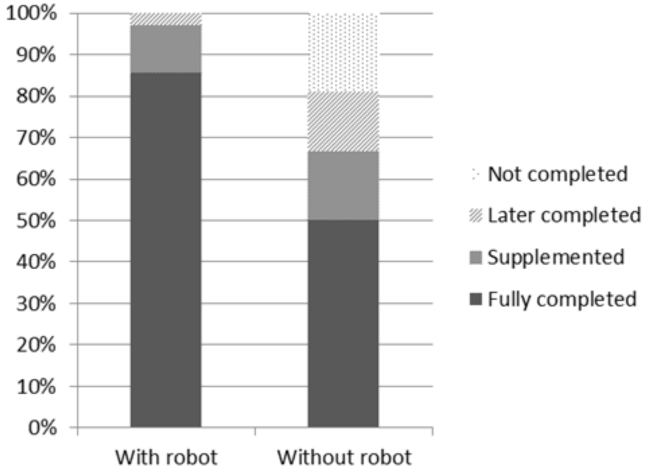



Figure 8.4: Diary actions performed with and without robot

### 8.5.2. ENGAGEMENT

We were unable to validate whether there was a difference in the time spent on the diary due to a gap in the log data. The time spent on the daily activities log was not recorded on the days without the robot, so the data from both conditions were unequal. However, we did see how the time spent on the diary changed over time. The first session with the robot typically took between 10 to 15 minutes, as did the second session at home without the robot. In later sessions, the time taken dropped and stabilized around 5-6 minutes for both conditions. This was due to the fact that by then the children knew their way around the diary and became more efficient in completing their task.

The younger children (aged 9) (N=2) were much more open and talkative than the older children (11-12) (N=4). Their answers were longer and their sessions typically took 1-2 minutes longer than those with the older children. They were more likely to ask the robot about its personal life than the other children were. Older children appeared to be less interested in the robot's life and did not ask it as many questions. They usually answered the robot with a simple "yes" or "no". There was no discernible difference between genders in interaction styles with the robot.

The children were very patient when they had to wait a while for Charlie to answer. This was apparent in the way that the children remained still and focused even when the robot did not do anything for a while. Overall the children were focused. They did not allow themselves to be distracted by background noises or siblings and focused solely on the diary and to a slightly lesser extent on the robot. When the robot spoke, their attention visibly shifted to the robot for a short while, but then quickly back to the diary. This was apparent from their gaze direction which was visible on webcam.

8

| Participant | With robot | Without robot |
|---|---|---|
| 1 | 60.4 | 0 |
| 2 | 36.6 | 15.7 |
| 3 | 86.4 | 13.5 |
| 4 | 78.8 | 0 |
| 5 | 165.6 | 141 |
| 6 | 70.8 | 50.8 |
| **Mean** | **83 (SD=40)** | **37 (SD=50)** |

Table 8.1: Average amount of characters in daily activities logs

### 8.5.3. Bonding

In the post-condition questionnaire, the children were asked to agree or disagree with a number of statements on a scale of 1 to 5, where 1 meant 'completely disagree' and 5 meant 'completely agree'. Figure 8.5 summarizes the results from the post-condition questionnaire.

The robot received high ratings on all questions. Especially high (average rating above 4) was the rating of the robot's trustworthiness and its human-like behavior. Furthermore, when we asked the children what they liked about Charlie, they said they liked the fact that it asked questions about the things they did that day, as well as share with them his own daily activities.

We also asked the children how they viewed the robot. Five out of six children chose the option 'Friend'. One child said that Charlie was more like a peer, and another said that it was a device/robot to him (but also a friend). The children rated the robot after every session. On average, the robot received an average rating of 4.2 on a scale of 1 (lowest) through 5 (highest).

Furthermore, we noticed the children often smiled or laughed at the things the robot said, and they waved back to the robot when it greeted them. Overall they were very friendly towards it. The most positive reactions were elicited when the robot immediately responded to the ratings for the emotions and daily activities. The children seemed to think that it really understood them. The level of detail in the questions some of the children asked was high, regardless of their age. For example, when Charlie said that it had played soccer with friends, they would ask it about the final score and who the goal keeper was. One child even asked about Charlie's birthday, and when it said that was February 2nd, she wished it a happy birthday in advance, much like humans would do to each other.



## 8.6. Discussion

### 8.6.1. Self-disclosure and diary use

Children completed their online diary more often when they were keeping it with the robot than when they were keeping it alone. This was to be expected, as the children were actively notified by the experiment leader on the days they were scheduled to fill in the diary with the robot. However, after the children had interacted with the robot, they also logged in the diary later that day to supplement the diary, on their own initiative.

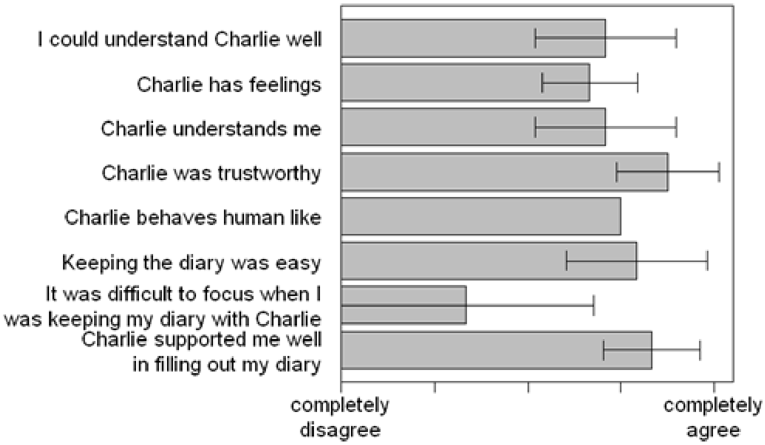



Figure 8.5: Post-condition questionnaire ratings (N=6)

This suggests that the robot was motivational by itself. A critical note with regard to the diary completion is that we called the children who forgot their appointments with Charlie. Considering the small sample size, we felt it was more important to have as many interaction sessions as possible to learn how the robot's self-disclosure influenced diary adherence. Moreover, children consistently filled in more information in their daily activities logs (the bottom part of the diary) when they were keeping their diary with the robot, than when they were keeping it by themselves. Thus, regardless of whether or not the children were reminded of their appointments, when they did interact with the robot, they filled in much more information. The robot's self-disclosure appeared to be a strong motivator for the child to disclose information in return. We can confirm the hypothesis that the robot encourages the child to share more information in his/her diary.



### 8.6.2. EMBODIMENT AND CHILDREN'S ENGAGEMENT

From the log data we gathered that the time taken to fill in the diary dramatically decreased after the first two sessions. This can be explained by the fact that during the first session the children listened to the robot explaining every diary section, while in subsequent interactions this explanation was optionally accessible (and rarely asked for). Due to the failed recordings of the sessions, the data on children's interaction behavior was elicited from notes made by the experiment leader during the study. These notes were taken systematically and helped explain some of the results, but unfortunately they could not be statistically validated. The observations showed a strong difference in active participation between the younger children, who were very open and talkative, and the older children who were less open and talkative. Possibly, the background story for Charlie was not convincing or appealing enough for the older participants in this study. This is in line with the commentary we received from an older child (13) during the pilot study we performed prior to the experiment ("I think the story is too childish for 11-12 year olds"). Overall, the experiment leader had the impression that the children paid close attention to the robot and the diary. They did not allow themselves to be distracted by background noises, siblings or parents. While some parents chose to supervise their child during the sessions, they did not appear to be a distracting factor. We did however see some mixed results in the answers on the question whether it was difficult to focus when they were keeping a diary with Charlie (in the post-condition questionnaire). But this might also be due to the slow internet connections of some children. Occasionally there were very large delays in the sound transmission, which made it difficult to understand what Charlie was saying. Since we did not have sufficient data on the children's engagement with the diary without the robot present, it is difficult to say whether the robot had a positive effect on the child's engagement in the activity. Therefore we cannot confirm the second hypothesis.

### 8.6.3. Robot's understanding and trust

Overall, the robot was received very well by the children. They greatly enjoyed interacting with Charlie. They believed it had feelings, was able to understand them, and that it behaved human-like, which are signs of empathy. However, we have to be careful in saying the conveyed sense of trust and understanding of the child was because of the robot (hypothesis 3), because occasionally part of the dialogue was typed by the experiment leader directly and converted to speech by the robot when there was no suitable answer in the dialogue model. Although most of the dialogue used was in fact pre-scripted, the comments that were made on the fly could have still somewhat skewed our results.

#### Future work

Future work in the area of robot-driven diary support could benefit from sentiment analysis, which refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information from source materials (i.e. the diary). Sentiment analysis would allow the robot to more accurately predict and interpret the emotional state of the child and choose the appropriate response. In our experiment, a human actor had to interpret the data and 'translate' this for the robot. Ultimately the goal of the ALIZ-E project is an integrated system which operates without human intervention. There still needs to be done a significant amount of work in the area of speech recognition and interpretation to achieve this goal.

This paper presented a first experiment in which we succeeded to include six children with diabetes, who interacted with the application over an extended period of time. Due to the small number of participants, a within-subjects design was applied. Based on the results, we will be able to get other hospitals interested and conduct a between-subjects experiment with more participants. Participation of diabetic children remains crucial, because children without diabetes do not have the intrinsic motivation or ability to fully complete a diabetes diary.

In the current experiment there was no explicit feedback about the relationship between blood glucose, mood and daily activities. In future work it would be interesting to add this type of feedback to the dialogue and to discover patterns together with the robot. We see opportunities for two follow-up experiments: (1) comparing children who keep a diary either with or without the robot, and (2) comparing their diary usage for a robot which does or does not exhibit self-disclosure, empathy and embodiment. Last but not least, the bonding effects of co-located and remote interactions should be tested: How the bonding transfers or evolves from co-located experiences with the robot in the hospital to remote interactions via webcam, and vice versa.

## 8.7. CONCLUSION

This paper investigated the contribution of a social robot to keeping an online diary together with diabetic children. A social robot can enhance the pleasure of the activity, and therefore the motivation of the child. Especially once the robot and the child really get to know each other, the child starts to consider the robot as a friend and he/she really opens up to it. Keeping an online diary together with a social robot can contribute to a better diary adherence. The robot utilizes aspects of physical embodiment, self-disclosure and empathy in its behavior and dialogue to achieve this goal. When we take into account that the diary use prior to this study was almost non-existent, this is a considerable improvement. Keeping a record of the values, emotional well-being and the daily activities allows the child to make meaningful inferences about the relationship between these three variables. This positively influences the child's self-reflection capabilities, which in turn contributes to his/her self-management. The robot could help the children overcome the initial hurdles of taking charge of their own diabetes self-management. The addition of a robot does not have to be detrimental to the children's feeling of independence, which becomes increasingly important as they reach puberty. This study provides a sound foundation for future research into robot characteristics and their effect on emotional support for chronically ill children.

## 8.8. ACKNOWLEDGMENTS

This work is funded by the EU FP7 ALIZ-E project (grant 248116). We would like to thank the diabetic children and medical staff from the ZRT hospital in Tiel (the Netherlands) and Diabetes Vereniging Nederland (DVN) for their cooperation.

# III

# SITUATED DESIGN RATIONALE FOR MULTIPLE USE CASES

This chapter contains:

- R. Looije., M.A. Neerincx, V. de Lange *Children's responses and opinion on three bots that motivate, educate and play*,in Journal of Physical Agents, 2(2), (pp.13-20), 2008. [146]
- R. Looije, M.A. Neerincx, J.K. Peters, O.A. Blanson Henkemans, *Integrating Robot Support Functions into Varied Activities at Returning Hospital Visits*, Int Journal of Social Robotics (2016) 8: 483. [149]

## Doing multiple activities

In this final Part two papers are discussed that look at the envisioned usage of the social robot. We established, that to keep engagement, the system should provide multiple activities which supports a feeling of autonomy. A selection of such activities with the robot has been evaluated in [146] and [149]. Figure III.1 shows the sDR of this part.

In this part we look at similar interactions as in Part II, but focused on the different activities where the robot could be part of. Because of the different activities that are available to the users in both papers there is a definite contribution to the method "Provide variation". In [146] this is extended by exhibiting social behavior. In assition, [149] actually integrates a selection of functions evaluated in previous experiments, and performs the evaluation in 'the wild'. The decisions to include a certain function or not are explained in the paper itself.

In the previous part we warned about taking results from one use case to another (e.g. self-disclosure by the robot during a game that is played for fun can have a different effect than self-disclosure during a diary task), in this part we are able to use functions over use cases and see how children react on this.

## Robots for autonomy

The chapters in this Part will answer the following two research questions:

1. chapter 9: Behaviors for the iCat to display different roles [146]
   - 1.1. Design question: How to create behaviors for a moderate expressive [253] iCat robot based on Motivational Interviewing [207] techniques?
   - 1.2. Hypothesis: Text, virtual and physical robot are for children, in an incremental order, increasingly motivating and educating. This can be explained by the incremental number of motivational interviewing techniques that can be implemented in the different interfaces.
2. chapter 10: Evaluating in the wild [149]
   - 2.1. Design question: What does experimentation in the wild add over controlled experiments that test isolated components of the robot one-by-one in a lab environment?
   - 2.2. Research question: Is the complete system is appreciated by children with diabetes, after multiple interactions, on the factors; autonomy, competence and relatedness.
   - 2.3. Research question: Does performing an experiment in the hospital with the real target users increases acceptation of all involved (children, parents and health care professionals)?

The overarching research question of this part is: What is the effect of different interfaces and multiple interactions and activities on the use of different roles, the performance and experienced enjoyment and bonding?

The two studies performed in the respective chapters yielded positive results. A clear added value of robots versus conventional text interfaces showed in chapter 9. This chapter also showed support for the different roles/activities a robot could engage in.

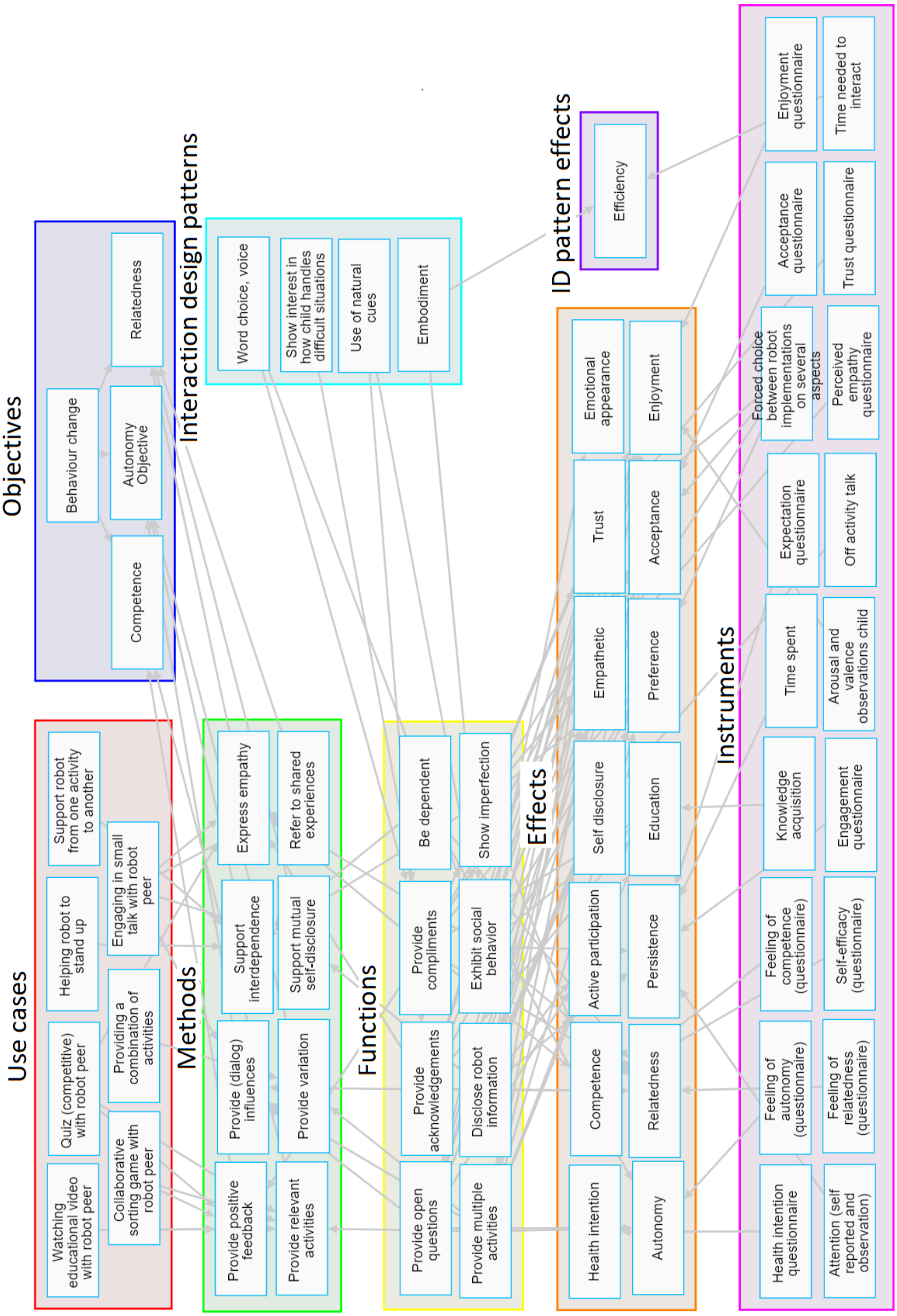



Figure III.1: situated Design Rationale of Part III

In the final experiment chapter 10 the efforts from the whole ALIZ-e project came together and resulted in a very interesting experiment that proved to have much added value. Unfortunately due to the large variation between children and software and hardware problems we could not make hard conclusions. On the other hand the observed interactions, next to conversations over a longer term with parents, health care professionals and the children, showed that the interaction with the robot was experienced positively and benefits could be envisioned and were experienced (e.g. lower threshold to discuss diabetes with health care professional).



Figure III.1 shows the sDR figure for these three studies. As in the previous parts the concatenated sDR supports the reasoning over the chapters. As the results of especially [149] were not clear we are not able to infer over the results, but it does show that it is possible to visualize a complex experiment like this in one picture showing the relations between objectives, expected effects and the implemented functions. It also shows the interrelations between these and therefore stresses the importance of using specific measures to decrease ambiguity.

# 9 | Children's responses and opinion on three bots that motivate, educate and play

**Abstract** *Social robots may help children in their daily health-care related activities, such as adherence to diet and exercises of diabetics. Based on a domain and literature study, we specified three support roles with corresponding bot behaviors: motivator, educator and buddy. These behaviors, such as showing attentiveness, could be implemented well in a physical character (the iCat robot), somewhat less well in a virtual character, and least well in a text interface. Twenty—eight to nine years old—children participated in a controlled experiment to evaluate the bots. They proved to value the support roles positively, in particular the buddy role. Objective and subjective data showed that they highly appreciated both the physical and virtual characters (more than the text interface). Furthermore, children proved to interact faster with the character than with the text interface. There is a clear added value of robots compared to conventional text interfaces.*

## 9.1. Introduction

Information and communication technology (ICT) in home, school and health settings has changed dramatically in the last two decades. For example for education it has been changing from one computer in a class that is hardly used, to computer usage by every school subject and the requirement to do homework on the computer. This use can be extended from homework tasks for school to physical exercise. These physical exercises might help to counter the increasing number of children suffering from obesity and diabetes. ICT technologies can thus aid in doing exercises [212, 24, 129, 96, 95, 243], giving social support [130, 136], and helping with lifestyle change [27, 145, 251, 28]. Research on persuasive technology [90] and affective computing [192] provides (partial) solutions, e.g. for the realization of social behavior, such as social talk and turn-taking [24, 129, 96, 95], and of empathic behavior, such as attentiveness and giving compliments [130, 136, 27, 243, 145, 28]. This research comprises supporting technologies that are more conventional text-based [243, 27, 28], and more innovative character-based virtual [212, 24] or physical [129, 96, 95, 130, 136, 145] "robots". The media equation [201] states that technology is higher appreciated when it exposes social behavior and is physically present. Consequently, one would expect that physical characters are appreciated more than virtual characters and text interfaces. This is confirmed in research comparing virtual with physical characters, as all results are in favor of the physical character [96, 8, 128, 257, 194]. In comparison with adults, children react to, and interact with, physical characters differently. Tanaka [235] found that children – after 27 lessons - interact with a physical character as if it was a peer instead of a toy. This can be caused by their tendency to heavily anthropomorphize the character. Draper [78] conducted research towards physical characters in the education of children. This research showed that a teacher teaches best, but that a physical character is better than a sound-tape with the lesson.

The paragraph above summarizes some research on persuasive technology, affective computing, virtual and physical characters. However, more research is needed for better understanding of the added value of robots compared to conventional text interfaces. First, there is a need for further theoretical foundation from psychology, pedagogy, persuasive technology and affective computing, to improve the development of a motivating and educating social companion. Second, there is a need for further empirical foundation, in which the different user interfaces are being evaluated in a comparative experiment with children.

In this paper, we address this by comparing a text interface, a virtual and a physical character that all implement the roles of educator, motivator, and (game)buddy as far as their dialogue and appearance characteristics allow for. Our general hypothesis is that a physical character is better at fulfilling these roles than a text interface and virtual character. We focus on the user experience [195]: how the children response to, and enjoy the interaction with the different interfaces.

## 9.2. Design of three bots for young diabetics

We chose the iCat from Philips (Figure 9.3a), in both physical and virtual form, to implement the behaviors for the concerning roles. This character was previously used in an experiment with older adults [145, 144]. During this experiment, participants evaluated five different interfaces: a text interface, a social and non-social virtual character, and a social and non-social physical character. User preference was measured for the different assistants on several factors, such as empathy, trust, and acceptance. The results indicated that socially intelligent characters are rated more empathetic than a text interface and a non social character. Moreover, the virtual character was appreciated more than the physical character both on the trustworthiness and the empathy dimensions [145]. Notwithstanding the positive results for the virtual character, half of the users indicated that they preferred the text interface while the other half preferred a social character. A possible explanation could be the anxiety that older adults have towards characters [222].

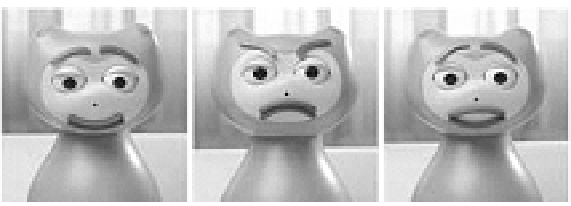


Figure 9.1: A happy, angry and surprised iCat

9

### 9.2.1. MEDIA EQUATION

People have the tendency to socialize information and communication technology [201], this is called the media equation. The more a device supports this tendency, the more people will like to use the technology. Furthermore, a physical character will have a greater social facilitation effect [96, 8, 128, 257] (i.e. people tend to perform simple tasks better in the presence of others [244]) than a virtual character [8]. Both the tendency to like social devices and the social facilitation effect support the idea that a social physical character is preferred as a personal assistant. Therefore, we distinguish three bots in this study:

1. Conventional text
2. Virtual robot (virtual iCat)
3. Physical robot (physical iCat)

### 9.2.2. DESIGN OF A PROTOTYPE FOR CHILDREN

The social characters and text interface developed for adults were taken as a starting point for the design of the prototype for children. The existing prototype was adapted for the use by children and made more automatic. We had to adapt the prototype because children ask for a different approach of both the design as well as the evaluation of the interface. During the design phase, special attention should be given to the different interests and cognitive abilities that children have in comparison with adults, which influence their interaction with the computer [49]. We looked specifically at cognitive, physical, and affective characteristics of children in the age group of 8-9. Children of this age are linguistically skilled and start performing several tasks independently. An example is diabetes where children start administering insulin and counting carbohydrates themselves.

Relating the cognitive development of children, interfaces should be *visually oriented* with not too much text and, just as for adults, *immediate feedback* is needed to keep the interaction natural and non-irritating. In relation to the physical development, Chiason and Gutwin [49] propose that interfaces for children should be tangible, such as the physical iCats, and that interfaces need not be cuddly in order to be engaging. Finally, research in affective computing shows that children like to have the possibility to be in control of the interaction with technology and that children stay engaged and motivated by providing them with occasional *entertaining events* [49]. Engagement and motivation can be stimulated by challenging and fun games, e.g. implemented in a (game) buddy [243, 117]. The (game) buddy ensures that users keep using the assistant, because it is fun [195].

In the evaluation phase, subjective measures are often used to get the opinion of the user about the tested interface. The opinion of children is important, because adults do not always understand what children want and why [199]. Doing a survey with young children is not easy. The children should be able to interpret all the questions correctly and make a considered choice between the answers. Another problem for the analysis is that children have the tendency to have extreme opinions on all the products they rate [199].

### 9.2.3. DIABETIC CHILDREN

In previous research a domain analysis of adults with diabetes was performed. We extended this analysis to the domain of children with diabetes, using diabetes as a case study. A diabetic nurse, play therapist, a patient who acquired diabetes on a young age and a game developer were interviewed. This analysis yielded insights in the differences and similarities between adults and children with diabetes and their computer technology usage. Both adults and children have a need for an educator who teaches them more about diabetes, because chronically ill have little knowledge about their disease [28] and therefore do not understand why they have to comply with certain advices. Furthermore, there is a need for a buddy that is a companion in coping with the disease. In addition, children were in need of help for counting carbohydrates, and one that helps keeping track of time to take their medication in time. An important remark was that the use of the device should be fun and challenging to improve the engagement and motivation. Eventually, diabetic children could be one of the first "serious" users of the envisioned personal assistant. Eating, physical exercise, and their joint effect on energy consumption are important issues for such children, and, therefore, 'core' elements for our study on robot assistance.

## 9.3. DESIGN OF THREE ROLES FOR THE BOTS

Based on the knowledge we gathered about diabetic children and their needs, a scenario was developed that includes personal assistance. Based on the scenario we chose three roles to be implemented in the prototype: educator, motivator, and game buddy. An extra advantage of implementing the motivator and educator roles is that the results can be compared to the motivator and educator role in the experiment for a personal assistant for older adults [145]. That experiment showed that these roles are appreciated when implemented in a social robot. We implemented the roles in the same three bots as in [145]: a chatbot, a virtual, and a physical robot. In contrast to the chatbot, the robots have the possibility to express facial and voice emotions.

### 9.3.1. MOTIVATOR



Both the motivator and educator are based on the Motivational interviewing theory, which by means of questions tries to facilitate increase in knowledge on persons' behavior *and* disease – in our case diabetes - thereby increasing the motivation to change. A therapist who can apply motivational interviewing successfully should be: empathetic [205] and trustworthy [170]. Motivational Interviewing is successfully applied in a text-based personal assistant, the HealthBuddy§, for chronically ill [27, 251]. We divide the properties of motivational interviewing into two roles, the motivator and educator role. The motivator role implements the properties that are linked to *how* things are said and done while the educator role focuses on *what* is said and done. This means that the motivator role looks at ways to make the assistant appear empathetic and trustworthy.

To make the assistant look empathetic we could find some skills with related behaviors to implement. We implemented three behaviors for three skills; reflective listening, positive regard, and attentiveness. The *virtual* and *physical iCat* are able to implement behaviors for all three skills, while the *text interface* can only implement behaviors for

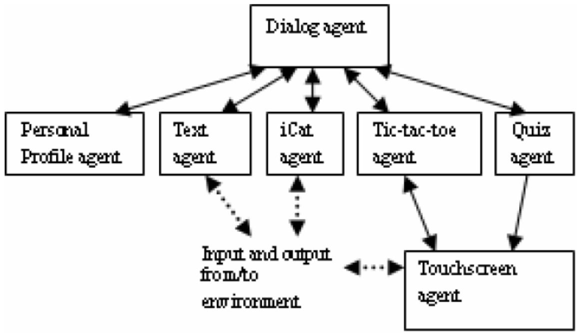



Figure 9.2: Agent structure

the positive regard skill.

Reflective listening behaviors that are implemented are: reacting positive or negative according to the event and asking questions when something is not understood. The behaviors that are implemented for positive regard are: give compliments when something is done correct and do not punish if something is done wrong. The behaviors for the last skill, attentiveness, are: look at the user, have an active listening expression, and sometimes nod.

It is very difficult to find behaviors that make an assistant look trustworthy; trust in an application is something that comes in time, but it can be stimulated. To enable trust, the dialog, mainly the form and content, can be made acceptable for the user. This can be done for example by taking the vocabulary of the user in account. Another way to receive trust, that the play therapist proposed, is to make the user comfortable (e.g. let the user play a game).

### 9.3.2. EDUCATOR

Motivational interviewing tries to increase the knowledge of a patient by educating the user. We implemented this in a quiz form that used educational videos on nutrition and/or exercise each followed by a multiple choice quiz question about the video to increase the knowledge of the user about the subject. The educator uses behaviors from the motivator to appear empathetic and trustworthy. It listens to what the user says, is happy when the user answers a question correctly, and just gives the reason for the correct answer when the answer is incorrect. The educator behavior was the same for the *physical* and *virtual iCat* and for the *text interface*.

### 9.3.3. GAME BUDDY

The game buddy role was chosen, because an assistant for children would definitely need a fun activity. Children need to stay engaged, and alongside of the serious tasks a personal assistant can offer them, some entertaining functionality is necessary.

A first prerequisite for the game buddy was to offer a familiar two player game that was not too difficult, did not take long, and was fun for a little while. In previous research

9

with the game of tic-tac-toe [253], children found it fun to play it with the iCat. Therefore, we decided to use tic-tac-toe in our prototype. Furthermore we based the personality of the game buddy on the personality that was preferred in the research of Verhaegh [253]: moderate expressive.

There was an algorithm that made sure that the level of the game was adapted to the user so that it became harder if the user won and easier if the user lost. The outcome of the previous game was stored in a user profile. We tried to keep the game challenging in this way.

The personal assistant in the game buddy role was empathetic (using the motivator behaviors, which were different for the *robots* and the *text interface,* see section III.A) towards the user; it gave compliments and was not over enthusiastic if it won a game. The personal assistant gave comments on the game; compliments ("nice move"), neutral remark ("now we are equal"), and congratulating remarks ("congratulations you won"). The comments were given taking three factors into account: Who made the last move, whether the situation is advantageous for the user, and if the game is in an end state.

Besides being complimentary the assistant was also attentive in the way that it asked the user if he/she would like to start, which symbol he/she preferred to use, and it looked at the game board when the attention of the user was there. Furthermore the assistant did not cheat, and left the user in control.

## 9.4. MULTI AGENT STRUCTURE

We implemented the prototype with the use of distributed agents that were in compliance with the FIPA standards [86]. The different roles were all implemented in their own agent so that the structure was modular. The modularity makes it possible to extend or adapt the system without changing the whole system. Furthermore, the use of agents makes the whole system easy distributable. Figure 9.2 gives an overview of the implemented agents. The agents are implemented in JADE.net [110] with the use of C#, because the communication framework was already implemented in C#.

9

The three different roles are implemented in different agents. The motivator is implemented in the dialogue agent (which is the central agent), deciding when what text and what expression should be used. The dialogue agent also poses the quiz questions and handles the answers. Secondly the tic-tac-toe agent implements the game buddy that decides when to do which move. Finally the quiz agent implements the educator role by starting up movies. The touch screen agent displays the movie and tic-tac-toe and sends the move of the user in tic-tac-toe back to the tic-tac-toe agent.

The text, touch-screen, and iCat agent receive and send information from and to the environment. The text agent represents the text interface, and the iCat agent represents

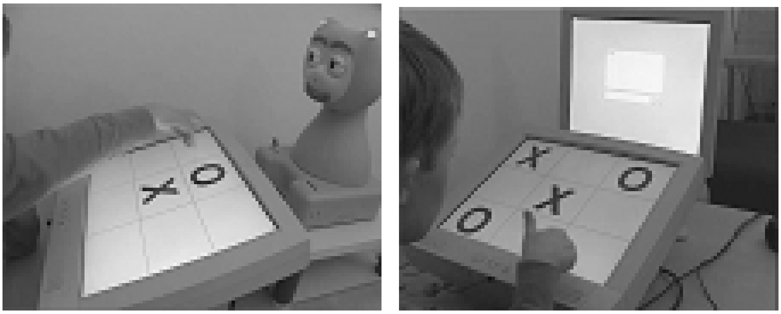
the iCat. Within the iCat agent, there is a module that handles the text input from the speech recognition that is performed by the experimenter. The last agent is the personal profile agent that holds information about the user, such as age, gender, lost and won games. This information can be used to adapt dialogue, game, and quiz.

| | | Question |
|---|---|---|
| Expectation | ifx1 | How nice do you think working with the robot/chatbot is going to be? |
| | ifx2 | What did you think of working with the robot/chatbot? |
| Engagement | ife1 | Would you like to use the robot/chatbot again? |
| | ife2 | Would you like to play another game with the robot/chatbot some time? |
| | ife3 | Would you like to play another quiz with the robot/chatbot some time? |
| | ife4 | Would you like to talk some more with the robot/chatbot some time? |

Table 9.1: Questions regarding fun

| Positive | Property | Negative | Property |
|---|---|---|---|
| Smiles | -Mouth angles directing upwards | Frowns | -Lowering the eyebrows |
| Laughing | -smile with unveiling of teeth | | |
| Concentration signs | -fingers in mouth<br>-tongue out | Signs of boredom | -Ear playing<br>-fiddling |
| Excitable bouncing | -Moving (slightly) back and forth in the vertical direction | Shrugs | -Moving shoulders quickly up and downwards |
| Positive vocalization | -Exclamations such as “cool”, “I like ..”, if made not directly towards the interface | Negative vocalization | -Exclamations such as “boring”, “don’t like”, if made not directly towards the interface |

Table 9.2: Positive and negative utterances that were counted



(a) (b)

Figure 9.3: Experimental setting

### 9.4.1. Wizard of Oz

The participants thought they were using a completely autonomous assistant, but the experimenter/wizard simulated the speech-to-text. The agents, text interface, and iCat were implemented in a way that the whole interaction between participant and personal assistant was autonomous (i.e., only the speech recognition was simulated via a person in another room, the so-called Wizard of Oz).

## 9.5. Evaluation

The three bots; chatbot, virtual robot, and physical robot, were implemented with the use of the predetermined roles and agents. After which they were evaluated. In this evaluation we tested if the participants thought of the bots as being empathetic, trustworthy, and fun, amongst others. Furthermore, we objectively measured positive and negative utterances and time spent at the interaction with the robot.

Based on literature about social actors and previous research our hypotheses were:

- (H1) The robots will be evaluated as more empathetic than the chatbot.
- (H2) Children will trust the physical robot most and the chatbot least.
- (H3) The physical robot is most attractive.
- (H4) The interaction will be faster with the robots.



### 9.5.1. Method

#### Participants

Twenty-four non-diabetic children took part in the experiment, that lasted around 1 hour and quarter, for which they were rewarded with a book token. The data of twenty children was usable (due to incompleteness and a child with a neuro-developmental disorder). The twenty children were all third

graders (i.e., fifth group of the primary school in the Netherlands), aged 8-9 (M age = 8.40, SD = 0.50).

*Setting:* The experiment was conducted in a room that resembled a living room. There was a table, on which touch-screen and iCat stood, or instead of the iCat a keyboard and computer screen stood (Figure 9.3).

### EXPERIMENTAL DESIGN

A within subject design was used for iCat vs. text interface, while there was a between subject design for physical vs. virtual iCat. This meant that all children used the text interface and the iCat for which the order of use was counterbalanced. Furthermore the children that used the virtual iCat did talk and played a game with the physical iCat at the end to get some additional information on their preferences for a virtual or physical robot.

### MEASURES

We limited the amount of questions to a minimum to keep the experimentation time reasonable.

*Fun:* The six questions regarding subjective fun (Table 9.1) were asked with the use of a smiley-o-meter [199], which is a five point Likert scale that uses smileys to represent the answers. We did also count the number of negative utterances and number of positive utterances and subtracted these from each other as a measure for observed fun. The utterances we counted are enumerated in Table 9.2.

*Acceptance:* Five different questions about acceptance were asked (Table 9.3). The questions were all posed on a five point Likert scale. We adapted the annotation of the scale to every question; An example of this is “Do you understand the robot” which has the scale “Never”, “Sometimes”, “Always”.

*Empathy:* For empathy four questions were asked (Table 9.3), the questions were also posed on a five point Likert scale and posed in the same way as the acceptance questions.

*Trust:* Three questions for trust were asked (Table 9.3). The questions were posed on a five point Likert scale similar to that of the acceptance and empathy questions.

*Efficiency:* The efficiency was calculated using the time of interaction with the interface. Because the virtual iCat and the physical iCat condition require some extra time caused by the “speech recognition”, this amount of time had to be subtracted. The subtraction of the speech recognition was done because in the future this will be done automatically and not by hand as was the case in this experiment. We calculated the efficiency by taking the total amount of interaction time minus the wizard time. This is around 6% of the total time.

*Learning effect:* The learning effect is related to the accurateness and completeness of the tasks. The effectiveness was therefore measured by the number of correctly answered quiz questions.

*Health intention:* Health Intention is interesting in relation with the motivational interviewing (change in lifestyle) approach we took. Therefore we asked questions about the attitude towards nutrition before the experiment and after the use of each assistant. The questions (Table 9.3) were based on the theory of Reasoned Action [87].

### PROCEDURE

Participants were told they participated in an experiment to evaluate personal assistants for children. They would work with a number of interfaces and have to fill in some questionnaires on what they thought of the interfaces.

They used the bots subsequently. First they answered a question about their health intention. And before using an interface, they answered a question about expected fun.

| | | Question |
|---|---|---|
| Acceptance | ia1 | Would you like to have the robot/chatbot at home? |
| | ia2 | Did you find it easy to work with the robot/chatbot? |
| | ia3 | Do you understand the robot/chatbot? |
| | ia4 | Which interface did you find easiest to use? |
| | ia5 | Which interface did you prefer? |
| Empathy | ie1 | Do you find the robot friendly? |
| | ie2 | Do you think the robot understands you? |
| | ie3 | Do you think the robot tells the truth? |
| | ie4 | Do you find the robot is curious about you? |
| Trust | iv1 | Do you think the robot tells the truth? |
| | iv2 | Would you answer honestly to the robot's questions? |
| | iv3 | Do you think the robot would tell your secrets to someone else? |
| Health Intention | hi1 | How many times a day would you like to eat fruit? |
| | hi2 | How many lollipops do you think you should be allowed to eat a day? |

Table 9.3: Questions regarding acceptance, empathy, trust, and health intention

They were told that when they would hear a beep, the interaction would start. The interaction with the interface followed a structured dialog, which was led by the interface. In the interaction, questions were asked by the bots and the participants were expected to answer on those. It was structured, since we wanted to let the participants experience more or less the same interaction, in order to be able to compare the results. In each condition, the dialog followed the same structure, consisting of three parts or tasks that represented the three different roles: motivator, educator, gamebuddy. First the assistant introduced itself (talking task/motivator), then a video quiz was played with the children followed by a quiz question (video quiz task/educator) and finally one or two tic-tac-toe games were played (game task/gamebuddy). After the interaction children were asked the five remaining questions on the experienced fun and the questions about trust, health intention (two after the first interface and three after the second), perceived empathy and three of the acceptance questions (ia1-ia3). In the end the children were asked what kind of roles or applications they would use the iCat for and ia4-ia5.



## 9.6. RESULTS

### FUN

The question about the fun expectation (ifx1) resulted in a significant difference between the physical iCat (mean = 4.6 out of 5) and the text interface (mean = 4.0 out of 5) (Mann-Whitney U (1,8)=20.5, Z=2.06, $p<0.05$). In addition, we compared the indicated value of fun per task within and between interfaces (ife2-4). The game with the physical iCat was valued significantly more fun (mean = 4.7 out of 5) than the quiz with the physical iCat (mean = 3.3 out of 5) (Sign test Z(1,8)=2.04, $p<0.05$). The same applied for the virtual iCat (4.8 vs. 4.0) (Sign test Z(1,9)=2.04, $p<0.05$) and the text interface (4.7 vs. 3.3) (Sign

| Argument | iCat | Text |
|---|---|---|
| Talking & no typing | 4 (20%) | |
| Talking | 3 (15%) | |
| Difficult to understand (speech) | | 3 (15%) |
| Typing | | 2 (10%) |
| No typing | 3 (15%) | |
| Difficulty reading | 3 (15%) | |
| Other | 2 (10%) | |
| Total | 15 (75%) | 5 (25%) |

Table 9.4: Reasons why children chose an interface (nr. & % of children)

test Z(1,18)=2.41, p<0.02). The game of the physical iCat was also experienced as more fun than the quiz of the text interface (4.7 vs. 3.4) (Sign test Z(1,8)=2.27, p<0.03). These results indicate that the game is considered more fun than the quiz.

The observed fun was measured by examining the result of the positive expression values minus the negative ones. In the talking task this gave significant differences between physical iCat (2.7) and virtual iCat (1.0) (Manova F(1,8) = 18.3) and between physical iCat (2.7) and text interface (0.0) (Manova F(1,8) = 7.0). When all expression values were taken together there was a significant divergence between physical iCat (10.9) and text interface (5.4) (Manova F(1,8) = 5.0). Another interesting measure is the total amount of fun utterances, which can be used to determine whether or not there are more positive utterances towards a particular interface. This measure provided two significant differences between both the virtual iCat (1.6 utterances) and the text (0.8 utterances) (Manova F(1,9)=7.0, p< 0.02) and between physical iCat (2.8 utterances) and text (Manova F(1.8)=8.7, p<0.001) So, children show more indicators of fun when talking with an iCat than with the text interface.

### ACCEPTANCE

Both acceptance questions about the ease of use (ia4) and preference (ia5), asked at the end of the experiment, showed significant differences between the different interfaces. The iCats were found easier to use than the text interface (Chi-Square (1,19) = 5.0, df = 1 p<0.03). The physical and virtual robots were found easiest to use, 70% and 80%, respectively. Similar results were found when asked for their preference. About 70% favored the iCats and 30% the text interface (Chi Square(1,19) = 4.1, df = 1 p<0.05) . The majority of the children stated the iCat to be more fun. The reasons they gave are summarized in Table 9.4. Children who performed their tasks with the virtual iCat were also given the opportunity to use the physical iCat. These children were also asked which of the three interfaces they preferred. The physical iCat appeared to be the most fun to work with. It was favored by 80% of the children, because it was real. Some additional comments were that its eyebrows and mouth could move. The remaining three questions regarding acceptance did not yield significant differences. All interfaces were rated high on acceptance: scoring 4.3, 4.5, and 4.4 out of 5 for the text interface, virtual iCat, and physical iCat, respectively. This indicates that all interfaces were very acceptable.

| Task | Mean iCat | Mean Text | One-way MANOVA | Sign. |
|---|---|---|---|---|
| Talking (physical iCat) | 27.6 | 60.6 | F(1,8)=15.5 | p<0.01 |
| Talking (virtual iCat) | 23.9 | 56.0 | F(1,9)=7.8 | p<0.02 |
| Game (physical iCat) | 122.6 | 187.3 | F(1,8)=9.6 | p<0.01 |
| Game (virtual iCat) | 120.5 | 171.9 | F(1,9)=11.7 | p<0.01 |
| Total (physical iCat) | 478.9 | 621.6 | F(1,8)=6.6 | p<0.03 |
| Total (virtual iCat) | 462.3 | 584.2 | F(1,9)=24.0 | p<0.001 |

Table 9.5: Significant results for efficiency (time on task)

#### Empathy

All the three interfaces had high scores on the empathy questions ranging from 4.0 to 4.2 out of 5: 4.2 for the physical iCat, 4.0 for the virtual iCat, and 4.1 for the text interface. All interfaces were thus perceived as empathetic. There were no significant differences between the interfaces.

#### Trust

The children rated all three interfaces high on trust 4.1 out of 5 for the physical iCat and the text interface and 4.3 out of 5 for the virtual iCat. Again there were no significant differences between the interfaces.

#### Efficiency

For the efficiency of the interfaces we looked at the duration of the complete interaction. Both the efficiency of the virtual iCat and the physical iCat differed significantly from the text interface (Table 9.5). A comparison between the iCat and virtual iCat did not provide any significant difference.

#### Learning effect

About 85% of the children answered the question, posed before the movie containing the information, correctly. This affirms that the children were already knowledgeable on the topic. On average the children answered 8.3 out of 10 questions correct. Thus no learning effects could be found.



## 9.7. Conclusion and discussion

The experimental set-up, in which only the speech recognition was simulated, worked well, and the physical and virtual robots were highly appreciated. We realized bots that could have meaningful and pleasant dialogues with children for their three roles. The interaction with the robots was significantly faster than with the chatbot and the physical robot was most fun to interact with. The game buddy role was important for the engagement with the personal assistant of the children. In contrast with the experiment with older adults [145], no significant differences were found for empathy. This can be explained by the high ratings the children gave to all three interfaces ("ceiling effect"). So, the proposed type of support for personal healthcare was well-accepted by the children in general.

This study compared three interfaces with their "natural" dialogue styles: a text-based chat-bot with two speech-based robots. You could say that we compared text to speech. We argue that a text interface for the characters would have been unnatural, because their appearance strongly suggests they have the ability to speak and listen. Correspondingly, speech dialogues are uncommon for the graphical, direct manipulation displays (windows).

In the short term, no significant discrepancies were observed regarding motivation and education between the different personal assistants. Therefore, a long-term experiment should be conducted in which engagement will play a larger role, because children will have to keep using the personal assistant for a longer period of time. Long-term effects of artificial agents in healthcare interventions are discussed in e.g. Marsella, Lewis Johnson, Bore [155] (education about cancer), Bickmore and Picard [26] (motivating to exercise), and Brave, Nass, and Hutchinson [38] (social support). These papers show the relevance of the educator, motivator and buddy roles for user support. The long term results suggest that virtual characters that exhibit affection are more enjoyable, more trustworthy, more supportive, and a better educator in comparison with no virtual character or a virtual character without affective abilities. Furthermore, learning results were better, and the participants were more willing to continue working with the social character. This literature focused only on adults. We would like to explore the long term effects on children and the effects of a physical character in comparison with a virtual character. In the healthcare domains we are looking into children with e.g. obesities, diabetes, and coeliac. These children should adapt their diet to stay healthy and are not allowed to eat the same as most children (i.e. a diabetic should keep track of his/her sugar intake). A buddy to cope with being different could be appreciated. Furthermore, the buddy could help educating them about their condition and motivate them to follow the physician's advice of the physician.

In the future the game buddy role should be extended to make it possible to play multiple games. Furthermore, the dialog agent should be able to handle more diverse interactions and preferably even conversations that were not anticipated by the programmer beforehand. As expected, the results showed that the quiz was valued as less fun than the game. Fun is very important to keep the children engaged, as we learned from the educational game developer during domain analysis. In the future, we would like to explore other educational methods that are perhaps more fun to use (this might eventually lead to a game educator).



In general, we can say that the children rated the interface properties high, which caused a small number of significant differences in the subjective measures. The objective measures also showed a preference for the robots, while their interaction was faster and exhibited more social behavior. They were excited about participating in the experiment and using the iCat. These results indicate that the iCat is an interface that attracts the attention and therefore can have positive effects on motivating and educating children while being a buddy, which is of importance when applying the robot in the healthcare domain. So, the motivator and educator roles that we developed are appropriate for both older adults (see [144]) and children, and the iCat is a good platform to implement and test such roles for both user groups.

## 9.8. Acknowledgment

We would like to thank all the children of "groep 5" from the Postiljon and their parents for their participation. Furthermore we thank IOP-MMI SenterNovem a program of the Dutch Ministry of Economics, for partially funding the SuperAssist project.

# 10 | Integrating robot support functions into varied activities at returning hospital visits: Supporting child's self-management of diabetes

**Abstract** *Persistent progress in the self-management of their disease is important and challenging for children with diabetes. The European ALIZ-e project developed and tested a set of core functions for a social robot that may help to establish such progress. These functions were studied in different set-ups and with different groups of children (e.g. classmates at a school, or participants of a diabetes camp). This paper takes the lessons learned from these studies to design a general scenario for educational and enjoying child-robot activities during returning hospital visits. The resulting scenario entailed three sessions, each lasting almost one hour, with three educational child-robot activities (quiz, sorting game and video watching), two intervening child-robot interactions (small talk and walking), and specific tests to assess the children and their experiences. Seventeen children (age 6-10) participated in the evaluation of this scenario, which provided new insights of the combined social robot support in the real environment. Overall, the children, but also their parents and formal caregivers, showed positive experiences. Children enjoyed the variety of activities, built a relationship with the robot and had a small knowledge gain. Parents and hospital staff pointed out that the robot had positive effects on child's mood and openness, which may be helpful for self-management. Based on the evaluation results, we derived five user profiles for further personalization of the robot, and general requirements for mediating the support of parents and caregivers.*

## 10.1. Introduction

### 10.1.1. Diabetes Type 1

The growing burden of chronic illness on health and health care has globally led to health policy responses increasingly referring to self-management. This applies to the increasing number of children and adolescents in Europe with a chronic illness. For example, the incidence of childhood type 1 diabetes mellitus (T1DM) in Europe, now ranging from 3.9/100,000 cases per year in Macedonia to 57.4/100,000 in Finland [187], is rising rapidly. In the below 5-year-old age group, there is a doubling time of less than 20 years [92]. T1DM is associated with serious physical and psychological complications

[63, 190], which may appear sooner or later, cause high morbidity and mortality, affect the quality of life, and increase health-care costs [93]. Complications can be prevented by performing self-management (e.g., monitoring blood glucose, recognizing symptoms and injecting insulin). However, self-management is not an easy goal to attain for young patients. First, it requires motivation and long-term perseverance, in order to become 'a way of life'. However, children's illness regularly causes feelings of embarrassment (approximately 25% of the youth involved in a study of Peyrot [190]), and negative effects on school performance and psychological well-being. Improving the way they feel about diabetes, might be a first step in improving the self-management. Second, the children need not only to learn to self-manage their lifestyle-related diseases to improve their situated health-related habits, but also to be prepared for the physical and social changes at adolescence. Third, the specific self-management goals of children and adolescents are strongly affected by a diversity of personal and environmental factors, such as the child's developmental stage, parent's support and health care providers. So, children and their social environment have to find a personalized strategy to establish pervasive self-management.

### 10.1.2. IMPROVING SELF-MANAGEMENT

There is a broad source of literature on theories that are relevant for self-management support: Changing behavior [153, 74, 206], persuasive design [90], gaming theory [62], education [256] and behavior change support systems [179]. These theories have some common principles. According to the first principle, intrinsic motivation is key and requires that someone feels in control of the situation (experience autonomy). This can be reached for instance by providing variation and influence of dialog. The second principle emphasizes the feeling of competence: The user should feel capable of reaching an objective. This principle originates from educational and gaming theory [256, 62], and from behavior change literature [153, 74, 206], stating that relevant activities and objectives should be provided, which are challenging and achievable, and for which positive feedback should be provided. The third and final principle concerns relatedness: Education and self-management are improved when there is a relation between tutor and trainee. The tutor can be a peer or teacher with whom a form of relatedness (or rapport) is build up [256, 153, 186]. The three factors: autonomy, competence and relatedness are the building blocks of the Self-Determination Theory (SDT) [74].

### 10.1.3. SOCIAL ROBOTS



Social robots show human-like (social) characteristics, e.g. they express emotions and use natural cues as gaze to share point of focus [91]. For prolonged self-management support, rapport should be build up between child and robot resulting in a positive effect on relatedness [23]. In Zhao et al. [186], several behaviors are identified to create rapport between an agent and a person. Examples are the initiation of mutual self-disclosure, praise and acknowledgement, and referring to shared experiences. It is interesting to note that these behaviors are also prescribed in behavior change methods, e.g., express empathy in Motivational Interviewing [161]. So, the social robot can be viewed as an embodiment of a behavior change support system [179]. Such robots are being used for behavior change support, for instance, to support persons with autism [204, 6], to acquire a healthy lifestyle [227], and to educate persons (e.g. [236, 239, 126]). A robot has a rich set of possibilities to incorporate behavior-change methods from social sciences, but the specific translation from these methods to a coherent and concise set of robot functions is complex and difficult to evaluate.

### 10.1.4. SITUATED COGNITIVE ENGINEERING

The European ALIZ-e project aimed at a social robot that 7 to 11 year old children could use recurrently and possibly help these children to progress on self-management (i.e., autonomy, competence and relatedness, [19, 18]; see section 1.2). An iterative situated Cognitive Engineering method was applied [172], to (i) derive use cases, requirements and claims for the self-management support (i.e. the design rationale), and (ii) build prototypes to test and refine the design rationale. The tests were conducted at schools and hospitals, focusing on specific parts of the design rationale, i.e. one or more “core functions” of the social robot that were hypothesized to have effect on relevant SDT-factors. For example, the idea that relatedness is stimulated by having a background story for the robot [250]. These functions were studied in different set-ups and with different groups of children (e.g. classmates at a school, or patients in a hospital). Often it was not (yet) required (for a first test and refinement cycle) to involve the target group, children with diabetes. This paper takes the lessons learned from these tests to design a general scenario, incorporating a variety of use cases. This way, an integrated set of core functions was prototyped and tested with children with diabetes in a hospital (i.e. the real target environment).

The next sections provide an overview of the earlier experiments conducted and their results. The current study incorporates the “proven” functions and makes use of the insights on the experimental setup that we built up in these experiments. The resulting social robot and scenario are evaluated with diabetic children in a hospital setting, studying the influence on autonomy, competence and relatedness. Furthermore, the perceptions and opinions of the children, their parents and their medical caregivers on the short and long-term are investigated. Conclusive evidence on the effects of the specific metrics could not be found, but the interactions with the children, parents and caregivers during the evaluation and afterwards gave valuable insights. Parents and caregivers became more enthusiastic over time and reported results in increased self-management and lower thresholds in hospital visits.

## 10.2. LESSONS LEARNED FROM PREVIOUS EXPERIMENTS



Over four years, several tests were conducted, in which children interacted with a social robot (Philips iCat or Aldebaran NAO) and performed one or several activities with the robot. These activities were designed to examine the effects of specific support functions, e.g. on specific learning objectives. Four educational activities were developed. The first was a Trivial Pursuit®based quiz in which robot and child played against each other. This educational quiz had a textual and competitive nature 10.1a. The second activity was an educational sorting game 10.1b, in which the child and robot classified objects in categories and could cooperate to reach the highest classification score. Due to its collaborative nature and visual orientation, the sorting game involved another learn-

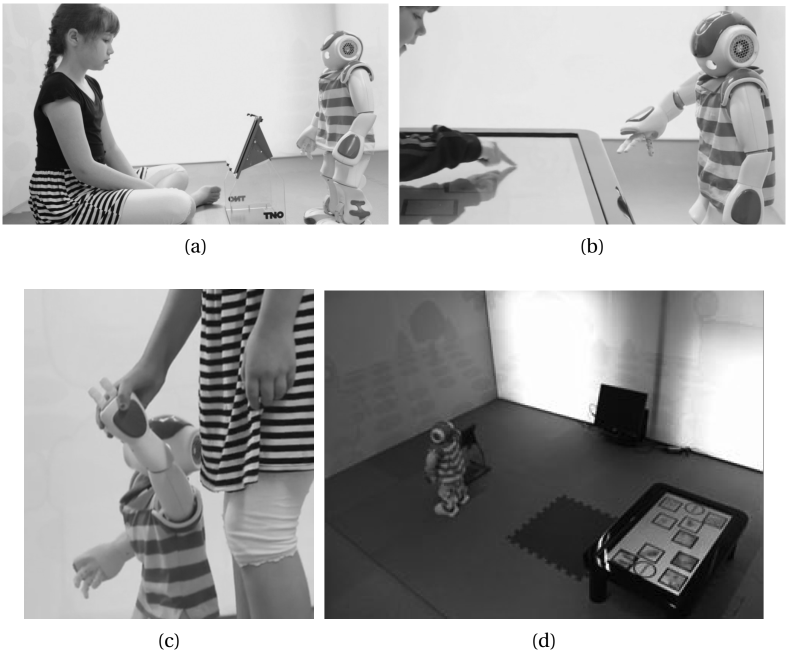



Figure 10.1: The experimental setup: (a)quiz, (b)sorting game, (c) walking with robot and (d) robot playground

ing style than the competitive and textual quiz (whereas, they could support the same learning objective). The third educational activity entailed different versions of movement games [208], which could address the same learning objective, but in a kinesthetic learning style. The fourth activity used educational videos that are both visual and aural. With this variety of activities, the social robot could support a variety of learning styles [89]. Next to these educational activities, there were "intervening" activities, such as small talk, to establish continuous child-robot interactions. All robot support functions were designed to address the objectives of SDT: autonomy, competence and relatedness. Table 10.1 provides an overview of the relevant experiments, their relations to the objectives of SDT, the context (setting and users) in which the experiment was conducted, the results and the transfer of these results into the integrated social robot (that will be tested subsequently).



According to SDT, a feeling of autonomy can be enhanced by providing choices. To stimulate this, the ALIZ-e project aimed at providing numerous activities that robot and child could do together. The quiz and sorting game were developed to support this. They both focus on education, but where the robot and child are playing against each other in the Trivial Pursuit®based quiz, in the sorting game they have to cooperate to get the highest score. In [98] it was shown that the possibility to switch between activities is beneficial for the motivation (see experiment 1 in Table 10.1).

The second factor of SDT, competence, can be supported by adapting the difficulty of the exercises to the child [111]. This adaptation proved to be beneficial for the motivation of the children (see experiment 2). It should be noted that the robot was not an expert in this interaction, i.e., the robot made the same amount of errors as the child [218]. Showing that the robot was not an expert was emphasized by making the robot exhibit thinking behavior [264]. Overall, this resulted in a positive experience of the robot (see experiment 3 and 4 in Table 10.1). In addition to competence, experiment 3 and 4 also addressed relatedness by encouraging self-confidence.

The third pillar of SDT is relatedness, meaning that the robot is experienced as a "pal". Firstly we made sure that the robot can exhibit recognizable emotions [55, 127] (see experiments 5 and 6). We also looked at adapting the robot to the personality of the child [249], but we found that personality is probably not a good aspect to adapt to (experiment 7). We still expect that adapting to energy level, and perhaps modulating the energy level of the child will support the relatedness, but this was not evaluated. We did evaluate the adaption of robot's emotional state to the state of the user and state of the situation (within boundaries) [242]. The results from this experiment showed that children who interacted with the robot that adapted its emotional state to the child and situation, showed more, and more positive, emotional expressions than children who interacted with a robot that did not adapt its emotions to the child and situation (experiment 8). However, recognizing child's emotions in an interactive situation is still very hard. Therefore, we studied the effects of remembering small facts about their life (e.g. name, hobbies, information provided in a previous session) [29]. This is rather easy to implement and proved to have a very positive effect on the children (see experiment 9). Another easy to implement functionality is that the robot tells something about itself (e.g. age, hobbies), which proved to increase the willingness of the children to disclose information about themselves [250] (experiment 10). Finally, we looked at the willingness of children to touch the robot [230]; experiment 11 showed that they are quite willing.

In addition to the conclusive results, interesting observations were acquired during the experiments that are relevant for the further development of the robot. For example, changing activities by the robot and the child themselves proved to be stimulating (e.g., to transfer from quiz to sorting game without the help of the experiment leader [98]; see results experiment 1 in Table 10.1). Another observation was that providing a confined, shared environment for the robot and child proved to reduce child's feeling of being observed and part of an experiment [230] (experiment 11).

Table 10.1: The 11 experiments that examined specific robot support functions for child's self-management with their relations to the Self-Determination Theory (SDT), the location of the experiment, the participants and the results.

| Nr. | Experiment | SDT objectives focus | Location | users (nr, age) | Results (implemented in current evaluation y/n) |
|---|---|---|---|---|---|
| 1 | Multiple activities (quiz, sorting game) in hospital setting [98] | Autonomy | Hospital | Non-diabetic children (13, age 7-11 ) - hospitalized | Multiple activities are beneficial (y) Child and robot should go from one activity to another (y) |
| 2 | Difficulty of math assignments adapted to performance child [111] | Competence | School | Children (20, age 9-10) | Adapting the difficulty of exercises to the child increases motivation (n) |
| 3 | Make the robot fallible in his answers [218] | Competence | School | Children (22, age 10-12 ) | Robot is able to adapt its performance to performance child (n) Fallibility is not proven effective, but theory is convincing (y) |
| 4 | Make the robot think about its answers. The robot takes some time and expresses thinking behavior [264] | Competence | School | Children (26, age 9-11) | The thinking robot is experienced as faster, more humanlike and more likeable, without decreasing perception of intelligence, trustworthiness and autonomy (y) |
| 5 | Make robot express recognizable emotions [127] | Relatedness | Research institute | Children (18, age 8-9) | Emotions are recognizable and add to likeability (y) |
| 6 | Compare capability to express emotions between robots [55] | Relatedness | School | Children (14, age 8-9) | Emotions of NAO are equally well recognized as thos of iCat (y) |
| 7 | Adapt robot personality behavior to personality child [249] | Relatedness | School | Children (16, age 7-9) | Personality was very hard to determine (no correlation between what child, parent and teacher indicated) so not a good factor to adapt to. (n) |
| 8 | Adapt robot emotions to child emotions and state of activity (within boundaries) [242] | Relatedness | School | Children (18, age 8-10) | Children that interacted with the robot that adapted its emotional state to the child and situation showed more, and more positive, emotional expressions than children that interacted with a robot that did not adapt its emotions to the child and situation (y) |
| 9 | Make robot remember small facts about their life (e.g. name, hobbies, information provided in a previous session) [29] | Relatedness | Hospital and At home | Diabetic children (30, age 6-12) - not hospitalized | Very positive effect on relation children towards robot (y) |
| 10 | Make robot disclose information about itself (e.g. age, hobbies) to stimulate disclosure of child [250] | Relatedness | At home | Diabetic children (6, age 9-12) - not hospitalized | Increased willingness of children to disclose information about themselves (y) |
| 11 | Ask the child to touch the robot in an interactive move session [230] | Relatedness | School | Children (22, age 9-11) | Most children like to touch the robot (y) The children like an enclosed environment where they are "more alone" with the robot (y) |

10

## 10.3. CONSTRUCTING AN INTEGRATED SET OF CHILD-ROBOT ACTIVITIES FOR HOSPITAL VISITS

Table 10.1 provides an overview of the 11 experiments that examined the specific robot support functions for child's self-management with their relations to the Self-Determination Theory (SDT), the location of the experiment, the participants and the results. The following subsections will elaborate on these results and describe how they will feed into the next version of the robot and the set of child-robot activities for returning hospital visits.

### 10.3.1. CHILD-ROBOT INTERACTION ENVIRONMENT

Based on the knowledge gathered in experiment 11 [230] we developed a physical setup for this evaluation. Firstly, we used the robot playground as used in [230] again. The playground (see Figure 10.1d) consists of three walls of 150cm high on which a robot landscape is depicted in soft grays. The floor consists of grey playtiles and one red and one blue depicting the positions the child should sit for the different games. All cables are hidden under the floor and behind the walls and two cameras are unobtrusively placed behind the walls so they just peek over it. The playground provides a shared environment for robot and child and since we did this experiment inside the hospital it also makes the surroundings more friendly. Furthermore, because children sit on the ground with the robot they are naturally on the same level as the robot, which is different when the robot stands on a table and the child sits at the table, which is also more static. Finally, the shared environment closes of the rest of the environment more, so the experimenter, who is in the same room, is easier to forget about.

### 10.3.2. CHILD-ROBOT ACTIVITIES

Next to the environment we made sure the interactions were in concordance with what we learned from previous experiences. The evaluation was a wizard-of-oz evaluation, which meant that the experimenter/wizard did the speech and state recognition of the child and there was a protocol that was followed that described the possible dialog and behavior actions. The wizard had some freedom to put in new text for the robot to say. The wizard had camera images from the playground, could switch from camera dependent on the activity, and had an elaborate wizard interface to direct the interaction. Overall, the activities consisted of three educational child-robot activities (quiz, sorting game and video watching), two intervening child-robot interactions (small talk and walking), and specific tests to assess the children and their experiences.



#### EDUCATIONAL ACTIVITIES

The child and robot could do three activities together, following experiment 1 that concluded that multiple activities are beneficial [98]. The two games as developed within the ALIZ-e project and an educational video. The quiz was based on Trivial Pursuit®. Child and robot each stand on opposite sides of a tablet in a kind of see-saw construction (see Figure 10.1a). The tablet is turned towards the robot and it can then ask the first multiple choice (A-D) question. After posing the question the robot turns the tablet towards the child and the child can answer, by saying the answer out loud (no touch).

The robot reacts on the answer and congratulates when it is correct and provides the argumentation when it is incorrect. There is no judgment when the answer is incorrect. Then the next question appears on the tablet and the child can pose it to the robot. The robot thinks about the answers it provides (experiment 4) [264] and makes errors (experiment 3) [218]. The game can be set up competitive, but we did not incorporate a scoring mechanism.

The sorting game shows pictures on a large touch screen (see Figure 10.1b), the pictures need to be swiped into one of two categories that are named/depicted on the sides of the screen. The categories are for instance “high in” and “low in” carbohydrates and pictures shown on the screen are “a salad”, “chips”, “bread”, “sweets”, “milk” etc.. Child and robot stand on opposite sides of the table and they can both, one at the time, swipe a picture in the correct category. The aim here is to get a high score together, so it is a collaborative game setup. During the game the robot acknowledges the actions of the child with exclamations as “too bad”, “you did great”.

The difficulty of both the quiz and the sorting game was not adapted to the users' performance although it was found to be effective (experiment 2) [111]. We did not do this because of a limited number of questions/assignments per session and a high variability between children. The questions/assignments were related to diabetes and thus relevant for the children.

The final activity is not a game, but an educational video the robot and child can watch together. The video is for instance about the symptoms of high blood glucose levels (a “hyper”).

After a certain number of questions of the quiz (8), or a certain amount of time with the sorting game (5 minutes) the robot initiated a change activity dialog. The child could then choose to proceed or change activity, although in the first and second session they had to do all activities so there was a time limit on how long they could do each game (10 minutes max). The child could also initiate the dialog to change the activity. When this was really soon after starting the activity, the robot tried to convince the child to do it a little longer (“just a few more questions”), otherwise it would agree on changing.

### SMALL TALK

Based on experiment 9 and 10 [29, 250] we incorporated small talk in the evaluation. At the start of the evaluation the robot asked the child some personal information: Name, age, hobby. The robot did also ask if the child had questions for the robot, so it could also answer questions about its age and hobbies. Furthermore, the robot asked at the end of the first and second session if the child had plans for the coming weeks (until the next session) and referred back to these in the next session. Finally, during the activities the robot asked questions about diabetes. The robot for instance said “The holiday period seems to be really hard to me, with all the candy and strange food, how do you deal with that?”. During the small talk and the activities the robot displayed emotions that correlated with the situation (experiment 5, 6 and 8 [55, 127, 242])



### WALKING

Because we did not want a detrimental effect on the interaction when switching activities, because of interference of the experimenter (experiment 1 [98]), we decided the child was responsible for getting the robot from one activity to another (see Figure 10.1c).

We thought this would work because experiment 11 [230] showed no hesitation of most children to touch the robot. We explained how to walk with the robot, but when some children started to lift the robot we also accepted this. Something else that came up after a few of the first sessions were finished, was that the robot fell over sometimes and most children felt the need to help it up. Therefore we added a function that made sure the robot would not hurt the child, shutting down the automatic stand up function and removing motor stiffness, so that the child could support the robot standing up. We also explained to the children how they could help the robot in standing up by putting it in sitting position.

## 10.4. Evaluation

In order to get a feeling of how diabetic children interacted with the NAO when different activities are offered and physical interaction is possible we carried out an experiment.

### 10.4.1. Evaluation method

#### Participants

17 diabetic children in the age of 6-10 (M=8.24, SD=1.25) participated in the experiment. They were selected by their diabetic nurses of the Meander Medical Centre (Amersfoort, the Netherlands) and on basis of the parents willing to come three times extra to the hospital. All children got the diabetes diagnosis more than a year and a half ago, the range was 23-108 months (M=51, SD=29,64). Most children used a pump to regulate their insulin intake (11), the others used insulin injections (6).

#### Materials

To execute the experiment in an adequate way the following materials are needed for the experimental setting: The child with the robot on the robot playground and the execution of the experiment including measurement material.

- Robot playground: playtile floor of $2 \times 3m^2$ with walls (Figure 10.1d)
- See-saw tablet holder, a device enabling turntaking by flipping the tablet
- Samsung Tablet
- 15” screen to watch little movies about diabetes
- 27” television touch screen with table legs, to play the sorting game
- Questionnaires
- Wizard Laptop
- Movie Laptop
- An extra screen to watch interaction
- Cameras to record interaction
- 3 NAO robots (2 minimum needed for third session and backup when technical failures occur)

### MEASURES

We used observations, tests and questionnaires to quantify and qualify the interaction with the robot.

**Tests**

*Knowledge test* This questionnaire is used to assess whether there is knowledge improvement. This test is filled out before the first and after the last interaction and consists of 32 knowledge questions (e.g. What is important for you to know about your physical education class? a) If you're going to do something fun, b) If it is active or calm what you're going to do, c) If you are going to play football, d) If you're clothes look good: b is correct). The questions one until eight occur in the first session of interaction on the tablet, questions nine until 16 in the second session and questions 17 until 24 in the third session (for the children who chose the quiz). When questions or answers were not understood or the children were not able to read they received help.

*Self-efficacy test*: The SE card-sorting questionnaire is used to assess the current autonomy of the child. To measure SE, a card sorting questionnaire based on Karoly and Bay [119] is used together with diabetes-care activities proposed by the diabetes specialists of the Meander Medical Center.

*Memory test* With the aid of a memorizing task we examine whether children memorize more information given by a familiar robot, as is expected when intrinsic motivation is higher due to a peer teacher that applies SDT strategies [175]. In the third session every child listens to two robot stories. One story is based on the English Wechsler Intelligence Scale for adults (Williams, 1997 [259]) and the other one thought up using the same build up. One story is given from the familiar robot (called Charlie) and the other story is provided by another NAO robot (called Robin), who is introduced as a friend of Charlie. This robot is exactly the same as Charlie, but has a different voice and wears a grey striped shirt. The order of the stories and the robots is counterbalanced. After each story there is a short recall memory test. First the children are asked to reiterate the story as best as they can (immediate free recall). After this they are asked nine questions about the story (Immediate cued recall). An example of such a question is: "what was the name of the lady in the story?"

**Questionnaires**

*Fun Questionnaire* To measure the pleasure and fun the children experienced the children filled in a Likert scale questionnaire about the robot and the activities. First there were three 7-point questions on fun with the robot, quiz and sorting game, after which four 4-point questions were asked related to different aspects of the robot. The questionnaires used were based on the Smileyometer from the Fun Toolkit of Read and MacFarlane [199].

*SDT Questionnaire* To measure the feeling of self-determination we asked the children 10 questions on a 7-point Likert scale. Question 2,3,8 and 9 were regarding feeling of competence, question 4,6,7,10 were about feeling of relatedness and question 1 and 5 were related to feeling of autonomy.

**Observations**

*Game preference* In the second session the children could say which game they preferred and were asked if they wanted to start with this game and in the third session they could only choose one game.

*Online analysis and offline video and logging analysis* For the analysis of the whole interaction in each session we used notes that were taken during the interaction, video analysis and analysis of the logs. We looked at walking, time with activities, game order, attention of child, interaction with robot (talking general, talking diabetic related, touching), reaction on technical failures, empathy with robot, and how much the experiment leader is involved.

## PROCEDURE

Every child had three sessions of about an hour in the hospital. These appointments were at least 14 days apart (see Figure 10.2).

| **Session 1** *(52 minutes)* | **Session 2** *(54 minutes)* | **Session 3** *(44 minutes)* |
|---|---|---|
| • Questionnaire for the parents *(appendix B)* | Self-efficacy questionnaire *(appendix G)(15 minutes)* | Smalltalk (and Walking) (2 minutes) |
| Knowledge test I *(appendix C)*(15 minutes) | | **Quiz/Sorting game (8 minutes)** |
| Walking and Smalltalk (10minutes) | | **• Story I told by Robin or Charlie • Questionnaire about story I • Story II told by the other robot • Questionnaire about story II** ***(appendix H)*** **(7 minutes)** |
| **Quiz (8 minutes)** | **Quiz/Sorting game (8 minutes)** | |
| Movie (2 minutes) | | Questionnaire about robot comparison *(appendix I)* (5 minutes) |
| **Sorting game (8 minutes)** | **Quiz/Sorting game (8 minutes)** | Questionnaire about robot III (5 minutes) |
| Movie (2 minutes) | | Knowledge test II (15 minutes) |
| Smalltalk and saying goodbye (2 minutes) | | |
| | Taking a photo with the robot (2 minutes) | Questions about joining further experiments *(appendix J)* |
| Questionnaire about robot I, II *(appendix F)* (5 minutes) | | |

Figure 10.2: Planning for the three sessions.

In the first session the NAO robot, called Charlie, is introduced as a robot that helps children to manage their diabetes but still has to learn many things about diabetes himself. The experiment leader explains the activities in short and shows how the children can walk with Charlie. The interaction with Charlie starts with small talk and walking followed by one of the games. With the quiz Charlie has to be put exactly in front of the bars on the ground to be able to turn the tablet. In each session at least eight questions are played so that after three sessions they practiced 24 of the 32 knowledge test questions (if they chose to do the quiz in the last session). The sorting game is on the other side of the playground on a large touchscreen. Several pictures are shown and the child and NAO have to put them in the correct category (on one of the sides of the display). Examples of categories are: hyper/hypo, low/high carbohydrates. During each game open questions related to diabetes are asked to support self-disclosure (e.g. "Did it ever happen to you that you had a hypo or hyper and did not notice? How come did this happen?”). In between the games in the first and second session the children can watch an one-minute movie about dealing with typical diabetes situations, which is presented on a 15” screen. Dependent on the time left another short movie can be presented to the child after the games. After the interaction with the robot the children always fill in a questionnaire concerning judgment of the robot and the games they have played. The first session starts with the quiz. In the second session the children are allowed to choose with which game they want to begin with but they have to play them both. In the third session only one game is played, chosen by the child, because of a new scenario where the children meet Charlie’s friend Robin. Both Robin and Charlie tell a short story after which the children have to do a test with free and cued recall about the story.

## 10.5. RESULTS

Below we will describe the results from the evaluation. These results are divided in results that can be directly derived from the instruments and observations used in the evaluation and in feedback we got afterwards.

The tests were analyzed using t-tests and the questionnaires using the non-parametric Wilcoxon and Friedman tests. Game preference was counted and compared between the second and third session. The video and logging analysis was performed using Grounded Theory as starting point. This was because the 17 children differed in age, phase in their illness and interaction with the robot so much that we couldn’t compare between them. What we could do was analyzing the data looking for similarities and differences, to create preliminary user profiles, on which the robot could adapt its interaction in the future. All videos and logging files were watched and we looked at similar behavior between the participants on aspects as speech and touch interaction (time spent, manner of interaction, extravert behaviour etc.)

### 10.5.1. TESTS

The self-efficacy test is excluded because most children had some difficulty filling it in. Furthermore, the test took too long to do a pre- and post test.

**Knowledge test** Questions 7, 8 en 18 are excluded because we noticed that multiple answers were correct. A paired sample t-test shows that there was a significant difference in knowledge acquisition between the pre- and post test for the questions that were presented during the experiment (1-24). First session M=11.35, SE=.77, second session M=13.7, SE=.66 and a paired t-test t(16)=5.6, p<0.001 (2-tailed). The final eight questions (25-32) did not show significant improvement t(16) =1.19, p=.25 with M=5.94 and SE=.34 for the first session and M=6.29 and SE=.44 for the second session.

**Memory test** We did an independent samples t-test to test whether there is a significant effect of the robots in the immediate free recall and in the immediate cued recall (see Figure 10.3). There are no significant differences assessed between the scores reached after the stories told by Charlie and the scores reached after the story told by Robin in the immediate free recall ($p = .114$, $p = .521$)and in the immediate cued recall ($p = .869$, $p = .306$).

Table 10.2: Fun questionnaire Means and (SD) (* 1 NA due to technical failure sorting game, ** for quiz: 2 NA and 6 children who filled in something while they didn't play the quiz, for sorting game: 3 NA and 6 children who filled something in while they didn't play the sorting game, *** 1 NA (missed question))

| Question (Scale) | Session 1 | Session 2 | Session 3 |
|---|---|---|---|
| How much fun did you find Charlie the robot? (1-7) | 6.5 (0.87) | 6.8 (0.43) | 6.8 (0.44) |
| How much fun did you find the quiz? (1-7) | 5.8 (0.88) | 6.2 (0.66) | 5.7 (1.16) ** |
| How much fun did you find the sorting game (1-7) | 5.9 (1.14) | 6.1 (1.12) * | 5.7 (0.73) ** |
| How friendly did you find Charlie the robot (1-4) | 3.9 (0.33) | 3.8 (0.39) | 6.1 (0.40) *** |
| How well could you play together with Charlie the robot? (1-4) | 3.6(0.51) | 3.8 (0.39) | 3.8 (0.47) |
| How "cosy" is Charlie? (1-4) | 3.7 (0.44) | 3.9 (0.33) | 3.8 (0.62) |
| How warm (hospitable) is Charlie? (1-4) | 3.5 (0.62) | 3.4 (0.61) | 3.8 (1.35) |

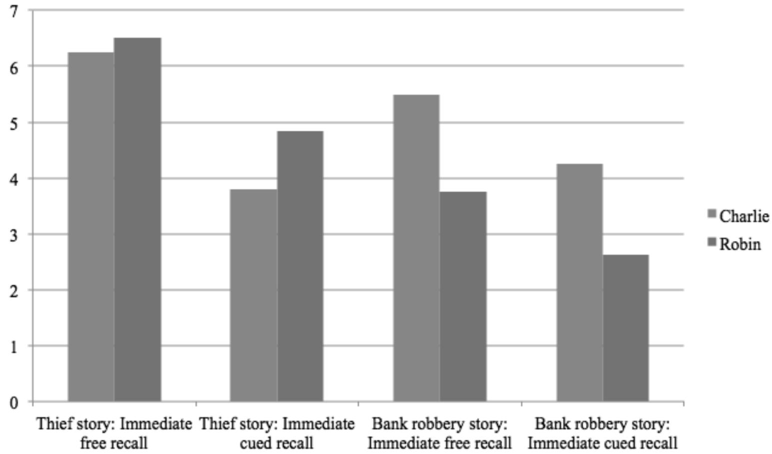



Figure 10.3: Story recall comparison between Charlie and Robin.

## 10.5.2. QUESTIONNAIRES

**Fun** We had separate questions on fun with robot, quiz and sorting game. Over the sessions these did not change significantly. The same was true for the separate questions on interaction with the robot (see Table 10.2).

**Self-determination** For the self-determination questionnaire we aggregated the questions related to competence, to relatedness and to autonomy per session.

*Competence* Overall, 49% of the children rated their feeling of competence a 7 (highest) and only 4% rated their competence under 4. In session 2 this was 56% and 7% and in the third session 50% and 4% . This means that no improvement was possible for almost half of the children and only very little for the children who scored initially under 7.

*Relatedness* We performed the same procedure as for competence and counted the number of times a 7 (highest) was chosen. 69% of the time children felt very related to the robot and only 6% chose a rating under 4. In the second session this was 76% and 1% and the third session 74% and 3%. So as with competence the ratings were already so high in the first session there was little room for improvement, 54% of the questions were rated a 7 on all three sessions.

Table 10.3: Example of 'Children who "just deal with it"'

| | |
|---|---|
| **Kind (pp12):** | Ik ga heel goed opletten, wat ik eet en ik kijk goed op de verpakkingen. En dan onthou ik dan. Als ik bijvoorbeeld bij Sinterklaas pepernoten wil eten, weet ik hoeveel in 50g zit en dan hou ik dat in mijn hoofd als ik dan de volgende keer 100g wil eten weet ik dat dubbele moet doen. |
| **Charlie:** | Oh, wat goed zeg! Nu ik dat weet, kan ik het ook aan andere kinderen leren. |
| **Kind (glimlacht):** | Oh dat is fijn! |
| *Translated* | |
| **Child:** | "I'm very careful with what I eat, look on the packaging and remember that. So If I want to eat ginger nuts at Sinterklaas for example I know how much sugar there is in 50g. I keep that in mind and when I want to eat 100g I know that I have to do twice as much insulin." |
| **Charlie:** | "Oh, great! Since I know that now, I can tell it to other kids." |
| **Child:** (smiling) | "Oh, that is good." |

*Autonomy* The autonomy was rated a 7 (highest) for 38% of the time in the first session (15% under 4), 44% in the second session (6% under 4) and 53% in the third session (6% under 4). Because of this increase we performed a Friedman test, but this was not significant p=0.29 (df=2, $chi^2$=2.45).

### 10.5.3. OBSERVATIONS

**Game Preference** In the second session 9 of the 17 children chose the sorting game as their favorite and 8 chose quiz and they also agreed starting with this game. In the third session 8 children chose to play the sorting game and 9 the quiz. 16 of the 17 children chose the game they preferred in the second session to play in the third. Only one child switched from sorting game to quiz.

**Video and logging analysis** From the video and logging analysis we extracted five user profiles as shown in the following and we did some additional observations.



**User profiles** The profiles were based on observations made during the experiment itself and observations from the videos afterwards. During the experiment the wizard, who was the same in almost all sessions, made notes about the behavior of the child in the experiment. Afterwards the same person identified some aspects, based on the notes and rewatching a few videos on which the children could be categorized in profiles. The scoring aspects were discussed with colleagues. Then taking these aspects all sessions were watched and scored. Aspects we looked at were related to dialog and actions of the robot, e.g. naming the child, falling. But also to dialog and actions of the child, e.g. reaction on falling of the robot, attention towards robot, time spent in activities, talking with robot (only telling or also listening), walking with robot, reaction on diabetes related questions. Finally we did also some general observations about the child, e.g. happy, open, shy, technology minded.

1. **Children who "just deal with it"** *(pp3, pp10, pp12, pp13, pp16)*: In this group there are children who know very much about diabetes and how to deal with it. They can tell about it in an open manner, even about the difficult parts (see Table 10.3). They seem to feel good and do many things on their own. In the group of children who "just deal with it" there are also children whose parents have diabetes. The children who indicated that their parents have diabetes seem to be much more relaxed and open for diabetes related questions and providing information to Charlie in a positive way. Diabetes for these children seems to be a shared (and not problematic) lifestyle together with a parent.
2. **Children who feel to fall outside the group** *(pp2, pp9, pp11)*: Children who seem to feel not that comfortable yet with having diabetes and the integration of it in their life belong to this group. Different reasons can be listed for this feeling. For example when children do not know enough about their diabetes, cannot connect the consequences of the diabetes to their feelings and are therefore more dependent on their parents. In the interaction this becomes clear by difficulty answering the open questions related to diabetes. They also see Charlie immediately as a friend, this is shown by having a picture of Charlie above the bed at home (pp2), having lots of empathy for Charlie when it falls (pp9) and more then passing interest in how many friends Charlie has (pp11).
3. **Children who are afraid to make errors** *(pp4, pp5, pp14, pp17)*: When children look away very often during interaction, give answers which are not consistent with their behavior or are ashamed to say anything, it seems that children react only like that because of someone listening or watching (for example Table 10.4). These children seem not that sure in what they know about diabetes and do not dare to say something, because it could be wrong.
4. **Children who are shy** *(pp7, pp8, pp16)*: These children take a longer time to tell something or do something with the robot. Often they whisper their answers, or just laugh a bit uncomfortable.
5. **Children who have difficulty with multitasking** *(pp1, pp5, pp6, pp17)*: Some children in this experiment were still very young and had difficulties with talking with Charlie and playing the games at the same time. Sometimes these children could not read the quiz questions themselves. The experiment leader plays a big role in these interactions. Social desirable behavior is almost unpreventable in those situations. In the most cases they also know less about diabetes than the other children and do less diabetes related actions on their own.



**Other observations** In general some children touch the robot from the first meeting on, curious about how it feels. Especially in the last session Charlie gets many questions of how it works. All children are interested in unpredictable facts about Charlie as for example the name and colors of Charlie's soccer club and the outcome of the last game. Furthermore, compliments seem to support all children: They react positively on them, some react more reserved whereas others give the robot compliments in return immediately.

Table 10.4: Example of 'Children who are afraid to make errors'

| | |
|---|---|
| **Kind (lijkt arrogant/onzeker en kijkt vaak weg) (pp4):** | Nou, als ik iets wil eten, dan spuit ik gewoon. |
| *Translated* | |
| **Child (seems arrogant/unsure, often looking away):** | So, when I want to eat something, I just inject insulin |

Walking the robot is not very easy, at least not to bump into anything, but it is appreciated by most and when it goes not fast enough they just carry it to the intended spot. Also the falling seemed to support most children in feeling useful, but not all children liked to help the robot after it fell. All children had to help the robot to the other activities and all children experienced at least one fall during their three sessions. For some children this occurred more often than for others. Our feeling was that although helping to stand up was beneficial the falling had a negative influence when it occurred often.

In the dialogs we saw some progression in what was disclosed towards the robot, they really wanted to tell the robot about their experiences in between the visits. Very noteworthy is that 4 children gave a present to the robot (drawing, paper craft, loom bracelet and World Football Cup goodie) (see for example Figure 10.4).

### 10.5.4. FEEDBACK AFTER EVALUATION

At the moment of completing this paper the experiment has finished a year and a half ago; since that time we received great feedback from parents and medical staff. Parents have told us of more independence since the three 20-minute interactions session. Medical staff tells us that children still ask when the robot returns and that they notice children are more at ease at the hospital since the experiment. In follow up contacts we noticed that parents, children and medical staff are more willing to participate in a follow up study than they were to participate in this study. This is also apparent in the fact that the Meander Medical Centre is now part of the H2020 project PAL that also looks at the use of the robot, in physical and virtual form, for children with diabetes.




Figure 10.4: Drawing and paper craft gifts

## 10.6. CONCLUSIONS AND DISCUSSION OF THE EVALUATION RESULTS

### 10.6.1. TESTS

After negative experiences with other questionnaires, we decided to use this self-efficacy questionnaire with the sorting cards. This method seems to work well: It encourages the children to think about their answers and vary them. But the questionnaire was not enough adapted to the target age and took too long to fill in. So although it did not have the desired result now, we would like to refine it and use it as pre- and posttest for self-efficacy in the future. In the Netherlands there is a list of “Know and Do” objectives for different age groups (6/7, 8/9, etc.); we are looking in to using this to measure the level of self-management. Of course we will also look for alternatives to measure variation in self-efficacy related to diabetes over time. It should be noted that parents and medical personnel indicated (after the experiment) that self-efficacy was improved. One of the parents for instance told that their daughter made more decisions on her own, like adapting the insulin before a meal because she wouldn’t eat a lot of it. The parent said that the fact the robot made errors did have a positive effect. Furthermore, although not significant, there was an increase in autonomy according to the questionnaire.

The knowledge test had good results, but improvements are possible. Some (more interesting) questions had multiple possible answers, because in many situations there are multiple solutions for the problem at hand for diabetics (this is just one of the things we want to learn the children). Also the reaction to high or low bloodsugar is dependent on the situation: Illness, stress, physical activity and food influence the bloodsugar and to keep the variation at a minimum it is necessary to know why the body reacted in this way to come to the best reaction. Furthermore, we noticed that children answered lifestyle questions truthfully. So when asked how they handled a situation like telling a parent of a friend they had diabetes, they did not provide the “correct” answer, which was very obvious (“I do x because then I show I’m the boss of diabetes”), but said they rather not tell because it would make them different. We were very suprised, but also happy with this. We rather have the answer about how they handle such a situation so that we can make them understand why they should change behaviour than that they provide the “correct” answer.

The memory test did not result in a significant difference between the familiar (Charlie) and unfamilair robot (Robin), but we did see some opportunities to improve the test. First we need to make absolutely sure that both robots are equally understandable, while speaking with different voices and we should use a validated, for the specific age group, verbal memory test.



### 10.6.2. QUESTIONNAIRES

All questionnaires suffered from the same problem, a ceiling effect. A score below the 6 was low which makes it impossible to have an increase over time. Next to this we saw that the sorting of cards in the self-efficacy test had a positive effect on thinking about a question, whereas some items of the questionnaires stimulated putting crosses automatically. This could be seen for instance in the questions of session 3 where many of the children (12 out of 17) answered questions about the activity they did not perform.

It keeps being a challenge to have questionnaires that are informative, but they are still an important measurement method, so we will keep adapting them and hope to create an informative questionnaire. Furthermore, we will look further into ways to decreasing the effect, like make the answering more tangible (e.g. no cross but moving something to the answer), more forced choice, implicit association tests, providing parents with questionnaires for some effects, longer evaluation periods, and more.

### 10.6.3. OBSERVATIONS

#### GAME PREFERENCE

It was nice to see that some children preferred the quiz while others preferred the sorting game. This encourages us to proceed with having different activities that are performed with the same robot to reach the same objective and that which activity is performed depends on the child's preference, state and current objective.

#### VIDEO AND LOGGING ANALYSIS

**User profiles** The user profiles indicated in this experiment are a starting point for us to focus on some parts of the interaction and see if we can recognize these same profiles in another experiment or that they need to be adapted. The profiles as they are now, are solely based on the interpretations of one coder and thus need to be verified. After a set of stable user profiles is identified we want to use these profiles in the future to make a fast adaptation to the user possible. Below we provide per user profile a first idea on how the user profile influences the adaptation.

1. **Children who "just deal with it"** *(pp3, pp10, pp12, pp13, pp16)* The robot can tell the children who are more uncomfortable with their diabetes how these children could deal with it. The children mention that the robot needs to know more and get a teacher role which can give them more self-confidence. This group is challenging for the interaction because in particular the children who are easily comfortable in the interaction with the robot are also the first who get bored by the robot and its games. Fortunately, this group seems to be interested in a robot and how it works. In the interaction with this group this could be taken advantage of. Although the children in this group are already quite confident with their diabetes they might benefit from short interactions to provide them with a bit more confidence to take the next step in self-management. This idea is fed by the feedback we got from some parents with children in this group.



2. **Children who feel to fall outside the group** *(pp2, pp9, pp11)*: For these children the robot has to be a real friend. Remembering what the children said adds great value. It seems to be nice especially for these children to share interests with the robot, for example playing cards (pp2) or wearing bracelets (pp11). The robot should combine friendship and dealing with diabetes. To not break the bonding with the child, the robot has to be careful with its questions and for example not ask a question like "What do you do with Santa-Claus, so many weird food, how do you deal with it?" in the beginning of the interaction to not bring the child in an unpleasant situation.

3. **Children who are afraid to make errors***(pp4, pp5, pp14, pp17)*: The robot can show the children that it doesn't matter to make errors by making errors itself. It can give the children self-confidence through playing the games and praise when the children did something good. The bonding can grow and the child can grow too.

4. **Children who are shy***(pp7, pp8, pp16)*: When children are very shy, the robot should be patient, and should play and walk with the children instead of talking too much. Some children need more time to talk about difficult issues. The robot has to try to estimate such children's state and help them managing their diabetes without being too pushy.

5. **Children who have difficulty with multitasking***(pp1, pp5, pp6, pp17)*: To improve self-efficacy and knowledge with children who have difficulty with multitasking, the robot should catch the attention and hold the attention of the children. That is very challenging especially because children are very good in ignoring other things when they are engrossed in something else. The bonding with the robot could grow in first instance via playing and later via dialogue.

### 10.6.4. FEEDBACK AFTER EVALUATION

The feedback after evaluation provided us with lots of information, but in a semi-structured manner. Our experiences during this experiment with small talk with parents and health care professionals when they were watching the sessions and afterwards has shown us the importance of involving them in a more structured manner. In the future we will do this by involving them more in the design and evaluation via focus groups, structured interviews, participation in the experiment and questionnaires.

## 10.7. GENERAL CONCLUSION AND DISCUSSION

### 10.7.1. MAIN OUTCOMES

Overall, the general scenario for educational and enjoying child-robot activities during returning hospital visits, proved to capture the lessons learned well. The children had very positive experiences in the three sessions of almost one hour (i.e., quiz, sorting game and video watching, and small talk and walking). The children, but also their parents and formal caregivers, showed positive experiences. Children enjoyed the variety of activities, built a relationship with the robot and had a small knowledge gain. Parents and hospital staff pointed out that the robot had positive effects on child's mood and openness, which may be helpful for self-management. Based on the evaluation results, we derived five user profiles for further personalization of the robot, and general requirements for mediating the support of parents and caregivers.

More specifically, personalization to developmental age, interests and objectives of a specific child, proves to be important for both the interaction as the questions asked. Furthermore, we should not only focus on improving self-efficacy of the child, but also on improving confidence of the parents in their child. Many of the parents were overprotective. Involvement of children, parents and medical staff is thus essential. Fortunately we have seen that formal and informal caregivers changed from skeptic to enthusiastic, based on the reactions of the children who showed increased self-management and more positive hospital experience. The robot showed to have a new role for self-management that is different from that of the caregiver and peer. If the long-term effects follow the same line is to be seen, the positive attention the children received now in relation to their illness can already explain many of the beneficial effects of the robot intervention. On the other hand, if we can have such an effect with three 20-minute sessions with a robot it is worth the effort.

### 10.7.2. IMPORTANCE OF EVALUATION "IN THE WILD"

Performing an evaluation with children with diabetes in a care environment provided us with knowledge and experiences we could not have acquired doing evaluations at schools. We noticed that diabetic children's experiences with the robot differed from "healthy" children. They seemed to be more open for social interaction with the robot and also the fact that the robot was not all-knowing and dependent on the child seemed to influence these children more than healthy children. This was the first evaluation the robot received gifts from children, which shows that there is some kind of bond/relationship forming. The shared space of child and robot added to this experience as did the dependence of the robot on the child when it fell or had to go to another activity.

Because the children were brought to the experiment by their parents who often waited in the same room as the experiment leader (outside the experiment room) it was the first time we could interact with parents for a longer period. We of course knew that parents of children in this age group are of a huge influence on the child, and that this might be even more so with chronically ill children, seeing it first hand does change how you look at this influence. There were parents who already said at the beginning that they did not know if their child could perform well in the evaluation and we saw this back in the shyness of the mentioned child that changed a lot during the three sessions. Furthermore, having a child with diabetes has tremendous influence on family life. So caretakers and social environment influence the child, but the child also influences his or her environment. In future research we will take the influence and experiences of family and social environment into account.

The evaluation took place in the room next to the coffee corner of the hospital staff involved in the care of the diabetic children. This was great because they could look through a window and see what was happening, but also talk to parents and experiment leaders while getting coffee and thereby getting a better feel of the aim of the robot. They could see the enjoyment of the children, and also see and hear that the robot will not substitute them.



One of the main challenges we found is that because of the bond the children seemed to form and the things they discussed with the robot it did not feel ethically right to strictly follow protocol. For example when a child discussed his or her problems with diabetes because of a birthday party the robot did not react with "I don't understand", but the wizard typed in a relevant comment for the robot to say. Due to this and technical problems, no session was the same and the applicability of inferential statistics was limited.

### 10.7.3. Future work

This evaluation showed that parents, medical staff and children enjoyed working with the robot and saw advantages of the use. The next step is now to develop a prototype that can stand alone, might also be used at home (in virtual form) because there are only a few hospital visits, and that involves all stakeholders. This means we need at least a solution to deal with speech recognition and dialog management, personalization on at least child interests, developmental age and objectives towards self-management, and evaluating effectiveness so that care institutions can argue for the costs of using the robot. Currently, these aspects are being addressed in the European H2020 project PAL (www.pal4u.eu).

## 10.8. Acknowledgments

First we would like to thank the children and their parents, who had to come three times outside of regular appointments. Without the Meander Medical Centre and the support of the diabetic team there, especially Roos Nuboer and Mirjam Schouten who made the whole team enthousiastic and made sure that our schedule was full and tight, this experiment would not have been possible. Finally, not only the authors of this paper contributed to this experiment, but also Joris Janssen was indispensable as experiment leaders and Bert Bierman for his never ending and never failing technical support. This work is (partially) funded by the EU FP7 ALIZ-E project (grant number 248116) and the H2020 PAL project (grant number 643783).

# IV

# Conclusions and Discussion

*All good things must come to an end*

# 11 | Conclusions and Discussion

Our vision of situated child-robot collaboration is described in chapter 1 (page 3): A child with diabetes interacts regularly with a robot pal and its avatar to attain, step-by-step, self-management objectives. The set of objectives to pursue is, in consultation with the child, provided by the health care professional. The envisioned child-robot collaboration is developed incrementally; in iterative design and evaluation cycles. This thesis focused on the first building block of the self-management support that was studied mainly in the ALIZ-E project [18]: the child-robot interaction in the hospital (i.e., the research and development on the second building block, the avatar, is being done in the follow-up PAL-project [31]).

The following research question was derived from our general vision and research focus:

- Which robot functionalities and behaviors support the motivation and competencies for self-management of diabetes by children in the age of 7-12 with diabetes type 1 (T1DM)?

Following the situated Cognitive Engineering methodology, we conducted systematic analyses of (1) the domain (i.e., operational and user demands), (2) the human factors (i.e., relevant behavior change theories and methods), and (3) the technical opportunities and limitations of the robot platform (i.e., the NAO robot and mobile technology). Subsequently, we constructed a general design rationale format that captures the main concepts from the analyses outcomes with their interrelationships: objectives, methods, use cases, functions, interaction design patterns, effects and instruments chapter 2. This design rationale captures relevant behavioral change theories and techniques with the corresponding robot behaviors and expected situated effects, which were worked out in specific design questions and hypotheses to be tested in separate experiments. In total, the thesis presents eight experiments divided over three parts: Part I Interaction design patterns, Part II Single use cases and Part III Multiple use cases. In this section, we present the conclusions of each chapter. Subsequently, we present the contributions and limitations of this thesis. The last section discusses future work.

## 11.1. Conclusions

### 11.1.1. situated Design Rationale: concepts and interdependencies

To answer the main research question presented above, we first had to define the core concepts of (1) children's diabetes self-management, and (2) the behavior support methods for motivation and competencies that can be integrated into robot functionalities and behaviors. Furthermore, we needed to define the interrelationships between these concepts for a concise and coherent analysis of the support-effect structures (called situated Design Rationale; sDR). The corresponding research question is:

> *RQ:* Which knowledge structure can capture the core design and evaluation concepts of behavior change support robots for children's diabetes self-management?

chapter 2 studied this research question, starting with a short discussion of relevant theories and methods on behavior change [160], on behavior change support systems [179] and cognitive engineering methods [172]. This literature analysis provided a set of seven concepts with their relations: *Objectives* can be achieved by *methods*, which are contextualized in specific *use cases*. A *function* serves the method and is shaped by *interaction design patterns*. The function and interaction design bring about specific effects for one or more use cases, which demonstrate the progress towards the objectives. The *effects* are measured with *instruments*, which are appropriate for the end-users (in our case children). See 2.1 for the generic view. To describe the sDR a tool that is able to describe relations and is able to visualize them is necessary.

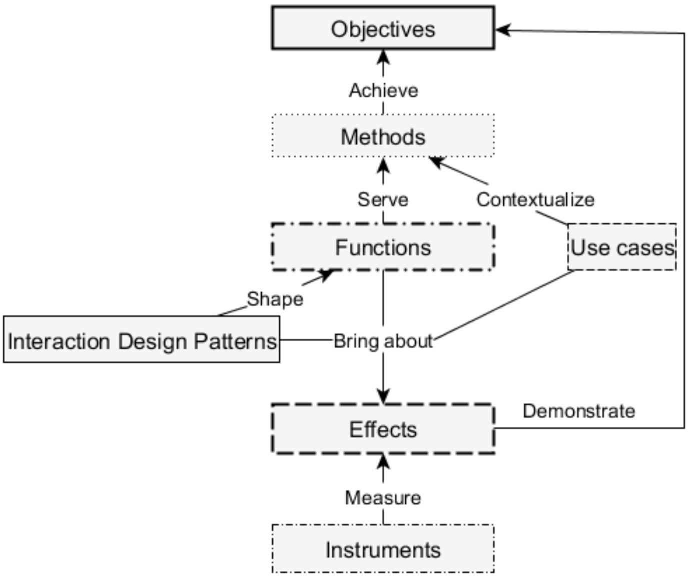



Figure 11.1: Generic concept map of the situated Design Rationale (sDR).

The development of the situated Design Rationale was an iterative process, running in parallel with the development of the PAL diabetes self-management support system. Like the development of design patterns [173], the sDR has been abstracted from a large number of case studies that were conducted in the ALIZ-E project [18]. Overall, the sDR proved to capture all main design and evaluation results. It helped to identify the unforeseen effects in experiments, which should be addressed in future research and development activities.

### 11.1.2. PART I: EMOTIONS IN DIFFERENT EMBODIMENTS

Emotions are a part of natural human-human interaction and have a substantial impact on self-management of lifestyle related diseases [161].

We applied multi-modal emotion expression models to the ALIZ-e robot (NAO from Aldebaran) and a previous used robot (iCat from Philips) and evaluated their recognizability and effects on children's behavior. The studies focused on robot's facial, vocal and posture behaviors.

#### CHAPTER 3: MULTI-MODAL EMOTIONS OF A FACIAL EXPRESSIVE ROBOT

> Design Question: How to model the four Ekman emotions of anger, fear, happy and sad in the face and speech of the iCat, so that they are recognizable for children?

Ekman's research on emotion expressions has been well recognized in human-robot interaction [82, 229]. Facial expressions are core in this work. We used the iCat, because it was a good robot platform to study such facial expressions [147]. For the design of emotions in speech, relevant research of Cahn [43], Murray [167], Schroeder[219] and Kroes [137] was used. Elements that were adapted, depending on the emotion the iCat should express were, in the face: Lips, eyebrows, eye direction and in the voice: pitch and speed.

In the experiment 18 children of 8 and 9 year old participated. Each child worked with two robots each displaying a different emotion condition (neutral, emotional face or emotions in face and speech). They were asked to perform three different tasks, 1) the iCat told a story with emotional events in first person and exhibited an emotion at the end of each event after which a forced choice was presented to the children, 2) the iCat explained the rules of a sport and asked multiple choice questions about this, 3) the iCat tried to motivate children to do a cognitive assignment (select blue marbles from a bowl with colored marbles) and a physical assignment (perform as many steps as you can within two minutes).

Each task evaluated a different aspect. Task 1 was used to evaluate the recognition rates of emotions between the different conditions. The results showed no significant differences between the three conditions, results were between the 78% and 88% .

> Hypothesis: Children will show better understanding, acceptance, trust, fun, empathy and performance when interacting with an iCat that expresses multi-modal emotions (i.e., increasing in the following order: no emotions, facial, facial-and-vocal).

The objective measures on performance; recognition rate (task 1), correct answers (task 2), and number of steps and found marbles (task 3), are not significantly influenced. The children did however answer the questions during task 2 faster in the speech and face condition (4.7 sec) than in the other two conditions (5.9 sec). The tasks were deemed more difficult in the speech and face condition and the robot less intelligible in the same condition. This reduced intelligibility was reflected by a reduced amount of trust the children had in the robot. The face and speech robot was preferred over the face only robot. The more emotion modalities did not lead to higher scores on the factors. Because of the result on intelligibility, we can conclude that the emotional voice was not well designed. Consequently, we cannot derive sound conclusions on possible effects of the speech functions (i.e., the shaping of the function into the design had serious shortcomings, see Figure 11.1).

### CHAPTER 4: BODILY EXPRESSIVE ROBOT VERSUS FACIAL EXPRESSIVE ROBOT

> Design Question: How to model the five Ekman emotions of anger, fear, happy, sad and surprise in the postures and LEDs of the NAO, so that they are recognizable for children?

Dynamic facial expressions are better to recognize than static expressions in synthetic faces [121]. In a similar way, we expect that dynamic bodily expressions are better to recognize than static ones. The NAO robot is a good platform to study bodily expressions. Unfortunately, it does not have an expressive face, but only some color LEDs for its eyes. Based on literature on dynamic body posture [22, 71, 59] and use of color [122] NAO behaviors were created to model five emotions (anger, fear, happy, sad and surprise). Note that we added the emotion of surprise to extend robot's expressiveness. Of most emotions, except surprise, different versions were created and evaluated. The evaluation was done by 8 participants (mean age 24.6) with the use of signal detection theory [232]. For each emotion the participants were offered a trial of each 12 emotions, so five trials in total. During each trial they had to spot the target (signal) emotion (that was present 4 times) and discard the non-target emotions (noise). This created a signal noise ratio of each (version of the) emotion. Based on this the decision of the selection of the emotions for an experiment with children was made.

> Hypothesis: Three factors influence the recognition rate of robot's emotions: (1) the recognition rate differs between robot embodiments, (2) the rate is higher when the emotions are expressed in a congruent context (compared to no context), and (3) the recognition improves over time.

For the experiment with children we used the selected emotions. Fourteen children of 8 and 9 participated in the experiment. The children interacted both twice with the iCat and twice with the NAO, all within one session. This procedure was repeated a week later. During half of the interactions a story was told on which the robot reacted with an emotion on certain points (within context). During the other half the robots just expressed emotions and said when they moved to the next emotion (without context). Children were asked which emotion was expressed (forced choice). There were two independent variables, 1) robot platform (iCat or NAO) and 2) context (with or without). These variables were randomized in the following manner. Each child had either first the two context conditions (randomized) with the NAO and then the iCat or the other way around. The evaluation showed that the recognizability of the emotions expressed by the iCat and the NAO is similar $F(1,1118)=1.24$, $p=0.27$). Only sadness had a significantly higher recognition rate for the iCat (95%) than for the NAO (68%). Context (semantic congruent information), and familiarity (first versus second time), supported recognition rates significantly, respectively ($F(1.1118)= 29.79$, $p=.00$) and ($F(1.1118)= 18.76$, $p=.00$). So, although their emotional expression modalities are completely different, they are similarly competent in exhibiting recognizable emotions.

#### CHAPTER 5: PHYSICAL VERSUS VIRTUAL EMBODIMENT OF A ROBOT

A recurring question in human-robot research is "Why a robot and not a screen?". In the paper discussed in chapter 5 we tried to answer this question with the expectation the social presence of a physical robot would have added value over a virtual robot.

To make a fair comparison of the embodiment factor of the NAO, it is important that both the physical and virtual NAO act and look the same. By making use of the software that comes with the NAO, we already had a good virtual version of the NAO with the same behaviors as the physical NAO. To make the similarity even greater the virtual NAO was presented on a 30 inch monitor and the speech was produced by the same text-to-speech software. This made the results of the comparison more valuable.

> Hypothesis: Children's performance, attention, trust, enjoyment and preference in quiz task are higher, when interacting with a physical NAO compared to a virtual NAO.

To evaluate the hypothesis an experiment was set up. Children had to play a quiz with questions relating four topics: food pyramid, energy balance, eating healthy and the heart. A topic was completed when three questions were answered correct. They did this quiz twice with one week in between. 11 children with a mean age of 11 participated in the experiment.

The results showed that performance and motivation were not affected by the embodiment (p=0.25; note that a ceiling effect occurred), but the robot did attract more attention (p<0.01). And, when forced to choose, the children preferred the robot over its virtual counterpart (8 over 2). This suggests that the use of a physical robot needs to be carefully considered, use a robot when it is necessary for one of the factors or intended behaviors and use it so it can display it's added capabilities in comparison to a virtual robot (e.g. walking, touching).

In conclusion, the added value of emotion expressions and the use of a physical robot for improved performance and motivation are not substantiated. Some possible conclusions can be drawn from this, for instance; 1) emotion in speech was not good enough, 2) the tasks were not designed to elicit differences between the robot versions on the interaction metrics or 3) The number of participants was too low and the number of variables too high. The recognizability of the emotions was good. Although the emotions were recognized with a forced choice measure, thus a high chance to bet correct, we have enough confidence in the design of the emotions for the NAO to use them in following experiments.

### 11.1.3. PART II: ROBOTS FOR COMPETENCE AND RELATEDNESS

In Part II we looked at three functions that contribute to child's feelings of competence and relatedness towards the robot. Competence and relatedness are two out of the three aspects that contribute to behavior change according to the Self-Determination Theory [74]. The final aspect, autonomy is discussed in Part III



#### CHAPTER 6: INCREASING MOTIVATION BY ADAPTING DIFFICULTY

> Design Question: How to challenge children, aged 9-10, within their dynamic individual capabilities (c.f. Zone of Proximal Development [256] and Optimal Challenge [62]) in a math and memory game with a robot?

The theories of Zone of Proximal Development [256] and Optimal challenge [62] prescribe that for children to stay intrinsically motivated it is important they are challenged, but within their capabilities. To do this we designed a game consisting of two components: making arm movement sequences and solving arithmetic assignments. The robot provided the assignments and gave appropriate feedback depending on the performance. We will only discuss the math game, because this was further developed than the memory movement sequence game. The math game was adapted using a (Bayesian) assessment of the child's individual level and the difficulty of the assignment. The robot could then give the child an assignment based on an educated guess of the level of the child and the difficulty of the assignment.

> Hypothesis: Child's motivation to play a math game with a robot is higher when the game is adapted to his or her dynamic individual capabilities.

The variable was the (non) adaptivity of the arithmetic assignments. Using assignments as the children also get in class, 29 levels (each with 10 assignments) were identified and checked by the teacher. Two versions were created: One working towards a group goal and the other towards a personal goal. In the group version, the level is increased by one after a mistake has been made (after the initial jumps of three levels per correct answer). In the personalized version the level is decreased by one after a mistake and then increased by one, to make sure the correct level is reached. Furthermore, children in the personalized version can reach level 29, while the children in the group version can reach level 20 (which is the correct level for this class).

The implementation shows an example of a translation of the theories of Zone of Proximal development and Optimal Challenge [256, 62]. The evaluation investigates whether the intended behavior is invoked.

Twenty children aged 9-10 participated in the experiment, the group was divided in two over the two conditions. The experiment consisted of three sessions of 20 minutes in which we kept track of progress and the motivation shown with a free-choice period of 5 minutes. During the free-choice period they were free to proceed with the assignments together with the robot or do something fun like reading a comic.

The results showed that the children who were challenged according to their performance were able to progress further with the personalized assignments and were more motivated to keep interacting with the robot as was shown with the free-choice period.

#### Chapter 7: Reciprocal emotion elicitation

> Design question: How to model robot's emotional expressions that represent: robot's current performance, match child's intro-extroversion trait, and adapt to child's performance and emotional state?

11

As can be seen in this design question many factors are taken into account to adapt robot's emotion. This means design decisions are made regarding each factor. Intro-extroversion is taken as a scale, because research has shown that people like a robot similar in personality [115]. An extrovert robot makes faster and bigger gestures. The child's emotion should also influence the robot's emotions, this emotion can be recognized from the child his/herself (by the wizard) or by the performance. Decisions were made in how far the robot mirrored the emotion, simulating emotional contagion [157]. The robot's valence and arousal model was adapted based on the inputs (robot and child performance, child's valence and arousal and child's intro-extroversion trait). Speech volume and gesture size were influenced by the extroversion parameter, but also by the arousal of the child (that influenced the arousal of the robot). The arousal furthermore influenced the eye color, head position, gesture movement, speech rate and speech fundamental frequency. The valence of the child influenced eye color, trunk position, head position and gesture movement. Next to this, there were emotional occurrences, for instance, after a sequence of correct answers by either child or robot, where the robot was simply happy (or another basic emotion). All inputs together led to arousal valence values for the robot that could be translated to a weighted behavior. By formalizing the rules and dependencies and keeping track of the design decisions this model is easy to adapt according to the evaluation results.

> Hypothesis: A robot with adaptive emotional expressions will "score higher" on relatedness factors in both behaviors (emotional expressivity of the child) and opinion (fun, acceptance, empathy, trust, preference and recognized emotional expressivity) in comparison to a robot without adaptive emotional expressions.

Children had more positive expressions when interacting with the adaptive emotional robot. The opinions of the children did not differ significantly for the different robot versions (the score was high in general). When they were forced to choose it was noted that the adaptive robot scored higher on empathy and emotion, but lower on trust, reminding of chapter 3, where the unintelligibility of the emotional speech had a deprecation of trust as result. The children particularly liked emotion through movement.

#### CHAPTER 8: STIMULATING MUTUAL SELF-DISCLOSURE

> Design question: How to design, within the context of a diabetes diary, self-disclosure and empathetic behavior by a robot based on mutual self-disclosure (e.g. [202]) and empathy theories [67]?

In the context of a diary, self-disclosure on a high intimacy level is not strange. Research has shown that when people engage in self-disclosure, they also expect the other to reciprocate this, at the same level of depth and breadth [2] People also like and trust other people that engage in self-disclosure more [56]. Other research shows that robots are better liked when they disclose affective information than when they only talk about tasks [228]. Empathy also supports self-disclosure by the other, as is often used in patient-centered therapy [205]. The robot was therefore implemented with self-disclosure dialog acts, e.g. about its favorite pets, and with empathetic behaviors, e.g. reacting on a bad day with "I'm sorry to hear that".

> Hypothesis: Empathetic behaviors and self-disclosure of a video-conferencing robot improve children's adherence to fill out their diabetes diary.

The results showed that a robot that provides self-disclosure and empathy behavior had a positive effect on adherence (50% vs. 85% fully completed). Older and younger children differed in their approach of the robot. Younger children were more interested in the robot. All children bonded with the robot and found it trustworthy and human-like. They really liked the sharing of information and the direct feedback by the robot. These results provide a contribution to the self-management of the diabetes [114].

### 11.1.4. PART III: ROBOTS FOR AUTONOMY

In the final part, Part III, different activities are combined in an integrated setup (over multiple sessions). The feeling of autonomy is addressed in this part, while keeping the functions of the previous parts that attributed to relatedness and competence.

#### CHAPTER 9: BEHAVIORS FOR THE ICAT TO DISPLAY DIFFERENT ROLES

> Design question: How to create behaviors for a moderate expressive [253] iCat robot based on Motivational Interviewing [207] techniques?

Three roles were defined that each had another combination of motivation interviewing techniques. These roles were motivator, educator, and (game) buddy. These three roles each link to one of the three factors relevant for behavior change according to self-determination theory [74]: autonomy, competence and relatedness. Based on Motivational Interviewing techniques, behaviors for three skills were implemented for the motivator; reflective listening (i.e. reacting according to situation), positive regard (i.e. provide compliments, don't punish) and attentiveness (i.e. look at the child, nod). For the text interface only positive regards was possible. In the educator role the positive regard from the motivator role is used, extended by providing the correct answer after an incorrect answer was given. This role was the same for all three interfaces. The game buddy made use from knowledge from gaming [62] to make a game hard, but not too hard. Furthermore, it used the motivator behaviors (that differed between the iCat interfaces and the text interface), it thus provided compliments, reacting according to the performance of the child and asked the child for preferences. The explicit relation between methods and implementation makes it easier to describe expectations and evaluate these.

> Hypothesis: Text, virtual and physical robot are for children, in an incremental order, increasingly motivating and educating. This can be explained by the incremental number of motivational interviewing techniques that can be implemented in the different interfaces.

In the experiment the children interacted with all three roles, the motivational role was interaction focused, the educational role was a quiz supported by educational movie clips and the (game) buddy role was the robot playing tic-tac-toe with the children. The children proved to value the support roles positively, in particular the buddy role. Objective and subjective data showed that they highly appreciated both the physical and virtual characters (more than the text interface). Furthermore, children proved to interact faster with the character than with the text interface. There is a clear added value of robots compared to conventional text interfaces, which can be utilized for improving self-management.

### CHAPTER 10: EVALUATING IN THE WILD

> Design question: What does experimentation in the wild add over controlled experiments that test isolated components of the robot one-by-one in a lab environment?

All the knowledge gathered during the ALIZ-e project was integrated in the final experiment. Different functions that showed a positive trend in previous research were implemented in this experiment. To support autonomy there were two activities: a quiz and a sorting game. A feeling of competence was supported by a robot that made errors and thought about its answers. Relatedness was supported by implementing emotional expressiveness in the robot, adapting these emotions to the emotions of the child, remembering small facts of the child by the robot, the robot disclosed information about itself, the children could touch and walk with the robot, and finally the setting was confined, making a world for the robot and the child. During the experiment we noticed some unexpected events, e.g. a robot falling and thereby increasing the feeling of competence of the child, or a child not stopping with talking or being so open about his/her diabetes that a short "I don't understand" was not ethically feasible. We could observe these unexpected events because of the "in the wild" setting of the experiment.

> Research question: Is the complete system appreciated by children with diabetes, after multiple interactions, on the factors: autonomy, competence and relatedness.

The results showed that the interactions had a positive effect on competence and children's feeling of relatedness towards the robot. They also experienced a feeling of autonomy because of the activity choice. The amount of choice was increased with each session and in concurrence with this, the feeling of autonomy also increased over the sessions. None of the results was significant, because of the small number of children and the high variety between them.

> Research question: Does performing an experiment in the hospital with the real target users increases acceptation of all involved (children, parents and health care professionals).

Children, parents and health care professionals were enthusiastic and provided us with all kinds of advice and new requirements, because we really involved them they felt taken seriously and were more accepting of faults in the system.

## 11.2. Social robots for child's behavior change: contributions and limitations

There is a lack of research on long-term child-robot interaction in realistic environments [18]. The ALIZ-e and PAL projects study child's experiences in repeated interactions with a social robot in the field. The design and evaluation activities provided new insights on the models and methods for behavior change of children, and the domain and operational demands and constraints for the desired behavior change. It should be noted that more research on specific aspects is needed (for some effects there is not yet a decisive conclusion)

### 11.2.1. The situated Design Rationale

The situated Design Rationale provides a refinement of the situated Cognitive Engineering methodology [172]. It describes a concise and coherent specification of the design rationale. Both the concepts and their relationships are specified and visualized. Furthermore, it supports incremental design and evaluation. There is a step-by-step progress in a manner that ensures keeping and reusing design solutions, decisions on design choices and evaluation methods. The method has been applied on the 4-year European project ALIZ-e resulting in one integrated overview of the project sDR with for each Part in this thesis a depiction of the specific sDR.

The sDR provides a tool for recognizing which concepts need to be refined based on the number of relations between that concept and other concepts. See `https://bit.ly/2RXxWNd`, for the complete sDR of this thesis. Figure 11.2 shows the analysis of the relations between these concepts. The x-axis represents the number of relations and the y-axis the number of concepts. There is one concept that has 21 relations, this is "Exhibit social behavior". The high number of relations is a strong indication this concept is too generic to be informative. Next to "Exhibit social behavior" there are 13 other concepts with 10 or more relations to other concepts. Based on experience and knowledge there is a need for a threshold of when the concept is specified enough and neither over- or underspecified.

This will support the use in the HRI community for structured (situated) theory development.

### 11.2.2. Evaluation methods for children

During the project we learned that it was not enough to use evaluation instruments especially developed for children (e.g. the fun toolkit [199]), to get the input we needed. We also needed questionnaires that looked at the different aspects of self-determination theory, and our implementations. This meant translating adult questionnaires to language that was understandable for children, and developing specific questions for the things we were interested in. Soon we noticed that we wanted to know more from the children than they had the attention span for. This lack of attention span resulted in automatically filling in the questionnaires, all answers "1" for instance. Furthermore, we had children having difficulty with reading the questionnaires and very creative children that made in-between options in Likert scales. All this resulted in looking for methods to decrease the questionnaire load, while still getting the results we needed to adapt the

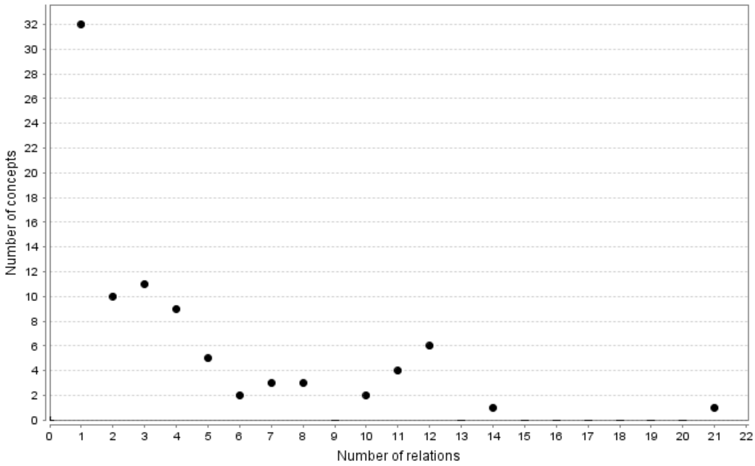



Figure 11.2: Depiction of the number of concepts that have a certain number of relations.

implementation. This meant more use of pictures, tangible questionnaires (to decrease the default answering) and other methods than questionnaires to measure important factors, because these all suffered from ceiling effects. One method, the free choice period [248, 111], in which children have full autonomy to choose between playing with the robot or doing something else, was deemed especially useful. Further research is needed, for instance in the use of games to reduce the use of psychometric instruments [189].

### 11.2.3. REPEATED MEASURES IN THE FIELD

By doing repeated evaluations in the field we got results we otherwise would not have had. The intrinsic motivation of children with diabetes to participate is for instance higher than that of their peers without diabetes. We also saw that their motivation to interact with the robot seemed higher, they really saw it as something positive related to their disease. This resulted, according to professional caretakers, in a lower threshold to visit the hospital and be open towards the professional caretakers even after the evaluations were finished.

The repeated interactions showed that it is very challenging to keep the interaction interesting with enough variety, personalization and (content of) activities. Because the setting was less controlled than a lab setting, there were quite some specific interaction errors, but many did not seem to have a detrimental effect on the overall interaction. The children saw it as an opportunity to help the robot.

### 11.2.4. EMOTION MODELS



Emotion models for different embodiments were developed and evaluated to be recognizable and invoke the desired behavior. These models were based on human emotion theories and tailored to specific robot platforms, making use of their strengths

For the iCat the emotion expression model was based on on theories on facial expressivity by humans (e.g. [82, 229] and emotion expressivity in synthetic speech (e.g. [43, 167, 220]).

For the NAO it was shown that human posture emotion research (e.g [22, 59]) can be transferred to the limited posture capabilities of the NAO while keeping recognizable emotions. The emotion expression was further enriched by using the ability of the NAO to show colored eyes, in research from [122] and [37] colors are connected to emotions and [113] used the colors of the NAO's eyes in earlier research to convey recognizable emotions.

In chapter 5 the differences between a virtual versus physical presence and the influence on intended behavior was evaluated. This is related to theories on social presence (e.g. [227, 151]). Lowenthal says that there is a continuum of social presence where there is on one side a feeling of relatedness and on the other end the feeling if someone/something is perceived as 'real' or having agency. The results shows that both the virtual and physical robot were seen as social actors, but perhaps the physical robot as having more agency.

Emotions expressed by the robot influenced how a child perceived the robot. It was seen as more social than a robot without emotions. This provides evidence that, similar to human-human interaction [161], a robot that expresses emotions is attributed social behavior and is easier to relate with. It is accepted more, deemed more fun and perceived as more empathetic. This can be enhanced by having the interactions in context and getting familiar with the robot over multiple interactions. And is further supported by having agency [151].

### 11.2.5. USE OF SELF DETERMINATION CONCEPTS

We based the design of supportive robot behaviors on human behaviors from behavior change theories, like we based emotional robot expressions on human emotion expression theories. The main theory we used was the Self-determination theory (SDT) [74]. This theory says that three things should be supported by behaviors of the therapist, or in our case the robot, which are: competence, autonomy and relatedness. To develop techniques to implement the behaviors matching these three factors we looked at methods from motivational interviewing [161], Zone of proximal development [255] and flow [62]. Methods from motivational interviewing support feelings of relatedness towards the therapist/robot by acknowledging, self-disclosure and mirroring, but also by giving choices, which also supports autonomy. A part of motivational interviewing is also setting achievable short term goals, the methods from zone of proximal development and flow further explicate how this can be done.

### 11.2.6. LIMITATIONS

**All evaluations** were limited by the participant numbers. Not many children participated, and the experiments were not long term. Because of this limited number of children we sometimes made extra efforts to involve them (call them), which made sure they participated, but confounded our results. Decisive conclusions on our hypotheses regarding the robot functions and their design can therefore not be made based on the current evaluation results.



Some comparisons, for instance between virtual and physical, were quite artificial, because the virtual and physical robot were programmed to be as similar as possible thereby not making full use of the different added values both can have (e.g. easier moving vs. ability to interact with the real environment). In chapter 6 the difference between the adaptive difficulty condition and non adaptive difficulty condition was an artificial difference that might have limited the outcomes. The adaptive difficulty condition was able to proceed far beyond the class level, while the non adaptive condition stopped at the class level. In a real class situation children who excel might not be challenged with exercises that are adaptive to their performance, but often they are provided with exercises that exceed the class level.

Sometimes the technical limitations had a negative influence on the outcomes. The emotional voice, in chapter 3 and chapter 7, was hard to understand and this influenced the feeling of trust towards the robot negatively.

**Content scarcity** resulted in difficulties to really personalize the educational content of quizzes and exercises regarding both subject and difficulty. This can be improved with more content. A complicating factor with making this content for diabetes is that many questions don't have a right or wrong answer, but an answer that is person and context specific.

**sDR** complexity increases fast. This is inherent of a tool that depicts the relations between all different concepts. Furthermore, the sDR, as described in this thesis, is derived from the current project by project members. The re-usability needs to be shown in other projects and by outside users. Because the need to extend and refine the sDR, a tool is needed. A first version is designed with the use of Cytoscape[1] and it would be good to integrate an improved tool within the sCE tool[2].

**NAO limitations** meant that not all desired behaviors and functionalities could be implemented or did not fit well. The NAO has no face, has a limited behavior repertoire, is slow, has a plastic body, etc.. This means the NAO can not show facial expressions, must keep its movements within it's balance point, can not easily go from one point to another and is not nice to cuddle with. Therefore we could not benefit optimally from the added value of a *physical* robot.

## 11.3. FUTURE WORK

**The situated Design Rationale** is applied to a specific domain and user group: self-management improvement of diabetic children in the age of 6-10. The main challenge for future work in this direction is to extend the use to other domains. We also would like to work in the future on the usability of sDR, there is a need to keep track of all the relations without losing usability. At the moment the sDR is either developed within yED[3] or Cytoscape[4], both have their pros and cons. In the future it would be nice to have



[1] http://www.cytoscape.org/
[2] http://www.scetool.nl
[3] https://www.yworks.com/products/yed
[4] http://www.cytoscape.org/

a tool that supports the visualization of an sDR in a user friendly (creating and viewing) manner that is integrated in the situated Cognitive Engineering tool[5].

An advantage of using sDR, is that the re-usability over the course of a project is high, objectives (should) stay the same, just as most of the related methods, effects and instruments. The functions and related interaction design patterns will change, but they should be related to the effects and methods that contribute to the objectives. An opportunity we see here is to connect software code to functions and ontologies to a whole project so that other researchers can reproduce the exact same experiment.

Another opportunity related to the re-usability over the course of a project, is the re-usability over projects. Because the objectives related effects of behavior change will not be different in different domains and for different user groups. The methods to reach these objectives and therefore the needed functions can, for instance, differ based on human factor aspects related to the user group (e.g. elderly, disabled children).

Next to behavior change we see opportunities for the use of sDR in entirely other domains where complex ICT solutions are designed and evaluated to reach certain objectives. An example of this, is for instance, the design and evaluation of a robot that supports social team cohesion in an urban search and rescue domain setting. The objectives and envisioned effects are different, mainly coming from group process research, as are the related methods, functions and interaction design patterns. However, sDR and the resulting concept map do support the design and evaluation of such a system.

**The evaluations** were limited in reliability because of the low participant numbers. We saw that doing evaluations out of the classroom and in the field provide much of the needed information, but decreases participant numbers even more, especially when taking into account interpersonal differences (time since diagnosis, age, current knowledge, personality etc.). In the future we would like to extend the number of hospitals participating in the evaluations to increase the number of participants. Another option is to extend the system with more activities and content so that children are more motivated to use the system for a longer period of time. This would make it possible to provide updates of the system during the evaluation, creating a within-subject setting. Finally, some general principles could also be evaluated with children that don't have diabetes, but have another chronic illness for which self-management is crucial. In comparison to healthy children, these children are expected to have, like the children with diabetes, a more intrinsic motivation to use the system.

**Social robots** are used within this thesis. The expectation is that the number of social robots will increase and that they will get cheaper. At the moment, one of the partner hospitals (Ziekenhuis Gelderse Vallei) has acquired NAO robots to use in their current practice. It would therefore be interesting to see if knowledge acquired within this thesis on the NAO and iCat is transferable to the daily practice and other robots. A quick look online shows that many social robots are created (e.g. Everest from Abilix[6], or Buddy from blue frog[7]). These robots are less expensive than the NAO and for the applica-



[5]www.scetool.nl

[6]http://en.abilix.com/index.php/robot/everest

[7]http://www.bluefrogrobotics.com/en/buddy-your-companion-robot/

tion environments (home, hospital, school) these costs are of utmost importance. It is therefore relevant to look into the transferability of knowledge to these cheaper robots. Furthermore, it is of interest to see why perhaps the more expensive robots are needed. What is it in their behavior that has added value? Or, can it also be done by a virtual agent? Which makes it even cheaper? Or is there an added value of using a social robot whenever possible and a virtual agent in other cases? Within the PAL project[8], one of the research questions is about transferability of relatedness towards a robot to its virtual counterpart.

But transferability of knowledge to other robots and agents is not the only thing of interest for the future of social robots. One of the main to be afraid of at this moment is that we will disappoint the general public, that has unrealistic expectations of what a robot can and cannot do [143]. An Asimo[9] or a documentary on a socially interactive robot[10] makes that people are unimpressed of what a robot can do when it is used in a real life setting without researchers in the vicinity. As the social robot research community we should make the general public more aware of the possibilities, but more important the restrictions of robots at the moment. Of course we want to sell our vision, but we should be careful that it is a realistic vision, otherwise people won't want to work with robots anymore when they finally can do what they envisioned.



[8] www.pal4u.eu
[9] http://asimo.honda.com/asimotv/
[10] http://bit.ly/2q8JfEP

# Dankwoord

Eindelijk, het einde is er nu echt. Wat ooit begon met als stip op de horizon, een wetenschappelijke carrière, is uitgemund in een lange barre tocht waarbij de wetenschappelijke carrière als stip is verdwenen. De nieuwe stip is nog onbekend en dat is misschien dan wel het belangrijkste wat ik geleerd heb. Het leven is niet altijd maakbaar en te plannen, vaak komen er gewoon dingen op je pad en dat wat op een afstand leuk lijkt is dat niet altijd in het echt. Ergens aan proeven en dan besluiten of het wel of niet voor jou is, is een goede manier om nieuwe dingen aan te gaan.

Het langer doen over een promotie heeft zeker voordelen. Ik zou niet zo tevreden zijn over de inhoud van mijn boekje als ik verder was gegaan met het eerste idee dat ik had.

Zonder Mark, mijn afstudeerbegeleider, mentor, promotor en geweldige collega was ik hier nooit gekomen. Aan het begin van mijn promotietraject gaf ik al aan dat ik mijn kinderwens niet zou uitstellen, maar we hadden beiden denk ik niet gedacht dat het zo snel zou gaan en er zelfs twee zouden komen. Mark, ik ben enorm dankbaar voor je relativering, je complimenten en je beschikbaarheid in nood.

Koen je bent wat later bij dit proefschrift betrokken, maar zonder jouw input was het niet dit verhaal geworden. Aan het begin hadden we nog wat moeite met afstemmen, omdat het bij mij natuurlijk trager ging dan je gewend was en ik ook niet altijd naar Delft kon komen vanwege werkzaamheden bij TNO. Later in het traject is dit helemaal goedgekomen en was ik erg blij met jouw pragmatische blik.

Beste promotiecommissie heel erg bedankt voor jullie tijd. Vanessa en Tony weten hoe lang ik hier mee bezig ben geweest en hoe blij ik ik ben dat het proefschrift hier ligt. Tony was de projectleider van Aliz-e en heeft dus veel van dichtbij meegemaakt.

De inhoud van dit proefschrift was niet tot stand gekomen zonder de inbreng de studenten die hun onderzoek voor het ALIZ-e project gedaan hebben bij TNO. Ik gebruik in dit proefschrift werk van Melanie, Iris, Anna, Chrissy, Myrthe, Esther, Vincent en Johanna. Het is leuk te horen dat een aantal van jullie al gepromoveerd zijn en zelfs al kinderen hebben. Naast deze groep was er nog een groep stagiairs en natuurlijk het ALIZ-e team in zijn geheel die gezorgd hebben voor de achtergronden en diepgang van dit proefschrift.

Een onvervangbaar persoon binnen het ALIZ-e en PAL-team was Bert Bierman. Bert zonder jouw zijn ontelbare experimenten gered door een laatste hack of snelle telefonische hulp. Daarnaast heb je met je luisterend oor en goede inzichten mij geholpen mijn prioriteiten goed te zetten.

Twaalf jaar heb ik met veel plezier bij TNO gewerkt, en dit komt zeker door mijn lieve directe collega's. Ik ga jullie niet met naam noemen, want dan mis ik er vast 1 of 2. Veel gekletst, gewandeld en altijd vol enthousiasme als er weer een nieuw speeltje was. Het is heel fijn om in een omgeving te werken, waar je opbouwende kritiek en complimenten krijgt en waar altijd een luisterend oor beschikbaar is.

Mijn kamergenootje Jacomien en mijn donderdagkamergenootje Rifca wil ik nog wel even noemen. Jacomien, ik denk dat we bijna 10 jaar bij elkaar op de kamer hebben gezeten. Ik vond de openheid tussen ons erg fijn. Rifca, jij kwam pas later, maar het was fijn om samen met jou aan PAL te werken, of niet te werken :P en gezellig te kletsen over leuke dingen en frustraties gerelateerd aan PAL en promotie.

Promoveren bij TNO is niet de standaard en daarom ben ik ook erg blij dat ik deze mogelijkheid heb gekregen en ook ondersteund ben in mijn keuze hier tijd voor vrij te maken. De combinatie van TNO met promotie is zwaar, maar ik ben nog steeds blij dat ik het op deze manier heb gedaan en niet vier jaar bij een universiteit. De afwisseling met mijn andere projecten heeft zeker dit proefschrift positief beïnvloed. Gedurende mijn promotie heb ik verschillende afdelingshoofden gehad, Adelbert, Myra, Patrick, Jasper en Sanne hebben mij gesteund in mijn beslissing, maar me ook keuzes gegeven.

Buiten TNO wil ik Geert-Jan en Ivana bedanken. Bij mijn eerste conferentie heb ik Geert-Jan leren kennen en voor ik het wist deden we twee Europese projecten samen. De eerste keer in Saarbrücken werd ik ook nog geweldig ontvangen door zijn vrouw Ivana, met wie ik daarna ook met veel plezier heb samengewerkt. Maar niet alleen in werk, maar ook privé sfeer zijn de gesprekken, emails en (netwerk :P) chats altijd erg fijn.

Zou ik nou wel of geen paranimfen doen? Daar heb ik lang over getwijfeld, maar uiteindelijk heb ik gekozen voor 2. Nanja, wij zitten in hetzelfde schuitje en ik hoop dat ik je net zo vaak heb kunnen helpen als jij mij. Carianne, jij bent mijn oudste en liefste vriendinnetje, al vanaf de kleuterschool kennen we elkaar. Ook al zien we elkaar niet vaak, we kunnen het altijd over alles hebben. Lief vriendinnetje, ik hoop dat we elkaar altijd weten te vinden.

Dan mijn familie, zonder mijn ouders en zusjes was ik niet geworden wie ik ben. Ik weet dat jullie vinden dat ik te ver weg woon, en ik mis jullie regelmatig. Maar ik weet dat als ik jullie nodig heb, dat jullie er dan meteen voor me zijn. Ik hou van jullie.

Sinds ik aan dit proefschrift begonnen ben, heb ik een mooi gezin gekregen. Anton, mijn rots in de branding was er al. Als ik deze basis thuis niet had gehad, dan had er vandaag een heel ander persoon voor jullie gestaan. Eentje met veel minder geluk in haar leven. Anton, mijn lieve lijfje, je bent er altijd voor mij. Zowel emotioneel als praktisch, zonder Anton was er geen interactieve digitale versie van de conceptmap geweest en had iemand anders mijn tekst nogmaals moeten controleren.

En natuurlijk mijn lieve Olivia en Elias, de spiegel die jullie voorhouden is regelmatig confronterend, maar ook zeker net zo motiverend om dingen waar ik al mijn halve leven tegen aan loop goed aan te pakken. Helaas zijn jullie niet gemakkelijk voor de gek te houden, want jullie geloven het gewoon niet als ik zeg dat ik zo lang over mijn promotie gedaan heb zodat jullie bij de verdediging mogen zijn. Ik ben gek op mijn 3 geweldige liefjes.

# Curriculum Vitæ

## Rosemarijn LOOIJE

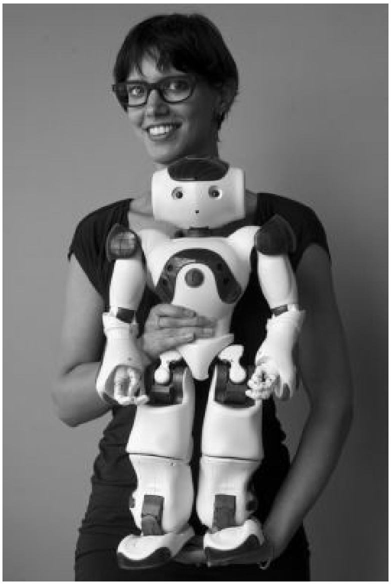


photo: Guus Dubbelman / de Volkskrant

**Rosemarijn Looije** (born November 18, 1982 in Haarlem) received her secondary education (VWO) at the Kennemer Lyceum Overveen from 1995 until 2001. She then studied Artificial Intelligence, with a specialization in Man-Machine-Communication, at the University of Groningen and received her Masters' title in 2006. Her graduation project, which she did at TNO Human Factors (Soesterberg), put her on the path of social robots and diabetes. The following 12 years she proceeded to extend this research in her work at TNO and for her dissertation. The main questions being throughout, how should robots and people work together to reach their shared goals, and what effect has a robot on the people it works with and their surroundings? Currently, she is employed at the University Medical Center Utrecht. Rosemarijn lives in Amersfoort with her partner Anton and their two children, Olivia and Elias.

# List of Publications


9. R. Looije, M.A. Neerincx, K.V. Hindriks, *Specifying and testing the design rational of social robots for behavior change in children*, Cognitive Systems research, 43 (pp250-265).

8. R. Looije, M.A. Neerincx, J.K. Peters, O.A. Blanson Henkemans, *Integrating Robot Support Functions into Varied Activities at Returning Hospital Visits*, International Journal of Social Robotics **8**, 483 (2016).

7. I. Cohen, R. Looije, and M. Neerincx, *Child's perception of robots emotions: effects of platform, context and experience*, International Journal of Social Robotics **6**, 507 (2014).

6. M. Tielman, M. A. Neerincx, J. J. Meyer & R. Looije, *Adaptive emotional expression in robot-child interaction*, in Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction (pp. 407-414). ACM 2014.

5. E. J. Van Der Drift, R. J. Beun, R. Looije, O. A. Blanson Henkemans & M. A. Neerincx, *A remote social robot to motivate and support diabetic children in keeping a diary*, in Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction (pp. 463-470). ACM 2014.

4. R. Looije, A. Van der Zalm, M. Neerincx, R.-J. Beun, et al., *Help, I need some body: the effect of embodiment on playful learning*, in IEEE Ro-Man 2012 (pp. 718-724).

3. J. B. Janssen, C. C. van der Wal, M. A. Neerincx & R. Looije. *Motivating children to learn arithmetic with an adaptive robot game*, ICSR conference 2011 (pp. 153-162).

2. J. Kessens, M. Neerincx, R. Looije, M. Kroes, and G. Bloothooft, *Facial and vocal emotion expression of a personal computer assistant to engage, educate and motivate children*, in Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009.

1. R. Looije., M.A. Neerincx, V. de Lange *Children's responses and opinion on three bots that motivate, educate and play*, in in Journal of Physical Agents, **2**, 2 (pp. 13-20), 2008.

# References

[1] ALIZ-e (2012). Aliz-e website. http://www.aliz-e.org/.

[2] Altman, I. and Taylor, D. A. (1973). *Social penetration: The development of interpersonal relationships.* Holt, Rinehart & Winston.

[3] Aronson, E., Wilson, T., and Akert, R. . S. P. P. E. N. J. (2005). *Social Psychology*. Pearson Education, New Jersey.

[4] Assor, A., Kaplan, H., and Roth, G. (2002). Choice is good, but relevance is excellent: Autonomy-enhancing and suppressing teacher behaviours predicting students' engagement in schoolwork. *British Journal of Educational Psychology*, 72(2):261–278.

[5] Banse, R. and Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 70(3):614–636.

[6] Barakova, E. I. and Lourens, T. (2010). Expressing and interpreting emotional movements in social games with robots. *Personal and ubiquitous computing*, 14(5):457–467.

[7] Barrett, L. F., Lindquist, K. A., and Gendron, M. (2007). Language as context for the perception of emotion. *Trends in cognitive sciences*, 11(8):327–332.

[8] Bartneck, C. (2003). Interacting with an embodied emotional character. In *Proceedings of the 2003 international conference on Designing pleasurable products and interfaces*, pages 55–60. ACM.

[9] Bartneck, C. and Forlizzi, J. (2005). A design-centred framework for social human-robot interaction. In *ROMAN 2004*, pages 591–594. IEEE.

[10] Bartneck, C., Kulić, D., Croft, E., and Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, 1(1):71–81.

[11] Bartneck, C., Reichenbach, J., and Breemen, A. (2004). In your face robot! the influence of a character's embodiment on how users perceive its emotional expressions. In *In Proceedings of the Design and Emotion, Ankara.*

[12] Baxter, P., Belpaeme, T., Cañamero, L., Cosi, P., Demeris, Y., and Enescu, V. (2011). Long-term human-robot interaction with young users. In *IEEE/ACM Human-Robot Interaction 2011 Conference (Robots with Children Workshop).*

[13] Beck, A., Cañamero, L., Bard, K., et al. (2010a). Towards an affect space for robots to display emotional body language. In *Ro-man, 2010 ieee*, pages 464–469. IEEE.

[14] Beck, A., Cañamero, L., Damiano, L., Sommavilla, G., Tesser, F., and Cosi, P. (2011). Children interpretation of emotional body language displayed by a robot. In *ICSR 2011*.

[15] Beck, A., Hiolle, A., Mazel, A., and Cañamero, L. (2010b). Interpretation of emotional body language displayed by robots. In *Proceedings of the 3rd international workshop on Affective interaction in natural environments*, pages 37–42. ACM.

[16] Beer, J., Fisk, A., and Rogers, W. (2009). Emotion recognition of virtual agents facial expressions: The effects of age and emotion intensity. In *Human Factors and Ergonomics Society Annual Meeting Proceedings*, volume 53, pages 131–135. Human Factors and Ergonomics Society.

[17] Bell, A. (2007). Designing and testing questionnaires for children. *Journal of Research in Nursing*, 12(5):461–469.

[18] Belpaeme, T., Baxter, P., De Greeff, J., Kennedy, J., Read, R., Looije, R., Neerincx, M., Baroni, I., and Zelati, M. C. (2013). Child-robot interaction: Perspectives and challenges. In *Social Robotics*, pages 452–459. Springer.

[19] Belpaeme, T., Baxter, P. E., Read, R., Wood, R., Cuayáhuitl, H., Kiefer, B., Racioppa, S., Kruijff-Korbayová, I., Athanasopoulos, G., Enescu, V., et al. (2012). Multimodal child-robot interaction: Building social bonds. *Journal of Human-Robot Interaction*, 1(2):33–53.

[20] Beran, T., Ramirez-Serrano, A., Kuzyk, R., Fior, M., and Nugent, S. (2011). Understanding how children understand robots: Percieved animism in child-robot interaction. *International Journal of Human-Computer Studies*, 69:539–550.

[21] Beun, R., Vos, E. D., and Witteman, C. (2003). Embodied conversational agents: effects on memory performance and anthropomorphisation. pages 315–319. Springer.

[22] Bianchi-Berthouze, N. and Kleinsmith, A. (2003). A categorical approach to affective gesture recognition. *Connection science*, 15(4):259–269.

[23] Bickmore, T. and Cassell, J. (1999). Small talk and conversational storytelling in embodied conversational interface agents. In *AAAI fall symposium on narrative intelligence*, pages 87–92.

[24] Bickmore, T., Gruber, A., and Picard, R. (2005). Establishing the computer–patient working alliance in automated health behavior change interventions. *Patient education and counseling*, 59(1):21–30.

[25] Bickmore, T., Schulman, D., and Yin, L. (2010). Maintaining Engagement in Long-term Interventions with Relational Agents. *Applied Artificial Intelligence*, 24(6):648–666.

[26] Bickmore, T. W. and Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 12(2):293–327.

[27] Bigelow, J. H., Cretin, S., Solomon, M., Wu, S.-Y., and O'Connell, M. (2000). Patient compliance with and attitudes towards health buddy. Technical report, DTIC Document.

[28] Blanson Henkemans, O., Fisk, A., Lindenberg, J., van der Mast, C. A., Neerincx, M., and Rogers, W. (2007). A computer assistant for supervision of self-care and improvement of health literacy. In *115th Annual APA Convention*. Washington DC: APA.

[29] Blanson Henkemans, O. A., Bierman, B. P., Janssen, J., Looije, R., Neerincx, M. A., van Dooren, M. M., de Vries, J. L., van der Burg, G. J., and Huisman, S. D. (2017a). Design and evaluation of a personal robot playing a self-management education game with children with diabetes type 1. *International Journal of Human-Computer Studies*, 106:63–76.

[30] Blanson Henkemans, O. A., van der Boog, P. J., Lindenberg, J., van der Mast, C. A., Neerincx, M. A., and Zwetsloot-Schonk, B. J. (2009). An online lifestyle diary with a persuasive computer assistant providing feedback on self-management. *Technology and Health Care*, 17(3):253–267.

[31] Blanson Henkemans, O. A., Van der Pal, S., Werner, I., Neerincx, M. A., and Looije, R. (2017b). Learning with charlie: A robot buddy for children with diabetes. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pages 406–406. ACM.

[32] Blanson Henkemans, O. A., van Empelen, P., Paradies, G. L., Looije, R., and Neerincx, M. A. (2015). Lost in persuasion a multidisciplinary approach for developing usable, effective, and reproducible persuasive technology for health promotion. In *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2015 9th International Conference on*, pages 49–56. IEEE.

[33] Boman, K. K., Viksten, J., Kogner, P., and Samuelsson, U. (2004). Serious illness in childhood: the different threats of cancer and diabetes from a parent perspective. *The Journal of pediatrics*, 145(3):373–379.

[34] Borchers, T. (1999). *Self-disclosure and Openness*.

[35] Borghouts, C., Buter, A., Dekker, J., Hoogenberg, E., Kopmels, D., and van Oostenbrugge, M. (2005). *Interactie in Rekenen*. Bazalt Educatieve Uitgaven.

[36] Borkenau, P. and Liebler, A. (1992). Trait inferences: Sources of validity at zero acquaintance. *Journal of Personality and Social Psychology*, 62(4):645–657.

[37] Boyatzis, C. J. and Varghese, R. (1994). Children's emotional associations with colors. *Journal of Genetic Psychology*, 155(1):77–85.

[38] Brave, S., Nass, C., and Hutchinson, K. (2005). Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent. *International journal of human-computer studies*, 62(2):161–178.

[39] Breazeal, C. (2003). Toward sociable robots. *Robotics and autonomous systems*, 42(3):167–175.

[40] Burkhardt, F. (2005). Emofilt: the simulation of emotional speech by prosody-transformation. In *Interspeech*, pages 509–512.

[41] Burki, T. (2015). Robots and type 1 diabetes. *The lancet. Diabetes & endocrinology*, 3(11):844–844.

[42] Butler, E. A., Egloff, B., Wilhelm, F. H., Smith, N. C., Erickson, E. A., and Gross, J. J. (2003). The social consequences of expressive suppression. *Emotion*, 3(1):48–67.

[43] Cahn, J. E. (1989). Generating expression in synthesized speech. Master's thesis, MIT.

[44] Cameron, D., Millings, A., Fernando, S., Collins, E. C., Moore, R., Sharkey, A., Evers, V., and Prescott, T. (2018). The effects of robot facial emotional expressions and gender on child–robot interaction in a field study. *Connection Science*, pages 1–19.

[45] Cañamero, L. and Fredslund, J. (2001). I show you how I like you - can you read it in my face? *IEEE Transactions on Systems MAN and Cybernetics*, 31(5):454–459.

[46] Cañamero, L. and Lewis, M. (2016). Making new "new ai" friends: Designing a social robot for diabetic children from an embodied ai perspective. *International Journal of Social Robotics*, pages 1–15.

[47] Cassell, J. (2000). *Nudge Nudge Wink Wink: Elements of Face-to-Face Conversation for Embodied Conversational Agents*, chapter Introduction, pages 29–63. MIT Press.

[48] Cassell, J., Vilhjálmsson, H. H., and Bickmore, T. (2001). Beat: the behavior expression animation toolkit. In *SIGGRAPH 2001*.

[49] Chiasson, S. and Gutwin, C. (2005). Design principles for children's software. Technical Report HCI-TR-05-02, Computer Science Department, University of Saskatchewan.

[50] Chidambaram, V., Chiang, Y.-H., and Mutlu, B. (2012). Designing persuasive robots : how robots might persuade people using vocal and nonverbal cues. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, HRI '12, pages 293–300, New York, NY, USA. ACM.

[51] Christophel, D. M. (1990). The relationships among teacher immediacy behaviors, student motivation, and learning. *Communication education*, 39(4):323–340.

[52] Clay, A., Couture, N., and Nigay, L. (2007). Emotion capture based on body postures and movements. In *International Conference on Computing and e-systems 2007 (TIGERA'07)*.

[53] Cockton, G. (2013). Diffusion of worth mapping: The worth of resource functions.

[54] Cohen, I., Looije, R., and Neerincx, M. (2011). Children's Recognition of Emotions in Robots' Face and Body. *Proceedings of the 6th International Conference on Human Robot Interaction, HRI 2011, Lausanne, Switzerland, March 6-9, 2011.*

[55] Cohen, I., Looije, R., and Neerincx, M. (2014). Child's perception of robot's emotions: effects of platform, context and experience. *International Journal of Social Robotics*, 6(4):507–518.

[56] Collins, N. L. and Miller, L. C. (1994). Self-disclosure and liking: a meta-analytic review. *Psychological bulletin*, 116(3):457.

[57] Colombo, R., Pisano, F., Mazzone, A., Delconte, C., Micera, S., Carrozza, M. C., Dario, P., and Minuco, G. (2007). Design strategies to improve patient motivation during robot-aided rehabilitation. *Journal of NeuroEngineering and Rehabilitation*, 4:3.

[58] Cornish, D. and Dukette, D. (2009). *The Essential 20: Twenty Components of an Excellent Health Care Team.* Rosedog Pr.

[59] Coulson, M. (2004). Attributing emotion to static body postures: Recognition accuracy, confusions, and viewpoint dependence. *Journal of nonverbal behavior*, 28(2):117–139.

[60] Coyle, D., Doherty, G., Matthews, M., and Sharry, J. (2007). Computers in talk-based mental health interventions. *Interacting with Computers*, 19(4):545–562.

[61] Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., and Petticrew, M. (2008). Developing and evaluating complex interventions: the new medical research council guidance. *Bmj*, 337:a1655.

[62] Csikszentmihalyi, M., Abuhamdeh, S., and Nakamura, J. (2005). *Handbook of Competence and Motivation*, chapter 32:Flow, pages 598–698. New York: The Guilford Press.

[63] Danne, T. and Kordonouri, O. (2007). What is so different about diabetes in children. *Diabetes Voice*, 52:16–19.

[64] Dautenhahn, K. (2002). Design spaces and niche spaces of believable social robots. In *Robot and Human Interactive Communication, 2002. Proceedings. 11th IEEE International Workshop on*, pages 192–197. IEEE.

[65] Dautenhahn, K., Ogden, B., and Quick, T. (2002). From embodied to socially embedded agents - implications for interaction-aware robots. *Cognitive Systems Research*, 3(3):397–428.

[66] Davis, H. A. (2003). Conceptualizing the role and influence of student-teacher relationships on children's social and cognitive development. *Educational Psychologist*, 38(4):207–234.

[67] Davis, M. H. (1983). Measuring individual differences in empathy: evidence for a multidimensional approach. *Journal of personality and social psychology*, 44(1):113.

[68] de Gelder, B. (2009). Why bodies? twelve reasons for including bodily expressions in affective neuroscience. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3475–3484.

[69] De Lange, V. (2007). The i-cat as personal assistant for children with diabetes. Master's thesis, Delft University of Technology.

[70] De Meijer, M. (1989). The contribution of general features of body movement to the attribution of emotions. *Journal of Nonverbal Behavior*, 13:247–268.

[71] De Silva, P. R. and Bianchi-Berthouze, N. (2004). Modeling human affective postures: an information theoretic characterization of posture features. *Computer Animation and Virtual Worlds*, 15(3-4):269–276.

[72] Deci, E. L., Eghrari, H., Patrick, B. C., and Leone, D. (1994). Facilitating internalization: The Self-determination theory perspective. *Journal of Personality*, 62(119-142).

[73] Deci, E. L. and Ryan, R. M. (2000). The" what" and" why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological inquiry*, 11(4):227–268.

[74] Deci, E. L. and Ryan, R. M. (2002). *Handbook of self-determination research*. University Rochester Press.

[75] Deckelbaum, R. J. and Williams, C. L. (2001). Childhood obesity: the health issue. *Obesity research*, 9(S11):239S–243S.

[76] Dehn, D. M. and Mulken, S. v. (2000). The impact of animated interface agents: a review of empirical research. *Int. J. Hum.-Comput. Stud.*, 52(1):1–22.

[77] Deshmukh, A., Aylett, R., Kriegel, M., and Vargas, P. A. (2012). Multiple embodiments for robots in heritage applications. In *First international conference robotic innovation for cultural heritage 2012*.

[78] Draper, T. W. and Clayton, W. W. (1992). Using a personal robot to teach young children. *Journal of Genetic Psychology*, 153(3):269.

[79] Driscoll, M. P. (2005). *Psychology of learning for instruction*. Pearson A and B, Boston [u.a.].

[80] Ekman, P. (1965). Differential communication of affect by head and body cues. *Journal of Personality & Social Psychology*, 2(5):726–735.

[81] Ekman, P. (1993). Facial expressions of emotion. *American Psychologist*, 48:384–392.

[82] Ekman, P. and Friesen, W. V. (1975). *Unmasking the face: A guide to recognizing emotions from facial cues*. Englewood Cliffs, NJ: Prentice Hall.

[83] Erden, M. S. (2013). Emotional postures for the humanoid-robot nao. *International Journal of Social Robotics*, 5(4):441–456.

[84] Fasola, J. and Mataric, M. (2011). Comparing physical and virtual embodiment in a socially assistive robot exercise coach for the elderly. *Center for Robotics and Embedded Systems, Los Angeles, CA*.

[85] Fasola, J. and Matarić, M. J. (2010). Robot Motivator: Increasing User Enjoyment and Performance on a Physical/Cognitive Task. *Proc. IEEE International Conference on Development and Learning (ICDL), Ann Arbor, MI, USA*.

[86] FIPA (2008). Fipa agent communication. http://www.fipa.org/repository/aclspecs.html.

[87] Fishbein, M. and Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*.

[88] Fishbein, M. and Ajzen, I. (2010). Prediction and change of behavior: The reasoned action approach.

[89] Fleming, N. D. and Mills, C. (1992). Not another inventory, rather a catalyst for reflection.

[90] Fogg, B. (2002). Persuasive technology: using computers to change what we think and do. *Ubiquity*, 2002(December):5.

[91] Fong, T., Nourbakhsh, I., and Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and autonomous systems*, 42(3-4):143–166.

[92] Gale, E. (2012). Epidemiology of type 1 diabetes.

[93] Garattini, L., Chiaffarino, F., Cornago, D., Coscelli, C., Parazzini, F., del COsti e Risorse del Diabete, S. G. R. R. E., et al. (2004). Direct medical costs unequivocally related to diabetes in italian specialized centers. *The European Journal of Health Economics, formerly: HEPAC*, 5(1):15–21.

[94] Ghai, S. and Sinha, R. (2015). Pitch adaptive mfcc features for improving children's mismatched asr. *International Journal of Speech Technology*, 18(3):489–503.

[95] Gockley, R. and MatariĆ, M. J. (2006). Encouraging physical therapy compliance with a hands-off mobile robot. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pages 150–155. ACM.

[96] Goetz, J., Kiesler, S., and Powers, A. (2003). Matching robot appearance and behavior to tasks to improve human-robot cooperation. In *Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003. The 12th IEEE International Workshop on*, pages 55–60. IEEE.

[97] Goffree, F. and Oonk, W. (2004). *Reken Vaardig: op weg naar basale en professionele gecijferdheid.* Noofdhoff Uitgevers B.V.

[98] Greeff, J. d., Janssen, J., Looije, R., Mioch, T., Alpay, L., Neerincx, M., Baxter, P., and Belpaeme, T. (2013). Activity switching in child-robot interaction: A hospital case study. In *5th International Conference on Social Robotics, ICSR 2013, 27 October 2013 through 29 October 2013, Bristol, 8239 LNAI, 585-586*.

[99] Grey, M., Whittemore, R., and Tamborlane, W. (2002). Depression in type 1 diabetes in children: natural history and correlates. *Journal of psychosomatic research*, 53(4):907–911.

[100] Grimm, L. G. (1993). *Statistical applications for the behavioral sciences*. Wiley.

[101] Grolnick, W. S., Gurland, S. T., Jacob, K. F., and Decourcey, W. (2002). The development of self-determination in middle childhood and adolescence. *Development of achievement motivation*, pages 147–171.

[102] Hess, U. and Blairy, S. (2001). Facial mimicry and emotional contagion to dynamic emotional facial expressions and their influence on decoding accuracy. *International Journal of Psychophysiology*, 40(2):129–141.

[103] Hidi, S. (2000). An interest researcher's perspective: The effects of extrinsic and intrinsic factors on motivation.

[104] Higashinaka, R., Dohsaka, K., and Isozaki, H. (2008). Effects of self-disclosure and empathy in human-computer dialogue. In *Spoken Language Technology Workshop, 2008. SLT 2008. IEEE*, pages 109–112. IEEE.

[105] Hirsh-Pasek, K., Golinkoff, R. M., Berk, L. E., and Singer, D. (2008). *A mandate for playful learning in preschool: Applying the scientific evidence*. Oxford University Press.

[106] Hirth, J., Schmitz, N., and Berns, K. (2011). Towards social robots: Designing an emotion-based architecture. *International Journal of Social Robotics*, 3:273–290.

[107] Hoffman, R. R., Johnson, M., Bradshaw, J. M., and Underbrink, A. (2013). Trust in automation. *Intelligent Systems, IEEE*, 28(1):84–88.

[108] Hollnagel, E. and Woods, D. (1983). Cognitive systems engineering: New wine in new bottles. *International Journal of Man-Machine Studies*, 18(6):583–600.

[109] IDF (2012). 5th idf diabetes atlas. Technical report, International Diabetes Federation (IDF).

[110] Jade (2008). Jade. http://jade.tilab.com/.

[111] Janssen, J. B., Wal, C. C. v. d., Neerincx, M. A., and Looije, R. (2011). Motivating children to learn arithmetic with an adaptive robot game. In *ICSR Conference*, pages 153–162.

[112] Jin, S. A. (2010). The effects of incorporating a virtual agent in a computer-aided test designed for stress management education: The mediating role of enjoyment. *Computers in Human Behavior*, 26(3):443–451.

[113] Johnson, D. O., Cuijpers, R. H., and van der Pol, D. (2013). Imitating human emotions with artificial facial expressions. *International Journal of Social Robotics*, 5(4):503–513.

[114] Jourard, S. M. (1964). *The transparent self: Self-disclosure and well-being*. Number 17. Van Nostrand.

[115] Jung, S., taek Lim, H., Kwak, S., and Biocca, F. (2012). Personality and facial expressions in human-robot interaction. In *Proc. 7tn ACM/IEEE Int. conference on Human-Robot Interaction*.

[116] Kafetsios, K. (2004). Attachment and emotional intelligence abilities across the life course. *Personality and individual Differences*, 37(1):129–145.

[117] Kahn Jr, P. H., Friedman, B., and Hagman, J. (2002). I care about him as a pal: Conceptions of robotic pets in online aibo discussion forums. In *CHI'02 Extended Abstracts on Human Factors in Computing Systems*, pages 632–633. ACM.

[118] Kanda, T., Sato, R., Saiwaki, N., and Ishiguro, H. (2007). A two-month field trial in an elementary school for long-term human-robot interaction. *IEEE Transactions on Robotics*, 23:962–971.

[119] Karoly, P. and Bay, R. C. (1990). Diabetes self-care goals and their relation to children's metabolic control. *Journal of pediatric psychology*, 15(1):83–95.

[120] Kashyap, C. D. and Vishnu, P. R. (2016). Facial emotion recognition. *International Journal of Engineering and Future Technology™*, 7(7):18–29.

[121] Kätsyri, J. and Sams, M. (2008). The effect of dynamics on identifying basic emotions from synthetic and natural faces. *International Journal of Human-Computer Studies*, 66(4):233–242.

[122] Kaya, N. and Epps, H. H. (2004). Relationship between color and emotion: A study of college students. *College Student Journal*, 38(3):396.

[123] Keller, J. M. (1999). Using the arcs motivational process in computer-based instruction and distance education. *New directions for teaching and learning*, 1999(78):37–47.

[124] Keltner, D. and Gross, J. J. (1999). Functional accounts of emotions. *Cognition & Emotion*, 13(5):467–480.

[125] Keltner, D. and Haidt, J. (1999). Social functions of emotions at four levels of analysis. *Cognition & Emotion*, 13(5):505–521.

[126] Kennedy, J., Baxter, P., and Belpaeme, T. (2015). The robot who tried too hard: Social behaviour of a robot tutor can negatively affect child learning. In *Proc. HRI*, volume 15.

[127] Kessens, J. M., Neerincx, M. A., Looije, R., Kroes, M., and Bloothooft, G. (2009). Facial and vocal emotion expression of a personal computer assistant to engage, educate and motivate children. In *3rd IEEE Int. Conference on Affective Computing and. Intelligent Interaction (ACII)*, Amsterdam, the Netherlands.

[128] Kidd, C. D. (2003). Sociable robots: The role of presence and task in human-robot interaction. Master's thesis, Massachusetts Institute of Technology.

[129] Kidd, C. D. and Breazeal, C. (2006). Designing a sociable robot system for weight maintenance. In *IEEE Consumer Communications and Networking Conference. Las Vegas, NV: IEEE*. Citeseer.

[130] Kidd, C. D., Taggart, W., and Turkle, S. (2006). A sociable robot to encourage social interaction among the elderly. In *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, pages 3972–3976. IEEE.

[131] Kim, A., Kum, H., Roh, O., You, S., and Lee, S. (2012). Robot gesture and user acceptance of information in human-robot interaction. In *Proc. 7th ACM/IEEE Int. conference on Human-Robot Interaction*, HRI '12, pages 279–280, New York, NY, USA. ACM.

[132] Kinzler, K. D., Corriveau, K. H., and Harris, P. L. (2011). Children's selective trust in native-accented speakers. *Developmental science*, 14(1):106–111.

[133] Komatsu, T., Abe, Y., Prendinger, H., Lester, J., and Ishizuka, M. (2008). Comparing an On-Screen agent with a robotic agent in Non-Face-to-Face interactions. In *Intelligent Virtual Agents*, volume 5208 of *Lecture Notes in Computer Science*, pages 498–504. Springer Berlin / Heidelberg.

[134] Korfbalverbond, K. N. (2008). Korfbal, speel 't volgens de regels. `http://www.knkv.nl/web/show/id=157092`. consulted on 02-02-2008.

[135] Kovacs, M., Goldston, D., Obrosky, D. S., and Bonar, L. K. (1997). Psychiatric disorders in youths with IDDM: rates and risk factors. *Diabetes care*, 20(1):36–44.

[136] Kriglstein, S. and Wallner, G. (2005). Homie: an artificial companion for elderly people. In *CHI'05 extended abstracts on Human factors in computing systems*, pages 2094–2098. ACM.

[137] Kroes, M. (2007). Emotional speech synthesis: a comparison of different methods. Master's thesis, University of Utrecht.

[138] Kruijff, G.-J. M., Janíček, M., Keshavdas, S., Larochelle, B., Zender, H., Smets, N. J., Mioch, T., Neerincx, M. A., Diggelen, J., Colas, F., et al. (2014). Experience in system design for human-robot teaming in urban search and rescue. In *Field and Service Robotics*, pages 111–125. Springer.

[139] Lask, B. (2003). Motivating children and adolescents to improve adherence. *The Journal of pediatrics*, 143(4):430–433.

[140] Laurenceau, J.-P., Barrett, L. F., and Pietromonaco, P. R. (1998). Intimacy as an interpersonal process: the importance of self-disclosure, partner disclosure, and perceived partner responsiveness in interpersonal exchanges. *Journal of personality and social psychology*, 74(5):1238.

[141] Lee, K. M., Peng, W., Jin, S.-A., and Yan, C. (2006). Can robots manifest personality?: An empirical test of personality recognition, social responses, and social presence in human-robot interaction. *Journal of communication*, 56(4):754–772.

[142] Leite, I., Martinho, C., and Paiva, A. (2013). Social robots for long-term interaction: A survey. *International Journal of Social Robotics*, 5(2):291–308.

[143] Ligthart, M., Henkemans, O. B., Hindriks, K., and Neerincx, M. A. (2017). Expectation management in child-robot interaction. In *Robot and Human Interactive Communication (RO-MAN), 2017 26th IEEE International Symposium on*, pages 916–921. IEEE.

[144] Looije, R. (2006). *iCat for you: a comparison between different personal e-health assistants*. PhD thesis, Master Thesis Rijksuniversiteit Groningen, Groningen.

[145] Looije, R., Cnossen, F., Neerincx, M., et al. (2006). Incorporating guidelines for health assistance into a socially intelligent robot. In *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on*, pages 515–520. IEEE.

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

[147] Looije, R., Neerincx, M. A., and Cnossen, F. (2010). Persuasive robotic assistant for health self-management of older adults: Design and evaluation of social behaviors. *International Journal of Human-Computer Studies*, 68(6):386 – 397.

[148] Looije, R., Neerincx, M. A., and Hindriks, K. V. (2016a). Specifying and testing the design rationale of social robots for behavior change in children. *Cognitive Systems Research*.

[149] Looije, R., Neerincx, M. A., Peters, J. K., and Henkemans, O. A. B. (2016b). Integrating robot support functions into varied activities at returning hospital visits. *International Journal of Social Robotics*, 8(4):483–497.

[150] Looije, R., Van der Zalm, A., Neerincx, M., Beun, R.-J., et al. (2012). Help, i need some body the effect of embodiment on playful learning. In *RO-MAN, 2012 IEEE*, pages 718–724. IEEE.

[151] Lowenthal, P. R. (2009). The evolution andinfluence of social presence theory on online learning. *Social Computing: Concepts, Methodologies, Tools, and Applications: Concepts, Methodologies, Tools, and Applications*, page 113.

[152] Lubsch, L. and Lynch, J. (2006). Caring for children with diabetes. *US Pharm*, 11:21–30.

[153] Markland, D., Ryan, R. M., Tobin, V. J., and Rollnick, S. (2005). Motivational interviewing and self–determination theory. *Journal of social and clinical psychology*, 24(6):811–831.

[154] Markopoulos, P., Read, J., and MacFarlane, S. (2008). *Evaluating children's interactive products: principles and practices for interaction designers*. Morgan Kaufmann Pub.

[155] Marsella, S., Johnson, W. L., and LaBore, C. (2003). Interactive pedagogical drama for health interventions. In *11th International Conference on Artificial Intelligence in Education, Sydney, Australia*, pages 341–348.

[156] Mbrola (2009). Mbrola synthesis. http://tcts.fpms.ac.be/synthesis/.

[157] McHugo, G. J., Lanzetta, J. T., Sullivan, D. G., Masters, R. D., and Englis, B. G. (1985). Emotional reactions to a political leader's expressive displays. *Journal of Personality and Social Psychology*, 49(6):1513–1529.

[158] McNally, K. A., Schefft, B. K., Szaflarski, J. P., Howe, S. R., Yeh, H.-S., and Privitera, M. D. (2009). Application of signal detection theory to verbal memory testing to distinguish patients with psychogenic nonepileptic seizures from patients with epileptic seizures. *Epilepsy & Behavior*, 14(4):597–603.

[159] McNeill, D. (1992). *Hand and Mind*. The University of Chicago Press.

[160] Michie, S., Johnston, M., Abraham, C., Francis, J., and Eccles, M. P. (2013). The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: Building an international consensus for the reporting of behavior change interventions. *Annals of Behavioral Medicine*, pages 1–15.

[161] Miller, W. and Rollnick, S. (2002). *Motivational interviewing: Preparing people for change*. The Guilford Press.

[162] Mitnik, R., Nussbaum, M., and Soto, A. (2008). An autonomous educational mobile robot mediator. *Autonomous Robots*, 25(4):367–382.

[163] Moon, Y. (2000). Intimate exchanges: Using computers to elicit self-disclosure from consumers. *Journal of consumer research*, 26(4):323–339.

[164] Moreno-Ger, P., Burgos, D., Martínez-Ortiz, I., Sierra, J. L., and Fernández-Manjón, B. (2008). Educational game design for online education. *Computers in Human Behavior*, 24(6):2530–2540.

[165] Mower, E., Mataric, M. J., and Narayanan, S. (2009). Human perception of audio-visual synthetic character emotion expression in the presence of ambiguous and conflicting information. *Multimedia, IEEE Transactions on*, 11(5):843–855.

[166] Muris, P., Meesters, C., and Diederen, R. (2005). Psychometric poperties of the big five questionnaire for children (bfq-c) in a dutch sample of young adolescents. *Personality and individual differences*, 38(8):1757 – 1769.

[167] Murray, I. R. and Arnott, J. L. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *The Journal of the Acoustical Society of America*, 93(2):1097–1108.

[168] Naito, H., Takeuchi, Y., Kim, J., Ge, S. S., Vadakkepat, P., Jesse, N., Al Manum, A., Puthusserypady K, S., Rückert, U., Sitte, J., Witkowski, U., Nakatsu, R., Braunl, T., Baltes, J., Anderson, J., Wong, C., Verner, I., and Ahlgren, D. (2009). Promotion of efficient cooperation by sharing environment with an agent having a body in real world. In *Progress in Robotics*, volume 44 of *Communications in Computer and Information Science*, pages 128–133. Springer Berlin Heidelberg.

[169] Narayan, K. V., Boyle, J. P., Thompson, T. J., Sorensen, S. W., and Williamson, D. F. (2003). Lifetime risk for diabetes mellitus in the united states. *Jama*, 290(14):1884–1890.

[170] Neerincx, M. and Streefkerk, J. W. (2003). Interacting in desktop and mobile context: Emotion, trust, and task performance. In *Ambient Intelligence*, pages 119–132. Springer.

[171] Neerincx, M. A. (2011). Situated cognitive engineering for crew support in space. *Personal and Ubiquitous Computing*, 15(5):445–456.

[172] Neerincx, M. A. and Lindenberg, J. (2008). Situated cognitive engineering for complex task environments. *Naturalistic decision making and macrocognition*, pages 373–390.

[173] Neerincx, M. A., van Diggelen, J., and van Breda, L. (2016). Interaction design patterns for adaptive human-agent-robot teamwork in high-risk domains. In *International Conference on Engineering Psychology and Cognitive Ergonomics*, pages 211–220. Springer.

[174] Nejat, G., Sun, Y., and Nies, M. (2009). Assistive robots in health care settings. *Home health care management & practice*, 21(3):177–187.

[175] Niemiec, C. P. and Ryan, R. M. (2009). Autonomy, competence, and relatedness in the classroom applying self-determination theory to educational practice. *Theory and Research in Education*, 7(2):133–144.

[176] Norman, D. (1986). *User-Centred System Design: New perspectives on human-computer interaction*, chapter Cognitive Engineering, pages 31–61. Hillsdale, NJ: Lawrence Erlbaum Associates.

[177] Norman, D. A. and Ortony, A. (2003). Designers and users: Two perspectives on emotion and design. In *Symposium on foundations of interaction design*.

[178] Novak, J. D. and Cañas, A. J. (2008). The theory underlying concept maps and how to construct and use them. *Florida Institute for Human and Machine Cognition Pensacola Fl, www. ihmc. us.[http://cmap. ihmc. us/Publications/ResearchPapers/TheoryCmaps/TheoryUnderlyingConceptMaps. htm]*, 284.

[179] Oinas-Kukkonen, H. (2010). Behavior change support systems: A research model and agenda. In *Persuasive Technology*, pages 4–14. Springer.

[180] Oinas-Kukkonen, H. and Harjumaa, M. (2008). A systematic framework for designing and evaluating persuasive systems. In *International Conference on Persuasive Technology*, pages 164–176. Springer.

[181] Oinas-Kukkonen, H. and Harjumaa, M. (2009). Persuasive systems design: Key issues, process model, and system features. *Communications of the Association for Information Systems*, 24(1):28.

[182] Okita, S. Y. (2013). Self & "Other's Perspective Taking: The Use of Therapeutic Robot Companions as Social Agents for Reducing Pain and Anxiety in Pediatric Patients. *Cyberpsychology, Behavior, and Social Networking*, 16(6):436–441.

[183] Olsen, D. R. and Goodrich, M. A. (2003). Metrics for evaluating human-robot interactions. In *Proceedings of PERMIS*, volume 2003, page 4.

[184] Ortony, A. and Turner, T. J. (1990). What's basic about basic emotions? *Psychological review*, 97(3):315.

[185] Palermo, T. M., Valenzuela, D., and Stork, P. P. (2004). A randomized trial of electronic versus paper pain diaries in children: impact on compliance, accuracy, and acceptability. *Pain*, 107(3):213–219.

[186] Papangelis, A., Zhao, R., and Cassell, J. (2014). Towards a computational architecture of dyadic rapport management for virtual agents. In *Intelligent Virtual Agents*, pages 320–324. Springer.

[187] Patterson, C. C., Dahlquist, G. G., Gyürüs, E., Green, A., Soltész, G., Group, E. S., et al. (2009). Incidence trends for childhood type 1 diabetes in europe during 1989–2003 and predicted new cases 2005–20: a multicentre prospective registration study. *The Lancet*, 373(9680):2027–2033.

[188] Peeters, M. (2014). *Personalized Educational Games-Developing agent-supported scenario-based training*. PhD thesis, SIKS, the Dutch Graduate School for Information and Knowledge Systems.

[189] Pereira Santos, C., Khan, V.-J., and Markopoulos, P. (2018). Profiling ethics orientation through play. *Behaviour & Information Technology*, 37(9):926–935.

[190] Peyrot, M. (2008). How is diabetes perceived? the results of the dawn youth survey. *Diabetes voice*, 53:9–13.

[191] Philips (2009). Philips robotics. http://www.research.philips.com/technologies/robotics.html.

[192] Picard, R. W. (1997). *Affective computing*, volume 252. MIT press Cambridge.

[193] Plant, R. and Ryan, R. M. (1985). Intrinsic motivation and the effect of self-consciousness, self-awareness, and ego-involvement: An investigation of internally-controlling styles. *Journal of Personality*, 53(435-449).

[194] Powers, A., Kiesler, S., Fussell, S., and Torrey, C. (2007). Comparing a computer agent with a humanoid robot. In *Human-Robot Interaction (HRI), 2007 2nd ACM/IEEE International Conference on*, pages 145–152. IEEE.

[195] Preece, J., Sharp, H., and Rogers, Y. (2015). *Interaction Design-beyond human-computer interaction.* John Wiley & Sons.

[196] Prensky, M. (2001). Fun, play and games: What makes games engaging. *Digital game-based learning*, pages 11–16.

[197] Rasmussen, J. (1986). *Information processing and human-machine interaction: An approach to cognitive engineering.* Elsevier Science Inc. New York, NY, USA.

[198] Read, J., MacFarlane, S., and Casey, C. (2002). Endurability, engagement and expectations: Measuring children's fun. In *Interaction Design and Children*, pages 189–198. Citeseer.

[199] Read, J. C. and MacFarlane, S. (2006). Using the fun toolkit and other survey methods to gather opinions in child computer interaction. In *Proceedings of the 2006 conference on Interaction design and children*, pages 81–88. ACM.

[200] Reeve, J. and Deci, E. L. (1996). Elements of the Competitive Situation That Affect Intrinsic Motivation. *Society for Personality and Social Psychology*, 22(1):24–33.

[201] Reeves, B. and Nass, C. (1996). *The Media Equation: How people treat computers, television, and new media like real people and places.* CSLI Publications and Cambridge university press.

[202] Reis, H. T., Shaver, P., and others (1988). Intimacy as an interpersonal process. *Handbook of personal relationships*, 24(3):367–389.

[203] Robben, S. (2011). It's NAO or Never! Facilitate Bonding Between a Child and a Social Robot: Exploring the Possibility of a Robot Adaptive to Personality. Master's thesis, University of Nijmegen.

[204] Robins, B., Dautenhahn, K., and Dickerson, P. (2009). From isolation to communication: a case study evaluation of robot assisted play for children with autism with a minimally expressive humanoid robot. In *Advances in Computer-Human Interactions, 2009. ACHI'09. Second International Conferences on*, pages 205–211. IEEE.

[205] Rogers, C. R. (1951). *Client-centered therapy: Its current practice, implications and theory.* Houghton Mifflin Boston.

[206] Rollnick, S., Mason, P., and Butler, C. (1999). *Health behavior change: a guide for practitioners.* Elsevier Health Sciences.

[207] Rollnick, S. and Miller, W. (1995). What is motivational interviewing? *Behavioural and Cognitive Psychotherapy*, 23(04):325–334.

[208] Ros, R. and Demiris, Y. (2013). Creative dance: An approach for social interaction between robots and children. In *Human Behavior Understanding*, pages 40–51. Springer.

[209] Rosenthal, M. S. (2009). *The Canadian type 2 diabetes sourcebook*. John Wiley & Sons.

[210] Russell, J. (1980). A circumplex model of affect. *Journal of Personality & Social Psychology*, 39(6):1161–1178.

[211] Russell, S. and Norvig, P. (2003). *Artificial Intelligence, a modern approach*. Pearson Education, New Jersey.

[212] Ruttkay, Z., Zwiers, J., van Welbergen, H., and Reidsma, D. (2006). Towards a reactive virtual trainer. In *Intelligent virtual agents*, pages 292–303. Springer.

[213] Ryan, R. M. and Deci, E. L. (2000a). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology*, 25(1):54–67.

[214] Ryan, R. M. and Deci, E. L. (2000b). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist*, 55(1):68.

[215] Ryan, R. M., Rigby, C. S., and Przybylski, A. (2006). The motivational pull of video games: A self-determination theory approach. *Motivation and emotion*, 30(4):344–360.

[216] Sabaté, E. (2003). *Adherence to long-term therapies: evidence for action*. World Health Organization.

[217] Salem, M., Kopp, S., Wachsmuth, I., Rohlfing, K., and Joublin, F. (2012). Generation and evaluation of communicative robot gesture. *International Journal of Social Robotics*, 4:201–217.

[218] Schadenberg, B. (2012). Modelling the user's skill and performance with the use of a bayesian rating system. Master's thesis, Rijksuniversiteit Groningen.

[219] Schröder, M. (1999). Can emotions be synthesized without controlling voice quality. *Phonus*, 4:37–55.

[220] Schröder, M. (2004). *Speech and Emotion Research: An overview of research frameworks and a dimensional approach to emotional speech synthesis*. Institut für Photetik, Universität des Saarlandes.

[221] Schulte, J., Rosenberg, C., and Thrun, S. (1999). Spontaneious, short-term interaction with mobile robots. In *Proc. of the Int. Conference on Robotics and Automation*.

[222] Scopelliti, M., Giuliani, M. V., and Fornara, F. (2005). Robots in a domestic setting: a psychological approach. *Universal Access in the Information Society*, 4(2):146–155.

[223] Sebire, S. J., Jago, R., Fox, K. R., Edwards, M. J., and Thompson, J. L. (2013). Testing a self-determination theory model of children's physical activity motivation: a cross-sectional study. *Int J Behav Nutr Phys Act*, 10:111.

[224] Shaarani, A. S. and Romano, D. M. (2008). The intensity of perceived emotions in 3d virtual humans. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 3*, pages 1261–1264. International Foundation for Autonomous Agents and Multiagent Systems.

[225] Shimomura, I., Matsuda, M., Hammer, R. E., and others (2001). Are eating disorders more prevalent in females with type 1 diabetes mellitus when the impact of insulin omission is considered? *Nutrition reviews*, 59(6):179.

[226] Shinozawa, K., Naya, F., Yamato, J., and Kogure, K. (2005). Differences in effect of robot and screen agent recommendations on human decision-making. *Int. J. Hum.-Comput. Stud.*, 62(2):267–279.

[227] Short, E., Swift-Spong, K., Greczek, J., Ramachandran, A., Litoiu, A., Grigore, E. C., Feil-Seifer, D., Shuster, S., Lee, J. J., Huang, S., et al. (2014). How to train your dragonbot: Socially assistive robots for teaching children about nutrition through play. In *Robot and Human Interactive Communication, 2014 RO-MAN: The 23rd IEEE International Symposium on*, pages 924–929. IEEE.

[228] Siino, R. M., Chung, J., and Hinds, P. J. (2008). Colleague vs. tool: Effects of disclosure in human-robot collaboration. In *Robot and Human Interactive Communication, 2008. RO-MAN 2008. The 17th IEEE International Symposium on*, pages 558–562. IEEE.

[229] Smith, C. and Scott, H. (1997). *A componential approach to the meaning of facial expressions*, volume 229, chapter 10, pages 229–254. Cambridge University Press.

[230] Solms, L. (2014). An exploration of the effects of touch on social bonding between robot and child (unpublished).

[231] Spiekman, M. E., Haazebroek, P., and Neerincx, M. A. (2011). Requirements and platforms for social agents that alarm and support elderly living alone. In *Social Robotics*, pages 226–235. Springer.

[232] Stanislaw, H. and Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior research methods, instruments, & computers*, 31(1):137–149.

[233] Steele, H., Steele, M., and Croft, C. (2008). Early attachment predicts emotion recognition at 6 and 11 years old. *Attachment & human development*, 10(4):379–393.

[234] Steinfeld, A., Fong, T., Kaber, D., Lewis, M., Scholtz, J., Schultz, A., and Goodrich, M. (2006). Common metrics for human-robot interaction. In *Proceedings of the HRI 2006*, pages 33–40. ACM.

[235] Tanaka, F., Cicourel, A., and Movellan, J. R. (2007). Socialization between toddlers and robots at an early childhood education center. *Proceedings of the National Academy of Sciences*, 104(46):17954–17958.

[236] Tanaka, F. and Kimura, T. (2010). Care-receiving robot as a tool of teachers in child education. *Interaction Studies*, 11(2):263–268.

[237] Tapus, A. and Mataric, M. J. (2006). Towards socially assistive robotics. *International Journal Robot Soc Jpn*, 24(5):576–578.

[238] Teixeira, P. J., Palmeira, A. L., and Vansteenkiste, M. (2012). The role of self-determination theory and motivational interviewing in behavioral nutrition, physical activity, and health: an introduction to the ijbnpa special series. *International Journal of Behavioral nutrition and Physical Activity*, 9(1):1.

[239] Tejada, S., Traft, N., Hutson, M., Bufford, H., Dooner, M., Hanson, J., Radler, A., and Mauer, G. (2006). Educational robots: Three models for the research of learning theories and Human-Robot interaction. pages 70–76.

[240] Tetteroo, D. and Markopoulos, P. (2017). Eud survival "in the wild": evaluation challenges for field deployments and how to address them. In *New perspectives in end-user development*, pages 207–229. Springer.

[241] Tielman, M. (2013). Expressive behaviour in robot-child interaction. Master's thesis, Utrecht University.

[242] Tielman, M., Neerincx, M., Meyer, J.-J., and Looije, R. (2014). Adaptive emotional expression in robot-child interaction. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pages 407–414. ACM.

[243] Toscos, T., Faber, A., An, S., and Gandhi, M. P. (2006). Chick clique: persuasive technology to motivate teenage girls to exercise. In *CHI'06 extended abstracts on Human factors in computing systems*, pages 1873–1878. ACM.

[244] Triplett, N. (1898). The dynamogenic factors in pacemaking and competition. *The American journal of psychology*, 9(4):507–533.

[245] Truong, K. P., van Leeuwen, D. A., and Neerincx, M. A. (2007). Unobtrusive multimodal emotion detection in adaptive interfaces: speech and facial expressions. In *Foundations of Augmented Cognition*, pages 354–363. Springer.

[246] Tsiouli, E., Alexopoulos, E. C., Stefanaki, C., Darviri, C., and Chrousos, G. P. (2013). Effects of diabetes-related family stress on glycemic control in young patients with type 1 diabetes systematic review. *Canadian Family Physician*, 59(2):143–149.

[247] Ultes, S., Dikme, H., and Minker, W. (2016). Dialogue management for user-centered adaptive dialogue. In *Situated Dialog in Speech-Based Human-Computer Interaction*, pages 51–61. Springer.

[248] Vallerand, R., Gauvin, L., and Halliwell, W. (1986). Negative Effects of Competition on Children's Intrinsic Motivation. *The Journal of Social Psychology*, 126(5):649–657.

[249] Van Dam, I. (2013). Meet my new robot best friend: an exploration of the effects of personality traits in a robot on enhancing friendship. Master's thesis, Universiteit Utrecht.

[250] Van Der Drift, E. J., Beun, R.-J., Looije, R., Blanson Henkemans, O. A., and Neerincx, M. A. (2014). A remote social robot to motivate and support diabetic children in keeping a diary. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pages 463–470. ACM.

[251] Van Dijken, G., Niesink, I., and Schrijvers, A. (2005). Telehealth in de verenigde staten. *UMC, Utrecht.*

[252] Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly*, pages 425–478.

[253] Verhaegh, J. (2004). icat, an intelligent user interface robot with character. an evaluation of robot personality and personality preference in child-robot interaction. *Unpublished master thesis, University of Maastricht.*

[254] Vosoiadou, S. (2003). How children learn? *Successful Schooling*, page 16.

[255] Vygotsky, L. (1978). Interaction between learning and development. *Readings on the development of children*, 23(3):34–41.

[256] Vygotsky, L. S. (1980). *Mind in society: The development of higher psychological processes.* Harvard university press.

[257] Wainer, J., Feil-Seifer, D. J., Shell, D., Matarić, M. J., et al. (2006). The role of physical embodiment in human-robot interaction. In *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on*, pages 117–122. IEEE.

[258] Wainer, J., Feil-Seifer, D. J., Shell, D., Mataric, M. J., and others (2007). Embodiment and human-robot interaction: A task-based perspective. In *Robot and Human interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium on*, pages 872–877. IEEE.

[259] Wechsler, D., Laicardi, C., and Orsini, A. (1997). *WAIS-R: Wechsler adult intelligence scale revised: manuale.* OS.

[260] Weiss, A., Bernhaupt, R., Lankes, M., and Tscheligi, M. (2009a). The usus evaluation framework for human-robot interaction. In *AISB2009: Proceedings of the Symposium on New Frontiers in Human-Robot Interaction.*

[261] Weiss, A., Wurhofer, D., and Tscheligi, M. (2009b). 'i love this dog' children's emotional attachment to the robotic dog aibo. *International Journal of Social Robotics*, 1(3):243–248.

[262] Westera, M., Boschloo, J., van Diggelen, J., Koelewijn, L. S., Neerincx, M. A., and Smets, N. J. (2010). Employing use-cases for piecewise evaluation of requirements and claims. In *Proceedings of the 28th Annual European Conference on Cognitive Ergonomics*, pages 279–286. ACM.

[263] Whittemore, R., Kanner, S., and Grey, M. (2004). *Evidence-based practice in nursing and healthcare: a guide to best practice*, chapter The influence of family on psychological and psychosocial health in youth with type 1 diabetes: a systematic review. Williams & Wilkins, Philadelphia.

[264] Wigdor, N. (2014). Conversational fillers for response delay amelioration in child-robot interaction. Master's thesis, University of Utrecht.

[265] Witte, K. (1994). Fear control and danger control: A test of the extended parallel process model (eppm). *Communications Monographs*, 61(2):113–134.

[266] Woods, S., Dautenhahn, K., and Schulz, J. (2004). The design space of robots: Investigating children's views. In *ROMAN 2004*, pages 47–52. IEEE.

[267] Woolf, B. (2007). *Building intelligent interactive tutors*, volume 3. Elsevier.

[268] Xu, J., Broekens, J., Hindriks, K., and Neerincx, M. (2013). Mood expression through parameterized functional behavior of robots. In *22nd IEEE RO-MAN*.

[269] Yanco, H. and Drury, J. (2005). Classifying human-robot interaction: an updated taxonomy. *International Conference on Systems, Man and Cybernetics*, 3:2841–2846.

[270] Zhao, R., Papangelis, A., and Cassell, J. (2014). Towards a dyadic computational model of rapport management for human-virtual agent interaction. In *Intelligent Virtual Agents*, pages 514–527. Springer.