Specifying and testing the design rationale of social robots for behavior change in children

Rosemarijn Looije[a,*], Mark A. Neerincx[a,b], Koen V. Hindriks[b]

[a] *TNO, Postbus 23, 3769ZG, Soesterberg, The Netherlands*
[b] *Interactive Intelligence Group, Delft University of Technology, Mekelweg 4, 2628 CD, Delft, The Netherlands*

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

*Please address correspondence to Rosemarijn Looije

Email addresses: rosemarijn.looije@tno.nl (Rosemarijn Looije), mark.neerincx@tno.nl (Mark A. Neerincx), k.v.hindriks@tudelft.nl (Koen V. Hindriks)

Preprint submitted to Cognitive Systems Research http://dx.doi.org/10.1016/j.cogsys.2016.07.002

supports replicating experiments or reusing parts of the design rationale.

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

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., [1]). 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 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. [2]) and evaluation (e.g. [3]) 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 [4] 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 [5]. 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?" [6] . 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 [7] can partially answer this question. sCE has been used in different domains, amongst which to systematically design and evaluate robot systems [8]. 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 sub-questions 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[1]). The structure of this paper is as follows: First in section 2, we provide background on diabetes, social robotics, behavior change and situated Cognitive Engineering. Second in section 3, we describe the sDR template, that describes the concepts and it relations, followed by the instantiation of sDR in section 4. In section 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 6.

2. Background

Type 1 diabetes has an enormous impact on the daily life of children with this illness as we will discuss in section 2.1. 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). The behavior of the robot should

[1] `www.aliz-e.org`

be based on knowledge from behavior change theories and systems (section 2.3), and the design of the robot should be based on a state-of-the-art design methodology (section 2.4). Based on this background we can conclude what is lacking to come to a concise and precise situated Design Rationale.

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 [9]. 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 [10]. 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 [11]. The life expectancy of children diagnosed by age 10 is 19 years shorter than that of healthy children [12]. There are also effects on psychological well-being, feelings of embarrassment and on school performance [13]. 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 [14]. Other research suggests that high family stress negatively affects glycemic control [15]. 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 [16, 17] where they are applied for autistic children, [18, 19, 20] for education and [21] 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. 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; [22, 23, 24]), 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 [25]) and their behavior. Bartneck et al. [26] 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). Fong et al. [1] 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. Dautenhahn [27] 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). Spiekman et al. [28] 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. Weiss et al. [3] 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 [29]), physiological measures, focus groups and interviews. Bartneck et al. [30] 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.

2.3. 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 [31]. 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 [4]. 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. [4] 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 Oinas-Kukkonen [6] 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 [6].

2.4. Situated cognitive engineering

The situated Cognitive Engineering (sCE) [32] 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 [33] and robots [8]. In sCE functions are incrementally developed. It can be viewed as a refinement of classical cognitive engineering methods [34, 35, 36], 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 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.

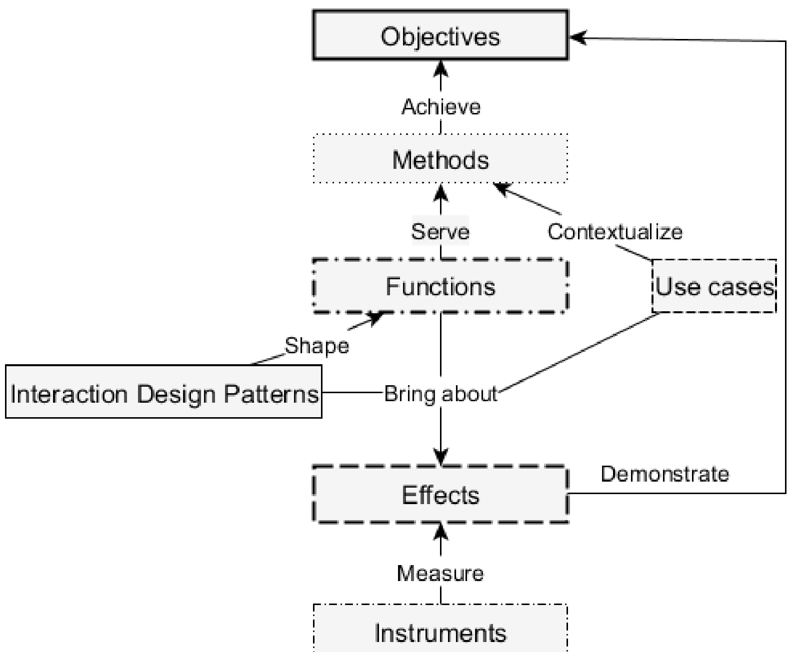

Figure 1: Generic concept map of the situated Design Rationale (sDR).

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.

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 1 the seven generic concepts and their relations are shown. In the following paragraph we explain how the generic sDR is developed.

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 [37] 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[2].

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 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*.

[2]Using yEd `https://www.yworks.com/en/products/yfiles/yed/` we created a general concept map of sDR 1 in which the concepts and their relations are visualized

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 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 [38] 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.

4. Instantation 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 6 we added the outgoing arrow form to the text of the functions.

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.

4.1.1. 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 [39], the Extended Parallel Process Model [40] and the Self-Determination theory [41].

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 dont 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 [42]. 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 [43]. 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 [44]).

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 [4]. As a decision had to be made we chose Self-Determination Theory as our starting point (see objectives in Figure 2), because this theory

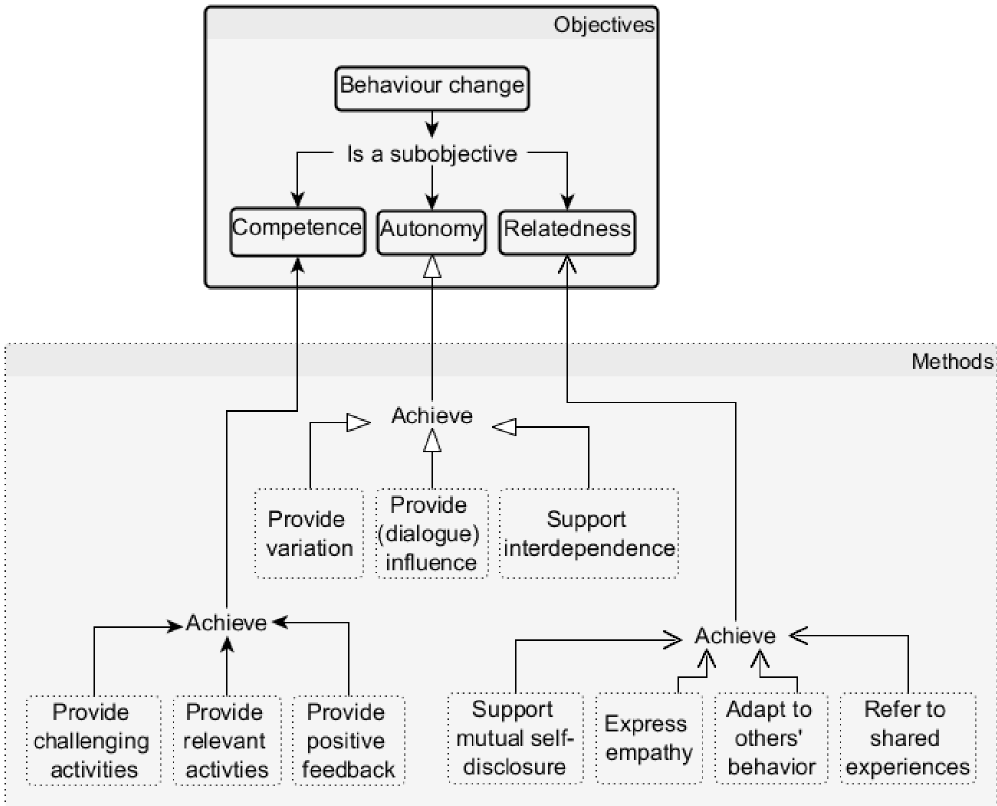

Figure 2: Objectives and methods that achieve them

is used not only in behavioral change but also in education [45], for children in the relevant age group (7-11) [46] and in games where it showed to be a predictor of enjoyment and future game play [47].

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 [48]. 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 [49] and techniques from MI have been implemented in a social robot for adults with diabetes [50].

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 1: Overview of the ALIZ-e use cases.

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

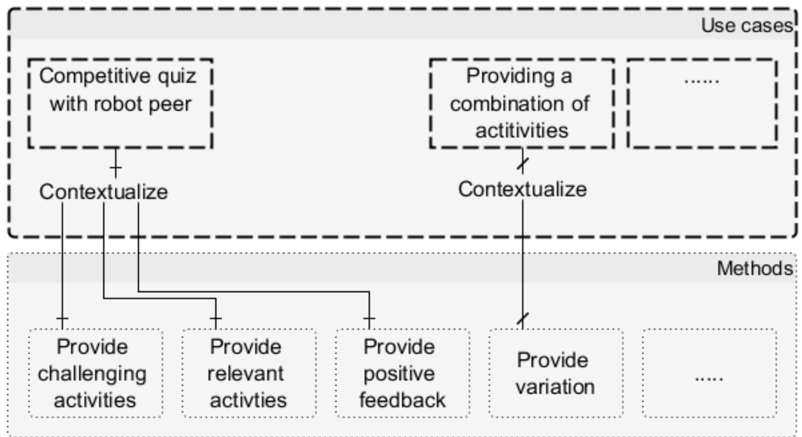

Figure 3: Use cases and how they contextualize methods.

methods are overlapping; for instance Vygotskys educational theory [51] and gaming theory [52] both endorse the importance of having challenging, relevant activities to support intrinsic (long-term) motivation. Literature supports the relation between challenging activities and self-determination theory [53]. 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 others 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 Zhao et al. [54] an overview of methods to reach rapport is provided.

Figure 2 shows the methods used within the ALIZ-e project and their relations to the objectives. All methods come from literature; MI [48], educational [51, 55], gaming [52] and relation theories [54]. In some of this literature the methods are explicitly linked to SDT objectives (e.g. [48, 53]), 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.

4.3. Use cases

The objectives, methods and (later on) related effects and measures wont 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 1) describing the interaction in more detail. For more information on how these use cases were incorporated in experiments we refer to [56], in which an experiment is described containing most of the use cases.

Each use case contextualizes the methods and provides situational context of the 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 3).

4.4. Functions

Based on the methods and use cases a selection of functions was implemented during the project. In Table 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 dont 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 4 this is shown, the functions related to the methods of Figure 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).

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: The different functions and an exemplification

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) [57] and at the effect of embodiment (virtual or physical) on the effectiveness of social behavior [58]. Figure 5 shows how the different aspects of the voice and body influence the emotion expression and thereby the social behavior.

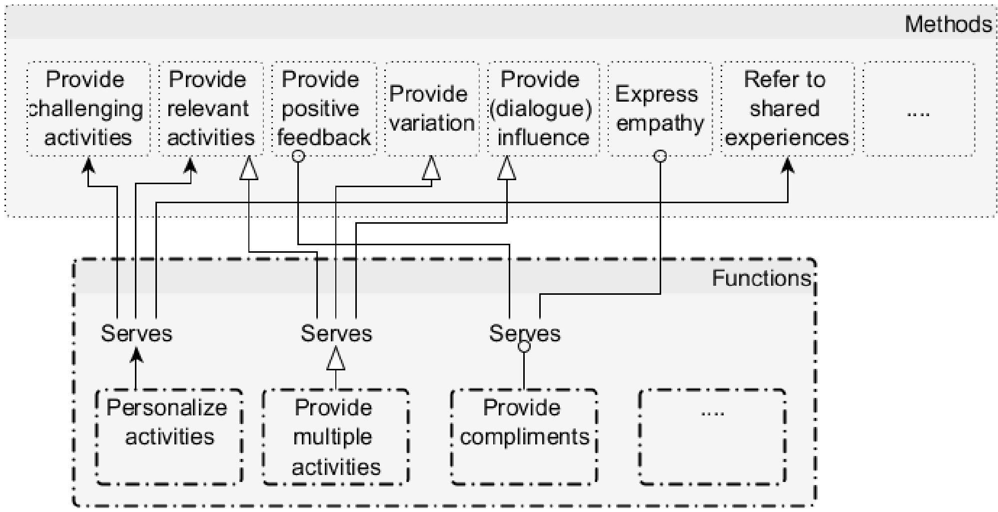

Figure 4: Methods and functions that serve them

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 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.

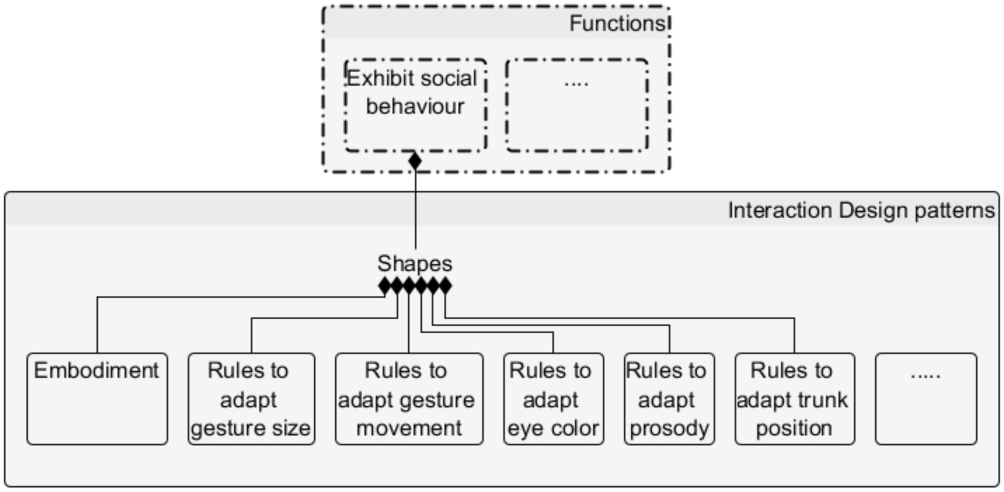

Figure 5: Interaction design patterns shape functions

In Figure 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.

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 7 it can be seen that although we would like to have ob-

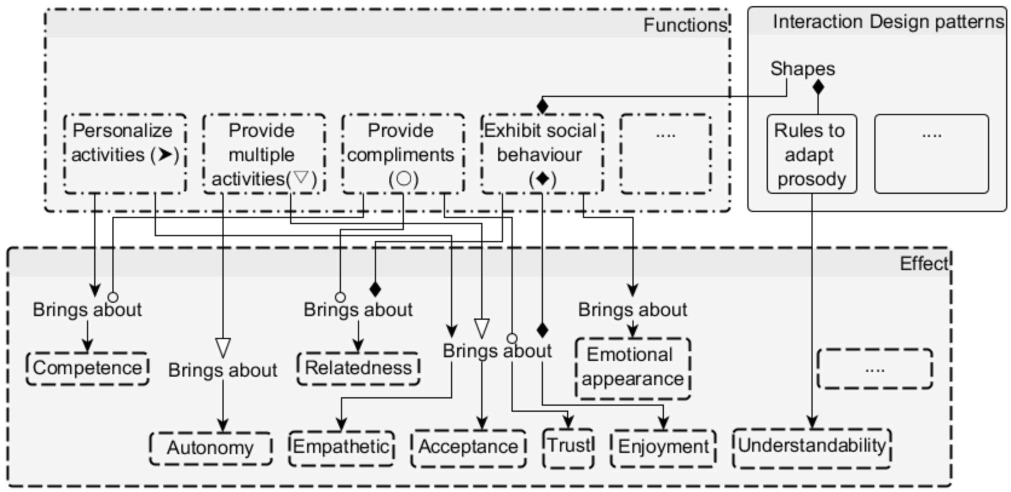

Figure 6: Functions and interaction design patterns bring about effects

jective 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.

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 robots 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 [59]. The objective this experiment focused on *relatedness* and

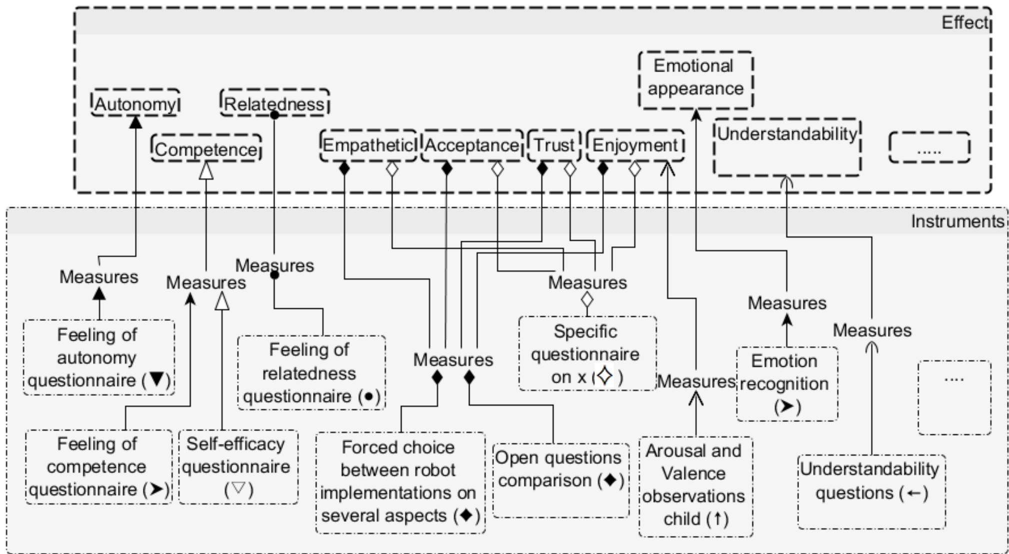

Figure 7: Effects are measured with specific instruments

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 8 shows the sDR of this specific experiment, we limited the number of relations in comparison to Figure 1 by excluding the relation between effects and objectives and use cases and functions, both can be derived by following the other relations.

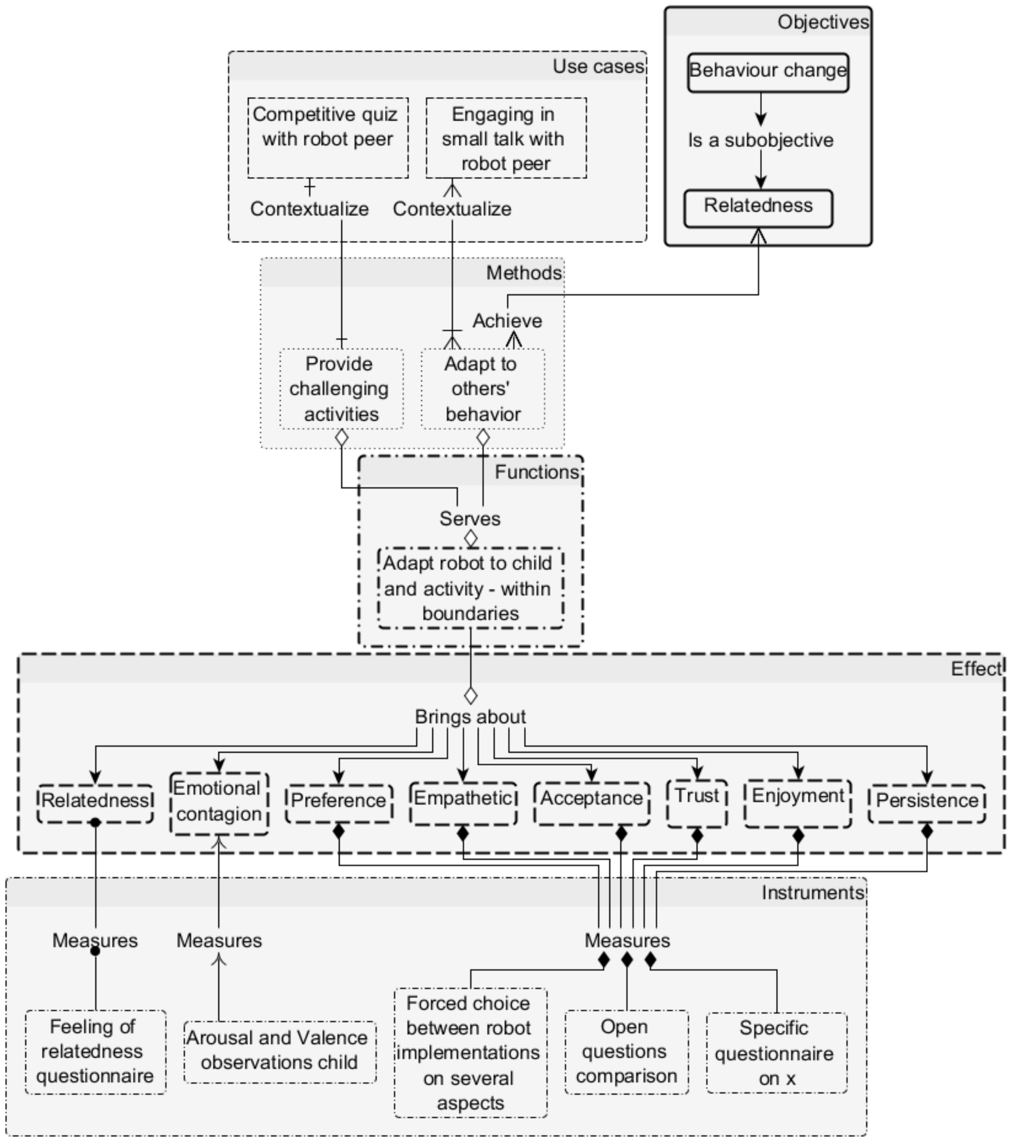

Figure 8: situated Design Rationale of emotional contagion experiment. The use cases "bring about" the effects, but for readability reasons we excluded this line from the overview as we did with the "demonstrate" lines from effects to objectives. [59]

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- affective 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.

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 [57]. 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 [60]; in addition, literature on trust of children in caretakers with an unfamiliar accent [61] indicates that understandability influences trust. We have to add understandability thus as a possible downside for prosody which can be measured asking directly about understandability and in concurrence look at effects on trust and acceptance. Figure 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

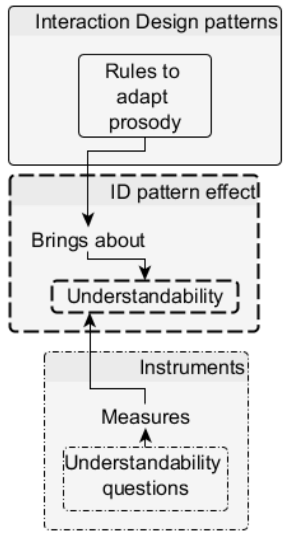

Figure 9: Refinement of sDR, based on emotional contagion experiment.

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.

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.

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://goo.gl/OHgUC8`). 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 projects progress. This relatively stability of the sDR supports adapting and extending.

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 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.

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`.

References

[1] T. Fong, I. Nourbakhsh, K. Dautenhahn, A survey of socially interactive robots, Robotics and autonomous systems 42 (2003) 143–166.

[2] K. Dautenhahn, B. Ogden, T. Quick, From embodied to socially embedded agents - implications for interaction-aware robots, Cognitive Systems Research 3 (2002) 397–428.

[3] A. Weiss, R. Bernhaupt, M. Lankes, M. Tscheligi, The USUS evaluation framework for human-robot interaction, in: AISB2009: Proceedings of the Symposium on New Frontiers in Human-Robot Interaction.

[4] C. Susan Michie DPhil, C. Marie Johnston PhD, C. Charles Abraham DPhil, C. Jill Francis PhD, M. P. Eccles, 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 (2013) 1–15.

[5] H. Oinas-Kukkonen, M. Harjumaa, Persuasive systems design: Key issues, process model, and system features, Communications of the Association for Information Systems 24 (2009) 28.

[6] H. Oinas-Kukkonen, Behavior change support systems: A research model and agenda, in: Persuasive Technology, Springer, 2010, pp. 4–14.

[7] M. A. Neerincx, Situated cognitive engineering for crew support in space, Personal and Ubiquitous Computing 15 (2011) 445–456.

[8] G.-J. M. Kruijff, M. Janíček, S. Keshavdas, B. Larochelle, H. Zender, N. J. Smets, T. Mioch, M. A. Neerincx, J. Diggelen, F. Colas, et al., Experience in system design for human-robot teaming in urban search and rescue, in: Field and Service Robotics, Springer, pp. 111–125.

[9] IDF, 5th IDF diabetes atlas, Technical Report, International Diabetes Federation (IDF), 2012.

[10] C. C. Patterson, G. G. Dahlquist, E. Gyürüs, A. Green, G. Soltész, E. S. Group, et al., 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 (2009) 2027–2033.

[11] T. Danne, O. Kordonouri, What is so different about diabetes in children, Diabetes Voice 52 (2007) 16–19.

[12] K. V. Narayan, J. P. Boyle, T. J. Thompson, S. W. Sorensen, D. F. Williamson, Lifetime risk for diabetes mellitus in the united states, Jama 290 (2003) 1884–1890.

[13] M. Peyrot, How is diabetes perceived? the results of the dawn youth survey, Diabetes voice 53 (2008) 9–13.

[14] K. K. Boman, J. Viksten, P. Kogner, U. Samuelsson, Serious illness in childhood: the different threats of cancer and diabetes from a parent perspective, The Journal of pediatrics 145 (2004) 373–379.

[15] E. Tsiouli, E. C. Alexopoulos, C. Stefanaki, C. Darviri, G. P. Chrousos, Effects of diabetes-related family stress on glycemic control in young patients with type 1 diabetes systematic review, Canadian Family Physician 59 (2013) 143–149.

[16] B. Robins, K. Dautenhahn, P. Dickerson, 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, IEEE, pp. 205–211.

[17] E. I. Barakova, T. Lourens, Expressing and interpreting emotional movements in social games with robots, Personal and ubiquitous computing 14 (2010) 457–467.

[18] F. Tanaka, T. Kimura, Care-receiving robot as a tool of teachers in child education, Interaction Studies 11 (2010) 263–268.

[19] S. Tejada, N. Traft, M. Hutson, H. Bufford, M. Dooner, J. Hanson, A. Radler, G. Mauer, Educational robots: Three models for the research of learning theories and Human-Robot interaction, pp. 70–76.

[20] J. Kennedy, P. Baxter, T. Belpaeme, The robot who tried too hard: Social behaviour of a robot tutor can negatively affect child learning, in: Proc. HRI, volume 15.

[21] E. Short, K. Swift-Spong, J. Greczek, A. Ramachandran, A. Litoiu, E. C. Grigore, D. Feil-Seifer, S. Shuster, J. J. Lee, S. Huang, et al., 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, IEEE, pp. 924–929.

[22] H. Yanco, J. Drury, Classifying human-robot interaction: an updated taxonomy, International Conference on Systems, Man and Cybernetics 3 (2005) 2841–2846.

[23] D. R. Olsen, M. A. Goodrich, Metrics for evaluating human-robot interactions, in: Proceedings of PERMIS, volume 2003, p. 4.

[24] A. Steinfeld, T. Fong, D. Kaber, M. Lewis, J. Scholtz, A. Schultz, M. Goodrich, Common metrics for human-robot interaction, in: Proceedings of the HRI 2006, ACM, pp. 33–40.

[25] S. Woods, K. Dautenhahn, J. Schulz, The design space of robots: Investigating children's views, in: ROMAN 2004, IEEE, pp. 47–52.

[26] C. Bartneck, J. Forlizzi, A design-centred framework for social human-robot interaction, in: ROMAN 2004, IEEE, pp. 591–594.

[27] K. Dautenhahn, Design spaces and niche spaces of believable social robots, in: Robot and Human Interactive Communication, 2002. Proceedings. 11th IEEE International Workshop on, IEEE, pp. 192–197.

[28] M. E. Spiekman, P. Haazebroek, M. A. Neerincx, Requirements and platforms for social agents that alarm and support elderly living alone, in: Social Robotics, Springer, 2011, pp. 226–235.

[29] V. Venkatesh, M. G. Morris, G. B. Davis, F. D. Davis, User acceptance of information technology: Toward a unified view, MIS quarterly (2003) 425–478.

[30] C. Bartneck, D. Kulić, E. Croft, S. Zoghbi, Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots, International Journal of Social Robotics 1 (2009) 71–81.

[31] P. Craig, P. Dieppe, S. Macintyre, S. Michie, I. Nazareth, M. Petticrew, Developing and evaluating complex interventions: the new medical research council guidance, Bmj 337 (2008) a1655.

[32] M. Neerincx, J. Lindenberg, Situated cognitive engineering for complex task environments, Aldershot, UK: Ashgate Publishing Limited, p. coming soon. Editor: Schraagen, J.M., Militello, L., Ormerod, T., & Lipshitz, R.

[33] O. A. B. Henkemans, P. van Empelen, G. L. Paradies, R. Looije, M. A. Neerincx, 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, IEEE, pp. 49–56.

[34] E. Hollnagel, D. Woods, Cognitive systems engineering: New wine in new bottles, International Journal of Man-Machine Studies 18 (1983) 583–600.

[35] D. Norman, User-Centred System Design: New perspectives on human-computer interaction, Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 31–61.

[36] J. Rasmussen, Information processing and human-machine interaction: An approach to cognitive engineering, Elsevier Science Inc. New York, NY, USA, 1986.

[37] J. D. Novak, A. J. Caas, 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/T heoryCmaps/TheoryUnderlyingConceptMaps. htm] 284 (2008).

[38] G. Cockton, Diffusion of worth mapping: The worth of resource functions (2013).

[39] M. Fishbein, I. Ajzen, Prediction and change of behavior: The reasoned action approach, 2010.

[40] K. Witte, Fear control and danger control: A test of the extended parallel process model (eppm), Communications Monographs 61 (1994) 113–134.

[41] E. L. Deci, R. M. Ryan, Handbook of self-determination research, University Rochester Press, 2002.

[42] R. M. Ryan, E. L. Deci, Intrinsic and extrinsic motivations: Classic definitions and new directions, Contemporary educational psychology 25 (2000) 54–67.

[43] E. L. Deci, R. M. Ryan, The" what" and" why" of goal pursuits: Human needs and the self-determination of behavior, Psychological inquiry 11 (2000) 227–268.

[44] W. Miller, S. Rollnick, Motivational interviewing: Preparing people for change, The Guilford Press, 2002.

[45] C. P. Niemiec, R. M. Ryan, Autonomy, competence, and relatedness in the classroom applying self-determination theory to educational practice, Theory and Research in Education 7 (2009) 133–144.

[46] S. J. Sebire, R. Jago, K. R. Fox, M. J. Edwards, J. L. Thompson, Testing a self-determination theory model of children's physical activity motivation: a cross-sectional study, Int J Behav Nutr Phys Act 10 (2013) 111.

[47] R. M. Ryan, C. S. Rigby, A. Przybylski, The motivational pull of video games: A self-determination theory approach, Motivation and emotion 30 (2006) 344–360.

[48] D. Markland, R. M. Ryan, V. J. Tobin, S. Rollnick, Motivational interviewing and self–determination theory, Journal of social and clinical psychology 24 (2005) 811–831.

[49] J. H. Bigelow, S. Cretin, M. Solomon, S.-Y. Wu, M. O'Connell, Patient compliance with and attitudes towards Health Buddy, Technical Report, DTIC Document, 2000.

[50] R. Looije, M. A. Neerincx, F. Cnossen, Persuasive robotic assistant for health self-management of older adults: Design and evaluation of social behaviors, International Journal of Human-Computer Studies 68 (2010) 386 – 397.

[51] L. S. Vygotsky, Mind in society: The development of higher psychological processes, Harvard university press, 1980.

[52] M. Csikszentmihalyi, S. Abuhamdeh, J. Nakamura, Handbook of Competence and Motivation, New York: The Guilford Press, pp. 598–698.

[53] W. S. Grolnick, S. T. Gurland, K. F. Jacob, W. Decourcey, The development of self-determination in middle childhood and adolescence, Development of achievement motivation (2002) 147–171.

[54] R. Zhao, A. Papangelis, J. Cassell, Towards a dyadic computational model of rapport management for human-virtual agent interaction, in: Intelligent Virtual Agents, Springer, pp. 514–527.

[55] A. Assor, H. Kaplan, G. Roth, Choice is good, but relevance is excellent: Autonomy-enhancing and suppressing teacher behaviours predicting students' engagement in schoolwork, British Journal of Educational Psychology 72 (2002) 261–278.

[56] R. Looije, M. A. Neerincx, J. Peters, Integrating robot support functions into varied activities at returning hospital visits, International Journal of Social Robotics (in press).

[57] J. M. Kessens, M. A. Neerincx, R. Looije, M. Kroes, G. Bloothooft, 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.

[58] 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, 2012 IEEE, IEEE, pp. 718–724.

[59] M. Tielman, M. 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, ACM, pp. 407–414.

[60] R. R. Hoffman, M. Johnson, J. M. Bradshaw, A. Underbrink, Trust in automation, Intelligent Systems, IEEE 28 (2013) 84–88.

[61] K. D. Kinzler, K. H. Corriveau, P. L. Harris, Childrens selective trust in native-accented speakers, Developmental science 14 (2011) 106–111.