Embodied Quantification of Self
Motivating and Informing Action in Self-Tracking

Victor Nihlwing
Andreas Philippi

Department of informatics
Human Computer Interaction and Social Media
Master thesis 1-year level, 15 credits
SPM 2017.08
Embodied Quantification of Self – Motivating and Informing Action in Self-Tracking

Abstract

Technical advancements allow for increasingly sophisticated methods of self-tracking. Despite this, the ways in which we interact with our numerical representations seem not to have progressed equally, making it challenging to use the data in meaningful ways. This prevents us from making the most of self-tracking in order to facilitate a healthier lifestyle and self-improvement. In this study, we show how Dourish’s Embodied Interaction can motivate acting based on self-tracked data, with the example of walking. We conducted evaluations with experts and users of a software prototype that is built on the notion of embodiment. Based on the results, we draw a number of conclusions about the usefulness of Embodied Interaction in this area: That digital applications can support physical activity through providing context, motivation and feedback; that self-tracking applications should focus on goals rather than data; that motivation might be increased by placing the users efforts in a context that transcends them as individuals; and that Embodied Interaction offers a rich field of possibilities which are yet to be discovered.

Keywords: Embodied Interaction, Self-Tracking, Quantified Self, Persuasive Technologies, Motivation

1. Introduction and Research Question

In 2001, Fogg and his research team examined 72 health-tracking applications for mobile devices, criticizing how their unimaginative implementation of “fairly mundane strategies” (Fogg, 2002, p. 186) fails to support users in their striving for a healthier lifestyle. Nevertheless, they acknowledged the general potential of mobile devices and corresponding health applications to offer a motivating and engaging user experience (Fogg, 2001). 16 years later, the topic of quantified self, i.e. collecting data about oneself, is considered to be one of the most important trends in healthcare, with, for example, 40% of the participants in a German study expressing high or rather high interest in the subject (Statista, 2017).

However, despite technical advancements in data gathering for self-quantification, the ways in which we consume the output seem not to have progressed equally. As Walker-Rettberg points out, large parts of the collected data lacks context and clear definitions, making it “useless, mere decoration, eye candy” (Walker-Rettberg, 2014, p. 75). While we tend to attribute a certain authority to what looks like facts, we miss that they still need interpretation, the knowledge for which is often lacking (Walker-Rettberg, 2014). Research
suggests that this is true not only for casual users, but also for ‘extreme users’, as they too appear to be confused about which data to collect and how to read it (Choe, Lee, Lee, Pratt & Kientz, 2014).

This issue has been addressed before from the perspective of persuasive technology. Persuasive technologies are systems that are designed with the intention to encourage a specific behavior or attitude (Fogg, 2002) and as such, they might be a natural choice when it comes to health applications and self-tracking. Fogg (2002) predicted this development and indeed, there has been numerous research in that area (e.g. Kelders & van Gemert-Pijnen, 2013; Daskalova, Ford, Hu, Moorehead, Wagonon & Davis, 2014; Ploderer, Reitberger, Oinas-Kukkonen & van Gemert-Pijnen, 2014; Sarzotti, Lombardi, Rapp, Marcengo & Cena, 2015). As we will elaborate later on, this research deviated from the traditional notion of self-quantification as creating a number-based digital representation of oneself, the so-called data double (Ruckenstein, 2014). Instead, they inform a more healthy behavior not through suggesting that the user is in absolute control (Walker-Rettberg, 2014), but by depicting their data double in terms of something else, i.e. a metaphor. An example for this is the study Fish’n’Steps (Lin, Mamykina, Lindtner, Delajoux & Strub, 2006) which illustrated the personal activity levels of participants as a fish that could have different sizes and moods.

Fish’n’Steps and similar alternative visualizations were found to improve motivation for activity, but often only did so if the user met specific preconditions. Put differently, participants struggled to attribute relevance for themselves to the presentations. It is here where we think that a different approach could be helpful, and this approach is Embodied Interaction. Embodied Interaction is a perspective on the relationship between people and technologies that emphasizes the participation of technologies in the world rather than being an addition to it – and that it is in such situations where meaning arises (Dourish, 2004). Its focus on turning “action into meaning as part of a larger system” (Dourish, 2004, p. 183) simultaneously hints at what might be a problem in the related research and offers a solution; perhaps motivation arises and reflects in user behavior when the self-tracked data is contextualized in the ‘right’ way. Hence, we attempted to assess the usefulness of theories of Embodied Interaction in the context of self-tracking in order to answer the following research question:

What are ways in which Embodied Interaction can help us to better inform and motivate action in the context of self-tracking?

To this end, we performed expert evaluations and empirical user studies on a software prototype. While the prototype incorporates elements that will seem familiar to readers who know about traditional persuasive technologies, its design rationale stems from Embodied Interaction. We are aware that Embodied Interaction is by no means a strict design framework. Hence, there might be different ways of approaching the same research question. However, in our design, we decided to follow the line of metaphorical representation of data. For the sake of this study, we focused on physical activity or, more specifically, counted steps. But the metaphor we use is grounded in the reality of global warming and relies on real and
open data provided by the World Bank¹. Thus, the design has a dualistic character: Depending on the user, it might encourage walking more with the side-effect of acting more environmentally-friendly, or it might encourage a more environmentally-friendly behavior with the side-effect of walking more. Both can arguably considered beneficial for the user.

Self-quantification has the potential to be a valuable tool for supporting a healthier lifestyle and self-improvement. Both, however, seem to be in need of different ways of how we understand and engage with them. With our work we hope to contribute to exploring such ways.

2. Theoretical Background

This section introduces the foundations that underlie the study. First, we give an overview about Embodied Interaction and its design principles. After that, we focus on two of these principles in more detail, before introducing a set of heuristics which will serve as a frame for discussing embodiment during the rest of this work.

2.1 Embodied Interaction

Embodied Interaction is a perspective on the relationship between people and technologies, that was first formulated by Paul Dourish in his book Where the Action is (2004). Dourish sees technology not merely as a matter of engineering, but also as an expression of specific philosophical stances that determine the way in which computer systems represent the world. In the case of Embodied Interaction, he mainly draws from the phenomenology of philosophers such as Husserl, Heidegger, Schutz and Merleau-Ponty. For Dourish, their work describes three important features of embodiment: Firstly, they share a notion of embodiment being some sort of participation in the world rather than just existing in a physical form. Secondly, there is a mutual effect between embodied actors and their environment in so far as that both are influenced by the other through the accomplishment of activities. Finally, it is from these interactions that meaning arises. Consequently, Dourish defines Embodied Interaction as “the creation, manipulation, and sharing of meaning through engaged interaction with artifacts” (Dourish, 2004, p. 126).

The term meaning, in this context, can be regarded from three different perspectives. Dourish distinguishes between the ontology, intentionality and intersubjectivity of meaning. Ontology concerns how we understand the structure of the world by interacting with it, and our ability to separate various entities, i.e. independent objects or beings. It also touches upon how we reflect and perceive the connections between different or various kinds of entities. When it comes to applying ontology to interface design, Dourish refers to the separate works of Norman (1988) and Gaver (1991) regarding the concept of affordances. It describes how artifacts in computer systems are designed to afford certain actions, and they reveal themselves mostly through interaction with them. A computer mouse, for example, affords two-dimensional movement and pressing the left and right buttons. Thus, affordance is about not only allowing, but encouraging certain kinds of activities in order for the user to understand how to interact with a system or the setting in which it should be used.

¹ World Bank Data: http://data.worldbank.org
Intentionality describes the ability to make a connection between one entity and another. Framed in the context of digital systems and considering that, as Dourish points out, computation is essentially representation, it is about supporting the connection between an entity and the representation of that entity. Intentionality therefore relates to the interpretation and understanding of technology: In what way is the meaning of an action communicated to the user? Is the action visible or is it its consequences that are displayed to the user? How should the action be represented to afford the intended interpretation? In accordance with Dourish’s understanding of phenomenology that meaning emerges from participation and action in the world, interpretation and understanding do not simply spring from providing information, but from applying it to a specific context and purpose.

The third perspective, intersubjectivity, implies that meaning is not understood by a single individual, but rather by the individual as a member of a community of practice. Such a community of practice shares a common view and an understanding with regard to how they implement and engage in practical activities. As knowledge is shared within the community, its members approach and interpret the activities in a similar and characteristic fashion. Intersubjectivity expands beyond affordance to also include the common expectations and the evolving values of the target group.

2.2 Design Principles of Embodied Interaction

Dourish’s (2004) considerations lead him to a set of design principles: Firstly, that computation is a medium where the actual content is not the information, but the way in which it is presented. Secondly, that the meaning of said content is determined by several factors such as social and practical context. This meaning, furthermore, is created and managed by users rather than designers, although it is the designer’s responsibility to afford a specific interpretation by the user. Further, embodied technologies are not just an addition to the world they represent, but an integral part of it; and, finally, they enhance action in that world with meaning.

In summary, Dourish describes embodiment as being:

“[… about engaged action rather than disembodied cognition, it is about the particular rather than the abstract, practice rather than theory, directness rather than disconnection.” (Dourish, 2004, p. 189)

As such, the concept of embodiment appears to be suitable to address the typical problems of self-quantification that we touched upon earlier: How users struggle to act upon the data they gathered about themselves. We consider especially two of the design principles to be relevant in this context: The notion of computation as a medium as well as the goal to turn action – the action of gathering data – into meaning. These design principles will be our focus and are explained in further detail below. However, since all principles stem from the same foundation, many of their ideas overlap. This makes it hard to clearly separate them. As a consequence, the other principles influenced our design decisions as well.
**Principle 1: Computation is a Medium**

Embodied Interaction emphasizes the participation of technologies in the world. This leads to a dualistic character of digital technologies: On the one hand, they do not actually contain the real-world entities they refer to, but a representation of them. On the other hand, that representation is not completely separate from its real-world counterpart. Some insight that a user gains from information in an application might lead them to act in a certain way, which in turn might reflect back on the state of the representation.

As we have shown, contemporary self-tracking applications are often weak at managing the connection between representation and participation. The notion of computation as being a medium introduces a different way of thinking about representations. As Dourish argues, computation offers an opportunity for “communication between a designer and a user through the medium of the system itself” (Dourish, 2004, p. 163). Dourish understands this communication mainly as modulation, i.e. the transformation of data in some way to carry specific information. The act of transforming is also referred to as *encoding*. It follows then, that the encoded representation is not so much about the initial data anymore, but about the meaning carried by the modulation and how it “affects practice, by transforming it, restricting it, or extending it” (Dourish, 2004, p. 164). However it is obvious that the encoded knowledge is of little use unless users are able to decode it again, and this is where the next design principle is helpful.

**Principle 2: Embodied Interaction Turns Action Into Meaning**

The second design principle we are focusing on is that “Embodied Interaction turns action into meaning” (Dourish, 2004, p. 183) in the ontological, intentional and/or intersubjective way. Dourish follows Heidegger’s reasoning who argued that technology – or, as Heidegger calls it, *equipment* – is merely a tool used to achieve an end or goal, rather than being a solution in itself. Thus, meaning is created through action, or interaction, with an artifact in a context to accomplish a purpose or reach a desirable result. It is therefore not found inside an application but in the way the application is used.

For Dourish, digital technologies forge meaning primarily in how they present knowledge. Intentionality should be supported by modulating and contextualizing raw information. The interpretation of this ‘meaningful information’ should be supported ontologically, i.e. by guiding the user towards the desired way of understanding through the design of the application. However both how information is modulated and how its interpretation is supported need to consider the intersubjectivity of meaning as well; it is not only the individual interacting, but also their community whose values are part of the lens through which the individual perceives the world.

2.3 Heuristics for Embodied Interaction

Evaluating a design artifact requires criteria of some sort. In HCI, it is common to use heuristics (Benyon, 2014) for that purpose, and some of the most widespread heuristics were introduced by Nielsen (1994). His *10 Usability Heuristics for User Interface Design* comprise principles that define features of good design with regard to usability, error prevention, visibility, aesthetics and consistency. The purpose of these heuristics is, on the one hand, to act as a framework to lean on when developing artifacts. On the other hand, they
provide a structure for evaluating existing designs. By comparing how well the design corresponds with the various principles, designers and evaluators alike can discover faults or parts of the system that are in need of further polishing before conducting user testing.

Due to the rather vague nature of Dourish’s design principles, we felt the need for an intermediary step between concrete design and the underlying theory of Embodied Interaction. Therefore, we formulated a set of heuristics for Embodied Interaction. Their purpose is to provide us as well as possible evaluators with common terms when discussing features of designs from the perspective of embodiment. These heuristics are not entirely new. Instead, they build on Nielsen’s (1994) already established heuristics, but reinterpret them against the backdrop of Embodied Interaction and, more specifically, the two design principles we explained earlier. In this process, we left out those of Nielsen’s (1994) original heuristics that mainly focus on usability and aesthetics. Although we still consider them important, we do not think that they are of primary concern for Embodied Interaction. Further, some of the original heuristics are merged into one of our heuristics, because we see them as different aspects of the same facet of embodiment. While this approach certainly narrows the focus and therefore runs the risk of missing out certain aspects due to the choice of heuristics, we think that it sufficiently covers the area of interest of this work. As a result, however, future research would need to critically examine the applicability of the heuristics with regard to the respective requirements.

Much of the difference between the original heuristics and the reinterpretation can be explained by extending the scope of Nielsen’s (1994) heuristics. Nielsen (1994) mainly focuses on the system and draws a distinction to the real world. Grounded in the concern of Embodied Interaction with participation in the world, we suggest to ‘soften’ this distinction as shown in Fig. 1. Instead, we consider not only the system, i.e. the application, itself, but also its user, the practice and context in which both are embedded. For the sake of lingual simplicity, we refer to these things outside the application as surroundings. With that in mind, the following sections introduce each heuristic and explain their foundations in Nielsen’s (1994) heuristics as well as Embodied Interaction.

![Fig. 1: Application and surroundings](image)
Heuristic: System Consistency

The system fits into the practice. It is an embedded part of that practice which is transformed, extended or restricted by the system. Thus, the system serves as a tool within the practice rather than being an end in itself.

Meaning, according to the design principle that Embodied Interaction turns action into meaning, is not inherent to a system. Instead, it arises in how the system supports and influences a practice. One of the ways in which it could do so is through what Nielsen (1994) describes as matching: By using user-oriented language and structure, the boundaries between the system and the surroundings blur and it becomes more integrated in it rather than feeling like an extension. The language and structure should be consistent across the system and its surroundings, as pointed out by Nielsen (1994). Another useful technique, taken from Nielsen’s (1994) heuristics as well, is to offer shortcuts for more experienced users.

Heuristic: Visibility

The meaning of actions is clearly understandable, even before the action is taken by the user. The impact on the surroundings of the application is visible, e.g. through constant feedback. The responses of the system to the actions are engaging for the user.

Nielsen emphasizes the importance of continuous feedback to inform the user about the status of the system. Extending this to the surroundings of the system, users need not only to be able to understand the effect of their behavior, i.e. the representation of their activity, but also how this reflects back to the world. Furthermore, visibility crosses time and space, encompassing the “visibility of the system’s behavior in response to some user activity, whether that activity is being enacted or simply considered” (Dourish, 2004, p. 165).

Heuristic: Encoding And Decoding Of Knowledge

Knowledge (e.g. information, data) is not simply displayed, but modulated. Modulation means that the raw information is changed in a way that benefits a specific understanding by the user. This could include putting it into a specific context. Users are able to understand the meaning carried by the modulation as well as the context. Thus, they can transfer the knowledge to their behavior in the real world.

According to the principle of understanding computation as a medium, the way knowledge is encoded carries the actual information. As described earlier, contextualizing the information is one way of making it meaningful. Encoded knowledge, however, can only be put to practice if the user is able to decode it again.
Heuristic: Communication

Meaningful interpretation of the knowledge through the user is supported by the system rather than relying on the user alone. The relations between the different parts of the system itself and the surroundings are understandable.

This heuristic can be seen as an extension of the previous heuristic. It specifies how the decoding of knowledge by the user should be supported. Referring to the concept of affordances and the phenomenological understanding that meaning resides in the world, the design should guide the user to the desired interpretation rather than just assuming that this happens automatically in the user’s mind. This analogous to Nielsen’s (1994) heuristic of preferring recognition over recall, thus offloading cognitive load to the environment, and could be supported by help and documentation, another one of Nielsen’s (1994) heuristics.

Heuristic: Error

The system helps to prevent, recognize, diagnose and recover from errors not only within the system, but also in the practice that it supports.

Finally, the goal of recognizing, diagnosing and recovering from errors can be almost entirely transferred from Nielsen’s (1994) heuristics. The difference, however, lies in the scope of the error. Instead of correcting errors in how the system works, embodied technologies should also consider errors in the practice which they are supporting.

3. Related Research

When it comes to software that tries to encourage certain behaviors, the field of persuasive technologies offers rich research. Hence, this section introduces the relation of persuasive technologies and self-tracking as well as examples for projects that try to increase motivation to be more physically active. After that, we summarize the current state of software in the field of self-tracking. Finally, we illustrate the concept of Embodied Interaction by giving examples of how it has been implemented so far.

3.1 Persuasive Technologies

Persuasive technologies comprise “interactive computing systems designed to change people’s attitudes and behaviors” (Fogg, 2002, p. 1). As such, an application like ours with the goal of motivating people to pursue a more active and environmental-friendly lifestyle can be considered persuasive. While the techniques used to alter user behavior may vary depending on the purpose of the application, Fogg (2002) predicted that, in one way or another, persuasion will be especially relevant when it comes to self-tracking. He argues that advances in digital and especially mobile technologies make – and will continue to make – it easier to gather data about oneself. That, he thinks, is intrinsically motivating due to the “natural human drive for self-understanding” (Fogg, 2002, p. 44). Furthermore, the feedback these applications provide can give their users an understanding of how well they do in a specific activity which, as a consequence, increases the chance that they keep engaging in it.
Indeed, the use of persuasion techniques in self-tracking has been subject to numerous research projects (e.g. Daskalova, Ford, Hu, Moorehead, Wagnon & Davis, 2014; Kelders & van Gemert-Pijnen, 2013). Some of them borrow elements that are known to increase motivation from games (gamification). *Fish’n’Steps* (Lin et al., 2006) is one of the earlier examples for such efforts. In their study, Lin et al. (2006) illustrated the number of steps a user took through a virtual pet fish. Its size and mood reflected the activity level of its ‘owner’. The fishes of the participants were grouped in several teams, each with their own fish tank, to add a sense of competition. The study concluded with an overall positive effect on motivation to engage in physical activity. This effect, however, was mostly raising awareness and as a result incorporating new habits for participants with low or moderate levels of activity, and not so much for those who already were quite active.

A more recent study was conducted by Ludden, Kelders & Snippert (2014). Similarly to Lin et al. (2006), they used a combination of persuasive techniques and game elements. Their goal was to motivate users to engage in positive psychology exercises to foster personal growth. This was metaphorically visualized through a journey. More explicitly than Lin et al. (2006), they were interested in how useful metaphors are for increasing user motivation. Overall, their findings are similar to the Fish’n’Steps project, discovering an increased interest in engaging in the respective exercises. This effect was attributed, at least in part, to the use of a metaphor. However looking at different user groups, they found a contradiction to the results of Lin et al. (2006), stating that they mainly succeeded in increasing motivation for participants that already were rather motivated prior to the study, while failing to attract attention from participants that had no particular interest in the subject before.

As we pointed out earlier, these rather creative advancements have not fully reached mainstream market yet. While indicating the possibility of increasing motivation through the use of persuasion and illustrative metaphors, the works of Lin et al. (2006) and Ludden et al. (2014) appear to struggle with addressing more than only very specific user groups. We think that our prototype circumvents this issue by using a metaphor that contextualizes personal behavior in terms of a reality that affects everyone.

Persuasive technologies are discussed not only with regard to the ‘how’, but also the ‘if’. Fogg (2002) argues that persuasion per se is not unethical, although certain intentions, methods and desired outcomes can make it so. An intention that can almost never be unethical, according to Fogg (2002), is persuading users to live a healthier life. However it is obvious that the view of what constitutes a healthy lifestyle can differ or change. One way to at least try to incorporate more diverse views is suggested by Karppinen & Oinas-Kukkonen (2013). They advise the combined use of guidelines, analyzing stakeholders and involving users in order to ensure that all interests are considered. Atkinson (2006) goes a step further and questions that there even is any case at all where persuasion is ethical. What she proposes instead is a re-conception of persuasion as education, i.e. presenting information and trusting in the user’s ability to arrive at the right conclusions themselves. From that perspective, one might argue that our design goes a step too far: While it uses real data to educate users, it also suggests a certain behavior as a consequence. Nevertheless, we think that this approach is ethically justifiable, since the prototype clearly communicates not only its purpose, but also the reasoning behind the advice it gives. In that sense, Embodied
Interaction provides an approach for persuasion through education by making information relevant for the user’s own behavior.

3.2 The State of Quantified Self Applications

Self-tracking is far from a new concept; people have been keeping logs and records about themselves or their habits for hundreds of years in different ways. However, the rapid advancement in wearable technology we have seen lately has made self-tracking increasingly accessible, and it is now easier than ever before for users to automatically collect and store large amounts of personal data (Walker-Rettberg, 2014). Today, there is a vast number of self-tracking technologies and devices to choose from, allowing people to track almost any aspect of their daily life, such as their amount of exercise or caffeine intake. The growing popularity in self-tracking resulted in the Quantified Self (QS) movement that emerged in 2007. It established the motto “Self Knowledge Through Numbers” (Choe, Lee, Lee, Pratt & Kientz, 2014). The general focus of the QS-movement seems to have been directed towards fitness and exercise, as a way to easily collect large amounts of personal data – with the assumption that it will help to improve the users’ performance in whatever activity is being tracked (Hoy, 2016).

In a study from 2006, Consolvo, Everitt, Smith & Landay set out to fight overweight and obesity among women, primarily by encouraging spontaneous activity using Houston, a personal fitness journal which enabled the users to share and track their progress towards a daily exercise goal (Consolvo et al., 2006). The study identified four requirements that should be kept in mind when designing applications focused on encouraging physical activities. These consisted of giving the user sufficient insight on their progress as well as providing credit when reaching set goals, implementing support for social influence, and, finally, considering the practical limitations that the user's lifestyle might impose (Consolvo et al., 2006).

Another group that have adopted QS technologies are parents, according to Walker-Rettberg (2014). Rather than tracking themselves, parents tend to use quantified self devices to measure the growth and well-being of their babies and children to make sure they are fed as well as grow and sleep properly. As Walker-Rettberg points out, these measures often cause concern for worried parents when the results temporarily deviate from what they expect. This concern, however, seems to be facilitated mostly by missing knowledge about how frequently certain types of data should be gathered on the one hand, and the respective QS applications encouraging excessive data collecting on the other hand (Walker-Rettberg, 2014).

In an extensive study that analyzed 52 video-recorded talks from QS meet-ups, Choe et al. (2014) managed to identify a number of common pitfalls when it comes to collecting, analyzing and interpreting data from self-tracking devices. Surprisingly, even ‘extreme users’ in the QS movement found it difficult to conduct self-tracking efficiently, struggling with aspects such as tracking too many things, analyzing the data correctly and making correct assumptions about their performance based on the data. Choe et al. (2014) conclude that even if users who are active in the community can learn from themselves and others mistakes
to improve their results, better designed self-tracking technology that clearly inform its users about their actual results would be beneficial.

### 3.3 Applications of Embodied Interaction

The majority of projects informed by Embodied Interaction today seem to lean towards the field of tangible computing. Zarin (2017) describes this as an attempt by designers, influenced by the theories of embodiment (Dourish, 2004), to move away from the screen-based interaction many current systems rely on, and approach a more practice-friendly design. One such example is the work by Sarzotti et al. (2015) on motivating users in self-reporting their mood through a tangible user interface (TUI). The TUI consisted of a wooden cube containing an Arduino system. Each of the six faces of the cube represented a different mood. The Arduino system was able to recognize the surface and the correlating mood facing up and would report any mood changes, for instance when the user flipped the cube, to a database. The ambition was to find an enjoyable and easy way of allowing users to track their mood, coming from the impression of how many current systems fail to motivate their users sufficiently, resulting in the users to simply forget or ignore the task of self-reporting.

In their paper *Touching the Dematerialized*, Campenhout, Frens, Hummels, Standaert & Peremans (2016) explore several ways to address tangibility. The emergence of mobile devices such as smartphones and tablets has, according to the authors, limited our interaction with these technologies “to button-pushing or a set of standardized gestures on a display” (Campenhout et al., 2016, p. 148). One design project featured in the paper consists of a physical, single-purpose alarm clock, meaning that the artifact has no other functions than those that are expected from a traditional alarm clock. To set a new alarm, the user has to physically alter the position of an alarm-module which will change the appearance of the clock, and thus indicate its current state simply through the clock’s present shape. The fact that the alarm-module is ‘loaded’ into the clock, as an effect of this action, aims at further enhancing the tangible aspect of the artifact (Campenhout et al., 2016).

Another example is Subramonyam’s Magic Mirror; a work-in-progress system which seeks to display self-tracking data from wearable devices “using the body as a reference frame” (Subramonyam, 2015, p. 1699). The system comprises a Kinect camera, a computer, a projector and a display. The user’s self-tracking devices send fitness data via a cloud server to the computer system where the data is stored. To access their data, the user approaches the Magic Mirror, i.e. the display where the video captured by the Kinect camera is projected. By performing certain gestures in front of the Magic Mirror, the corresponding data of the user will be projected to the screen, sometimes enhanced by additional graphics. If the user, for example, would place their hand above their heart, the Magic Mirror will visualize the current heart rate, as well as a beating heart. Other data that can be visualized includes burnt calories, steeps taken and sleep, each initiated by a unique gesture (Subramonyam, 2015). Further, a color-coded silhouette is projected around the users ‘mirror image’ to indicate whether the user has performed well (yellow-green silhouette) or needs to improve their behavior (orange-red silhouette).

Although the tangible approach seems to be the most common when it comes to the present research in the field, it is not the only way to employ Embodied Interaction. As
embodiment strives to blend our tools and devices into the setting and adapt them to fit better into our practices, we are still to discover new ways for applications to utilize its full potential (Campenhout et al., 2016).

4. Research Process

The goal of our research was to evaluate if and how the ideas taken from Embodied Interaction are useful for informing and motivating action in the context of self-tracking. We assessed the usefulness of these theories implicitly through an interactive prototype that incorporates them. Prototypes are typically used to communicate or test concepts during the development of a product (Benyon, 2014). A prototype can be characterized by defining its intention, i.e. whether it is meant to address the role the final product will play in its users’ lives, the look and feel of the product or its technical feasibility (Houde & Hill, 1997). However as Lund (2003) points out, this notion of prototyping seems to exclude other aspects such as the evaluation of a theory, even though especially design artifacts in HCI usually have some kind of claim, i.e. a theory or certain assumptions, inherent in them (Carroll & Kellogg, 1989). Thus, it should be possible to evaluate these claims as well.

As suggested by Lund (2003), we combined artifact-oriented methods, namely expert evaluation, with empirical user testing in order to learn both about the anticipated as well as the actual user experience.

4.1 Ethical Considerations

While conducting both types of evaluation and processing the data, we strived to create a welcoming atmosphere for our participants as well as follow the guidelines for ethical research issued by the Swedish Research Council (Hermerén, 2011). Thus, we informed our participants about the following aspects prior to any evaluation: The purpose of the study; that any participation was voluntary and could be ended at any time; that their identities and the collected data would be handled confidentially; and finally that the data would only be used in the context of this research. We also assured them that we are interested in their critical feedback regarding our design, rather than evaluating their ability to work with it.

4.2 Overview

The steps that led to the empirical testing with users can be summarized as follows:

1. **Initial Prototype:** A first version of a design prototype that is based on the ideas of Embodied Interaction was prepared. This step was already done before this work started.

2. **Expert Evaluation:** Two experts independently evaluated the initial prototype based on the heuristics established in section 2.3.

3. **Iteration on the Prototype:** The feedback given by the experts during the evaluations was considered and contributed to a number of changes made to the prototype.
4. **Interviews with Users:** The revised prototype was finally presented and evaluated with potential users during three pair interview sessions.

This process is comparable to the Deming cycle (Brüggemann & Bremer, 2012), also known as *Plan–Do–Check–Act:* A design or changes to the design are first *planned,* then quickly implemented and tested (*Do*). The results of that test are *checked* to be aligned with the goals before the design is introduced to a broader audience (*Act*). In that sense, the initial prototype already comprises the first *planning* and *doing* of the design. The result was checked during the expert evaluations. Since the prototype is not intended to be publicly available, we skipped the *Act* step and went through another iteration of *Planning* changes, implementing and testing them with users and evaluating the results. Each step is explained in more detail in the following sections.

### 4.3 Initial Prototype

*ClimateQuest* is a prototype that connects a step counter with climate change data taken from the World Bank's database. It encourages a healthier behavior in terms of being more physically active by prompting the user with challenges based on climate change data. Thus, the design has a dualistic character: Depending on the user, it might encourage walking more with the side-effect of acting more environmentally-friendly, or it might encourage a more environmentally-friendly behavior with the side-effect of walking more.

The design is meant as a smartphone application for people that are generally interested in the topic. After launching the application, the user is presented with a chart and some accompanying explanation on one specific data set, both world-wide as well as specific to their current location (Fig. 2). After reading the information, they are presented a challenge, i.e. a task that is explicitly related to the information and the set of data they just saw (Fig. 3). Once they accepted the challenge, they work towards solving it (Fig. 4). A real-world implementation of the design would consider external information such as step count from a fitness tracker to decide if the goal of the challenge was actually reached. Once the challenge is solved, the user sees the data from the beginning again (Fig. 5). But this time, it is related to their personal achievement. The application calculates a hypothetical third value, besides the world-wide average and local value. This third value tells the user what the data would look like if everyone incorporated that challenge into their lives. Thereby, it demonstrates the user how individual actions affect us all on a large scale.

In the following sections, we summarize the motivation behind the prototype design. In order to make it more accessible, we use the structure provided by the heuristics for Embodied Interaction we explained earlier.

**System Consistency**

The application fits into the practice in the sense that it is run on a smartphone – a device that is owned and carried by almost everyone and often already used for tracking personal data. The Health app on newer iPhones, for example, has a built-in step counter that works continuously in the background. Similarly, the sample challenge in ClimateQuest is solved automatically by counting steps, as opposed to requiring manual input by the user. This leads to an experience where the user is unlikely to spend a lot of time in the application itself, but
is motivated in their everyday life to walk more. Thus, the user’s practice of getting around is influenced by the application.

**Visibility**

Accompanying information and descriptions on every screen explain the reasoning behind the actions (Fig. 2 and 3) as well as the result (Fig. 5). The labels on the buttons are formulated in a direct, conversational way that is both engaging and makes clear what will happen when pressing it.
As the effect that one person can have on climate change is so small, it is hard to give feedback on the real impact of the user’s behavior on the environment. Here, the hypothetical third value mentioned earlier is used to nevertheless emphasize the responsibility of everyone.

**Encoding and Decoding of Knowledge**

The display of the data in the prototype is transformed so that scales and labels from the charts are removed. The idea behind this was that it might not be that important to know the exact data if the graphs are meaningful enough to convey trends. In that case, exact knowledge is omitted in favor of a less cluttered user interface, while the general feeling for what is shown is kept. This transforms the numerical data to a simpler version, carrying the information of trends and developments rather than details. The user’s actions (of walking) are connected to the data – and thereby embedded in an overarching context – in two ways: Firstly, by the descriptions on the screens; and secondly again by putting their efforts into relation with the data through the aforementioned hypothetical third value.

**Communication**

Again, the descriptions on each screen support the understanding of the data in a specific way rather than relying on the user’s ability to figure the meaning out on their own. This is supported by a consistent color coding throughout the application. Parts of the graphs related to the user directly, for example, are colored blue, while the corresponding parts in the descriptions are highlighted in the same color.

**Error**

While the rather strict architecture of the prototype itself allows for few errors within the application, the overall purpose of the design is to remind its users of their responsibility. In that sense it implicitly prevents error by encouraging more awareness in everyday behavior.

**5. Data Gathering and Findings**

This section describes our data gathering methodology, results and reflections for each of the steps in our process.

**5.1 Expert Evaluation**

Previous to conducting user testing, we evaluated the existing prototype with expert evaluations. The aim was to provide us with quick feedback about the quality of the prototype before engaging in the more time-consuming user testing – an approach that has been referred to as *discount evaluation* (Preece, Rogers & Sharp, 2002). We were mostly interested in two kinds of feedback: On the one hand, we looked for ways in which we could improve the implementation of Embodied Interaction theory in the design artifacts. On the other hand, we wanted to know about general problems in the user interface that could distract participants during the user testing from the features we want to evaluate. In both cases, expert support can help to reveal problems that we as the creators of the design are blind for due to our involvement in the underlying theories and thoughts (Benyon, 2014).
Because of the different types of feedback we were looking for, we conducted two separate evaluations: One with a domain expert in the area of quantified self and a second one with a usability expert. Both sessions were preceded by sending a briefing document (see appendix 1) to the experts as well as giving them access to the prototype. The briefing document contained information about the background of our study, the purpose and process of the evaluation, and how the data was going to be used. Furthermore, it listed and explained the heuristics we established earlier, so we could use them during the evaluation to structure the feedback. The prototype was introduced using a scenario which puts it into a broader context (Benyon, 2014). During the evaluation, the experts navigated through the prototype on their own while keeping the scenario and heuristics in mind. They noted down their reflections regarding the various heuristics in an online form. Afterwards we discussed their answers together to make sure we were understanding their notes correctly. This part of the evaluation was recorded with the consent of the experts just in case we would be in need of more data in addition to the forms.

Although our aim was to perform the evaluations as similar as possible, a few differences are worth mentioning. Unfortunately, the domain expert had overlooked the briefing document included in our initial email conversation, and was consequently less familiar with the system and the heuristics. In addition to that, we noticed a difference in how the experts dealt with the concept of heuristics. The usability expert was more familiar with this method and therefore took very precise and extensive notes. The domain expert, on the other hand, submitted only brief notes, but provided us with additional information during the discussion. Despite these differences, we took measures to ensure the trustability of the data (Graneheim & Lundman, 2004). Firstly, expert evaluations are a well-established practice in HCI, and getting the help of experts with two different domains of expertise promises a broader perspective. Both contribute to the credibility of the data. Secondly, the information provided in this document on the background and conduct of the work should help readers determine whether the results are transferable to their own research. Finally, both of us individually conducted the same analysis of the data before we merged our results and discussed the differences. This was to lower the possibility of missing important aspects.

To analyze the results of the evaluation sessions, we performed a qualitative content analysis. In doing so, we mostly followed the approach described by Harr, Nyberg, Berggren, Carlsson & Källstedt (2016), which in turn is based on the work of Graneheim & Lundman (2003). Similar to Harr et al., we considered this a suitable method because we were interested in qualitative information and, more specifically, in the categories and themes that are explicit (manifest) or implicit (latent) within the information. We are also aware that any qualitative analysis of a text is to a large extent subjective interpretation. In that context, we think that this structured approach helps to maintain a certain level of objectivity while making the rationale behind our interpretation accessible to the reader.

The process, in analogy to Harr et al. (2016), consists of identifying meaning units, shorten and rephrase them into condensed meaning units and label each of them with a code. The codes can then be grouped in subcategories and categories. Finally, the categories are divided in themes, which express the latent meaning underlying the categories. Examples for this coding process can be found in appendix 3. In our case, the codes are the individual
comments on the design. The categories group codes that refer to the same parts of the design. The themes describe the intent of the comments, for example if they are remarks, praise or parts that need improvement. On a side note, this approach can be seen as a way of abstracting content from the original format of the data. In that sense, it helped us to make the results of the two expert evaluations comparable, despite them being present in different format.

**Results**
The analysis of the expert evaluations yielded a number of conclusions about the current state of the application. During the sorting of the codes, three overarching themes emerged: *Praise* for the current state, *room for improvement* and *remarks*.

In general, the experts were intrigued and positive about the overall concept and its implementation in the prototype. Most of the codes extracted from the data were related to the Praise theme, such as “[...] I really like your idea of your application, it’s really, really interesting”. However, more interesting in the context of this study is where the experts saw room for improvement. Both experts criticized a confusing use of icons in the prototype that might distract the users evaluating it. Another aspect that was brought to our attention was the strict flow of navigation in the app, which could potentially result in the users feeling ‘locked’ in the system. Further, they mentioned that this strict flow might also be perceived as inconvenient and confusing when users decide to navigate back and forth. They also noted that some elements, namely the buttons to proceed to the next page, were not aligned with the layout on the previous screen, giving the prototype an inconsistent feel. One of the experts also pointed out that solving the challenge was too obviously a simulation as well as that the simulation contained a bug on top of that. This raised the concern that the prototype might feel too unfinished in order to be evaluated in the way we planned. Finally, some minor issues that were found suggested a slight change in certain phrasings.

When it comes to the way in which Embodied Interaction informed the design, the main concern was that the data visualization might be perceived as too abstract due to lack of context: “I don’t know how you designed it but this seems a very [...] abstract kind of thing, it doesn’t have any numbers”. We were therefore encouraged to display the data in another, more tangible way. Other feedback covered the way the application ties into practice and everyday use: How, for example, would the application behave throughout the day or when it works in the background? And how would it retain long-term motivation?

These findings gave us a selection of aspects to iterate on and refine before conducting user testing sessions. In the next section, we discuss how we rated the different aspects of the feedback in terms of severity and how they influenced changes in the design.

### 5.2 Prototype Iteration
The changes made to the original prototype are based on the feedback we gathered from the analysis of the expert evaluations. As a first step, we collected all the codes from the *Room for improvement* theme that emerged in the content analysis. After removing the duplicates, we grouped the codes that referred to the same underlying problem. These were then prioritized using the MoScOW method before deciding on how to integrate the necessary adjustments in the prototype.
MoSCoW is a structured, albeit flexible, approach to sorting requirements in the four categories Must Have, Should Have, Could Have and Won’t Have. It is difficult to obtain a universally accepted description of the method and these categories in particular have been criticized for being only vaguely defined (Wiegers, 2013). Nevertheless, we found them to be a useful guideline as long as the project members share a common understanding of their meaning while discussing which code should be classified how. Hence, we agreed on the following definitions, which are based on Benyon’s (2014) interpretation:

- **Must Have** – Faults and problems that are ‘show stoppers’ and have to be fixed in order to be able to get useful information from the upcoming user testing.
- **Should Have** – Important, but the prototype would still be functioning without them.
- **Could Have** – Minor fixes that will be included if they do not turn out to require a lot of effort.
- **Won’t Have** – Might be interesting for future work, but will not be considered in the current iteration.

Having the categories defined, we also needed to take a certain perspective, since this might determine the categorization of some of the codes. Due to the goal and the timeframe of this work, we decided to focus on what we need for the user testing. In that context the importance of an improvement would be relative to two factors: Firstly, how they could increase the ‘embodiment’ of the experience; secondly, whether they would distract users by unnecessarily confusing them about aspects we are not actually interested in, such as technical problems or low-impact icons.

### Results

Table 1 shows the grouping and classification of the codes. In the Must Have category, we included the problems of the challenge solving in the prototype feeling very artificial and also causing a software error, as well as the meaning of the icons being inconsistent and unclear. Although not directly addressing our actual research subject, they appeared to distract the experts from the factors we were interested in more than other problems. Therefore, we

<table>
<thead>
<tr>
<th>Must Have</th>
<th>Should Have</th>
<th>Could Have</th>
<th>Won’t Have</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prototype immersion broken; Bug in simulation mode</td>
<td>Knowledge too abstract; Improve relatability to data; Graphs too abstract</td>
<td>Gradually revealing information</td>
<td>Background work indicator; Leaving the application; Unclear effect of leaving challenge</td>
</tr>
<tr>
<td>Unclear icons; Icon consistency</td>
<td>Screen flow too linear; Home button; Feeling of being locked-in; Added complexity</td>
<td>Button placement changing</td>
<td>Recovery</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unclear language</td>
<td>Multiple challenges</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Longterm perspective</td>
<td></td>
</tr>
</tbody>
</table>

*Table 1: MoSCoW prioritization of results*
replaced the *Simulation* button by a more subtle message prompting the user to manually start the challenge as soon as they are ready. The step counter would then, not like in the initial prototype, slowly start counting in the speed of actual steps and only speed up after a certain amount of steps. This enables users to actually walk around a little bit while the step count increases approximately in the corresponding speed, but without requiring them to actually walk 10,000 steps during the evaluation. Furthermore, the icons were replaced by icons that are more consistent both within the application as well as in comparison with the icon language that is common in other software.

The codes listed in the Should Have category are more closely related to the embodiment of the design. As one of the experts explained, the first approach of removing all numeric values and communicating trends alone might be too abstract, since it does not provide any point of reference at all. However simply adding absolute values such as the amount of CO2 to the graphs would not help this, because, arguably, most people will not be able to relate those to their effects either. Therefore, we decided to pick up the expert’s suggestion of introducing a more tangible measurement unit in the form of a polar bear mascot. The goal was to support a more concrete interpretation of the encoded knowledge. Relating global warming and the annual amount of arctic ice melting to the roaming radius of polar bears might seem a bit of a stretch, but could actually lead to more approachable numbers than the raw climate change data. Another concern which the experts brought up was the very restricted freedom for navigating within the prototype. In the original version, users were only allowed to navigate back and forth within a fixed sequence of screens causing a feeling of being locked-in. It could be argued that this is not a very practice-oriented solution and, furthermore, raised concerns that the communication in the current design might only work due to its low complexity. Extending the challenges would have exceeded the scope of this
work, but we found a compromise by adding more flexibility in switching between the screens as well as a new home screen. The screen contains a list of past challenges which suggests that there is more to discover without actually implementing it (Fig. 6). The Could Have category includes minor adjustments that were not considered crucial by the experts, but also did not require a lot of effort to implement. We placed the polar bear mascot on a small ice shelf in the progress screen and connected the size of the ice shelf to the step count, so that it grows as the progress increases (Fig. 7). This acts as an instant feedback, connecting the previously introduced measurement unit to the action of walking. Put differently, the user’s effort is put into context already while solving the challenge, not only afterwards in the results. Furthermore, we changed the alignment of the buttons on each screen to stay in the same position for the sake of consistency, and rephrased some descriptions that one of the experts found to be hard to understand. Finally, the Won’t Have category mostly comprises technical requirements such as the ability of the application to keep running in the background and to restore its state in case it crashes. These would be considered more important in a real world implementation, but are negligible for our purpose. As mentioned earlier, we also decided to not incorporate any additional challenges due to the high effort of not only designing, but also implementing them. In a similar vein, we made the decision to only focus on the short-term use of the application, since the scope of this study did not allow for a long-term evaluation anyway.

5.3 User Testing
After updating the prototype, we conducted three user testing sessions with two participants in each session. The participants were recruited among the students of Umeå University. They were required to have some pre-existing knowledge or interest in self-tracking, so they would be able to relate our prototype to their own experience. The participants consisted of four males and two females between 23 and 27 years. Again we selected a qualitative approach since we wanted to learn in-depth about the experiences and thoughts of the users while and after they worked with the prototype. Each session included three parts: In the first part, we asked both participants some introductory questions to learn about their background, their physical activity level and their attitude towards – and previous experiences with – self-tracking.

After that, the participants were invited to explore the prototype together as a co-discovery and think aloud while doing so. According to Benyon (2014), co-discovery is an informal method that is useful to learn about first impressions. But, since the participants were chosen to have already known each other on a friendly basis prior to the evaluation, the discussion that usually emerges between the two while conducting the experiment leads to deeper exploration and more in-depth thinking about the situation than it usually would for single persons (Benyon, 2014). As our main focus was to learn more about how they perceive the application in order to draw conclusions about the underlying theories, our hope with this was to elicit feedback that was less superficial than just first impressions.

Once the participants finished the exploration part, we conducted semi-structured interviews (Benyon, 2014). The choice of this technique allowed for a natural flow of conversation, again to facilitate in-depth thinking. At the same time, we had a set of topics to
cover defined beforehand (see appendix 2). This list allowed us to make sure that we touched upon all relevant topics and served as a guideline for not letting the conversation go off track too much. The topics comprised questions that addressed the embodiment aspects of the prototype, such as if the participants think the design fits into their practice, whether they were able to correctly decode the knowledge and if that knowledge would influence their behavior in terms of their motivation. Each session took around 45 minutes and was done by two researchers; one conducted the interviews while the other took notes. All sessions were, with the consent of the participants, audio-recorded and transcribed. Finally, we performed another qualitative content analysis based on the transcripts (see section 5.1). This time, however, we focused only on meaning units that did not obviously address the looks or usability of the prototype alone without relating to Embodied Interaction.

**Results**

Overall, the application was perceived very positively. Almost all participants stressed how they felt captivated by the way their efforts were contextualized. One stated, for exampled, that he appreciated how the application provided reasons for why it works the way it does:

> “I really liked to have a motivation between all, for everything. That’s good, because sometimes that helps to keep people motivated. To see that it actually helps that you’re doing it.”

Another participant emphasized the importance of how the information and challenge are gradually built up, since she perceived that to be in accord with her mental expectation. Almost all participants were also quite explicit about how they felt to be specifically motivated by the fact that their activity was related to climate change rather than focusing only on themselves. Participants expressed this view in different ways, such as: “It angles away from the health perspective and is more about the global perspective. That is more [...] attractive to me”, “on a large scale, this is what everyone will benefit from” and “when it’s climate change I think that’s a thing that everyone should take a part of”. Especially the polar bear mascot seemed to serve as an important point of reference in this context and was the main aspect that stuck in the heads of the participants. Consequently, most participants were positive that they would actually use such an application. This was, as multiple participants said, not least due to the low interaction and tracking effort that the application needed. Further, one participant who was initially sceptic towards self-tracking could see how the application could help him to overcome this skepticism and get more involved:

> “I could probably see myself starting to use it, get more involved and then go into the more advanced part. And start getting more interested in filling out all the numbers. Progressively.”

More importantly, participants were positive that the information they got from the application would influence their motivation to be more physically active. In this context, it was also brought up by one participant that he “was aware of how we can affect the climate before, [...] but [...] maybe it could motivate [him] to actually stick to it”. Examples included one participant mentioning that she sees herself opting for walking when taking the bus.
would be the more convenient choice. Other participants could imagine altering their schedule or commuting habits to include more walking. Only one participant felt not compelled to continue using the application and did not see how it could have any impact on his behavior.

Criticism often addressed not what information was included in the design, but how it was presented. Albeit with varying emphasis, most participants expressed in one way or another how they would want the application to contain more “explicit fun”. Some just suggested to split the text in smaller, easier digestible parts that are illustrated with pictures: “If you split it up, maybe it’s easier to read.” On the other side of the spectrum, two participants advised to place a major focus on the gamification elements and turn the measurement unit of the polar bear into a virtual pet which demands to be ‘fed’ by physical activity. Another concern, which we regard as serious, was the impression of being manipulated: On the one hand, the metaphor relates the user’s physical activity to how it would help counteracting climate change. On the other hand, the relations are clearly not that simple and just walking more will not have an immediate effect. This concern, however, was raised only by one participant. Another participant in a different session had a remark that touched the same issue, although he interpreted it in a more positive way: “It would probably be very hard to measure, but it’s [...] the morale behind it, you have to try”. Apart from that, all participants already started thinking beyond the scope of this study and suggested improvements in how the finished product could fit better into their everyday life. It was frequently pointed out that a bigger user base also has a broad range of different lifestyles and -situations which the application would need to accommodate with more variation and freedom in choosing challenges.

Finally, participants repeatedly talked about how they would like to see social features. As one participant formulated it: “When it’s climate change I think that’s a thing that everyone should take a part of.” Features that were suggested included sharing personal successes, teaming up with friends and challenging others.

Discussion
During the evaluations, we only received a very small number of remarks that refer to the usability of the prototype itself. Therefore, we think that the expert evaluations helped us to remove most of the ‘superficial’ flaws of the prototype so that the participants focused on the actual subject of our research – improving motivation to engage in physical activity. The results of the evaluations indicate that the prototype achieved this goal. But more specifically, we were interested in if and how Embodied Interaction can inform design to this end. Based on the data we gathered, we suggest that it can; and in the following, we lay out some ways in which we think it does so.

Firstly, the use of a metaphor appears to have a positive effect on the willingness of the users to engage in the activity promoted by the application. This observation is in line with the related research we introduced earlier (Lin et al., 2006; Ludden et al., 2014). However, Embodied Interaction argues that all information, including metaphors, has to be made meaningful and can only be so through participation in the world with a specific aim (Dourish, 2014). This notion motivated choosing a real-world metaphor as opposed to a ‘fantasy’ setting in our design, like it was the case in the works of Lin et al. (2006) and Ludden et al (2014). Indeed, it appears to make a difference in how users arrive at a feeling of
relevance. The metaphor of climate change is not necessarily the only way of achieving this. Nevertheless, it seems to be particularly useful in this context: On the one hand, it enabled the participants to relate their efforts to something that transcends them as individuals rather than placing themselves at the center of attention. On the other hand, it allowed them to put knowledge into practice which they already had, even though they might not have consciously enacted it. The results of the evaluations indicate that this was more appealing to them than simply recording and seeing the number of their steps. This, in turn, raises questions about the purpose of self-tracking: If it is self-knowledge (Choe et al., 2014) and self-improvement (Hoy, 2016), perhaps these ends can be achieved not solely through numbers, but through giving them “meaning as part of a larger system” (Dourish, 2004, p. 183) as aimed at by Embodied Interaction. This would allow for more streamlined designs that avoid the problems of deciding which data to collect about oneself and what to do with it (Choe et al., 2014). The downside of this is the risk of being perceived as manipulative, as it was the case with one of the participants. It is therefore important to find the right balance between realism and reality.

Secondly, Campenhout et al. (2016) recommend an integration of the physical and the digital world in order to create engaging products. In our study, we did not follow their advice of designing more tangible and specialized devices, but chose a different way of integration. Our design takes on the function of a 'lens' through which users are encouraged to act in a certain way and receive relatable feedback on their behavior. The real interaction, however, happens in the physical world under the influence of the feedback. Most participants of the study found this approach of focusing on practice rather than using the application, as it is promoted by Embodied Interaction, to be very pleasing. For us, this indicates that a practice-oriented integration of the physical and the digital can have a persuasive effect on users.

Thirdly, we think that most of the criticism of the prototype can be addressed by starting at the approaches we have so far and developing them further. Hence, while we do not want to dismiss the concerns as negligible, we do not see them as contradictory to the ideas of Embodied Interaction either. The main negative feedback we received revolved around how the information was presented, i.e. encoded. The results of the evaluations indicate that, while the decoding worked well on the cognitive level, meaning did not necessarily emerge for all participants to the same degree. This is pointed out by Dourish (2014) in one of the design principles of Embodied Interaction – ultimately, meaning is created and managed by users, and merely suggested by designers. According to Dourish, it follows that the responsibility of the designer is to focus ...

“[…] on the resources that a design should provide to users in order for them to appropriate the artifact and incorporate it into their practice” (Dourish, 2014, p. 173).

In the case of ClimateQuest, this could indeed – as recommended by some participants – mean to place a greater focus on enhancing the metaphor with game elements or introducing social features. Determining suitable measures, however, would require further research with a broader perspective on Embodied Interaction.
Nevertheless, these conclusions come with some caveats. It happened that all participants of the study came, in one way or another, from a background in digital product design. We are aware that this might have had an influence on their attitude during the evaluations. However, their answers during the background interviews in the beginning of each session align with the research on problems with self-tracking (Choe et al., 2014; Walker-Rettberg, 2014). Thus, we think it is reasonable to assume that the participants face the same issues as other potential users, and we therefore regard their answers as valid. Apart from that, the scope of this study only allowed for one evaluation session with each user. Although our findings indicate a feeling of increased motivation for physical activity, it is not guaranteed that this will actually lead to a change in behavior in their everyday lives. Again, this impression would need to be confirmed by further research, but we are confident that it is at least a promising approach.

6. Conclusion

The aim of this study was to assess if and how Embodied Interaction can inform new and engaging ways of using the data we track about ourselves. To this end, we conducted expert evaluations as well as empirical user studies on a prototype that was based on ideas of embodiment. Both evaluations had the purpose of generating insight about the use of the theory of Embodied Interaction. So, what are ways in which Embodied Interaction can help us to better inform and motivate action in the context of self-tracking?

- **Support practice**: Embodied Interaction emphasizes that digital systems are means to an end. Physical – and arguably most other – activities that are commonly tracked happen in the physical world. In our study, we showed how digital applications can participate in these activities and support them through providing context, motivation and feedback. Further research is needed to find out whether the increased feeling of motivation as it was expressed by the participants would actually be enacted in their everyday lives.

- **Focus on goals rather than data**: Embodied Interaction is about connecting action to meaning. On the example of step-counting, our results suggest that making this connection in self-tracking applications can increase motivation to act upon the data we gather about ourselves. How the connection is made in the concrete case, however, entails the question of what we want to achieve by quantifying ourselves, rather than the question of what technology allows us to do.

- **Provide a less self-centered context**: Embodied Interaction sees the role of computation in being a medium to communicate knowledge. Knowledge is made meaningful by contextualizing and making it relevant to users. The results of our study indicate that the choice of context can influence the emergence of motivation. In contrast to how applications in quantified self employ the user’s data double or fictional
metaphors for this purpose, we suggest focusing on a bigger picture that transcends the user as an individual.

- **There is more to be discovered:** As we elaborated earlier, we think of the criticism on the prototype, which was expressed during the evaluations, as shortcomings of our design, but not as contradictory to Embodied Interaction. Further research with a broader perspective on embodiment could help to explore these issues.

On a related note, the heuristics that we formulated as a frame for discussing Embodied Interaction in design throughout the study turned out to be a valuable tool. They provided a common ground for discussions not only for us, but also for discussing matters of embodiment with other actors that are less familiar with the area. However, as we mentioned as well, these heuristics are rather tailored and might need to be adapted if they are to be used in different projects.

During one of the evaluations, a participant suggested what could be seen as a business-perspective of embodiment: ‘What’, he wondered, ‘if the revenues of a real-world implementation of the application would in parts be used to fund environment-related charities?’ This is an intriguing idea of how an embodied system like ours could have an impact in the world that goes beyond the metaphorical – and, if you want to see it that way, beyond the potentially manipulative. As Camenhout et al. (2016) rightly observed, it is up to researchers and designers to discover new ways of utilizing the potential of embodied technologies. Funding charities with revenues might be one – admittedly unconventional – such way. We think that our study discovered another promising path and we hope that there is a chance for it to be explored further in the future.

**Acknowledgements**

There is a number of people that supported us along the way, and we would like to take a moment to express our gratitude for their efforts. Firstly, our supervisor Andreas Lund at the Department of Informatics at Umeå University – thank you for your valuable feedback on our thoughts, process and writing. From the same department, we would like to thank our course coordinator Rikard Harr and Fatemeh Moradi for their thoughtful input in preparation of and during this work. And last but not least, to our fellow students in the Human Computer Interaction and Social Media program at Umeå University – thanks for any proofreading, discussing ideas and support you gave us during the last months!
Appendix 1: Expert Evaluation Briefing Document

1. Introduction

The following heuristics are based on the ten general principles for usability and interaction design created by Nielsen. The heuristics are developed for examining design proposals or, like in our case: a prototype, in order to identify and remove factors that might interfere when non-expert users are testing the artifact. Since our work is based on the field of Embodied Interaction as defined by Dourish, a few heuristics have been altered or added, inspired by the design principles listed in his book *Where the action is*. Embodied Interaction itself is not referring to a specific technology or some sort of guidelines, but rather a perspective of the relationship between users and systems, according to Dourish.

Finally, a few definitions are required in order to explain what we mean by certain terms. If we talk about system, we mean – in contrast to how the term is used in the original heuristics – not only the application itself, but also the context in which it is used. As depicted in the figure below, the context comprises the user and the activity that the user engages in. Meaning, then, is what emerges when a person interacts with and/or uses the system to achieve a purpose or goal.

![Diagram of System, Activity, User, Application, and Context]

2. Process

To get started, please read the rest of this document in preparation of the evaluation session. You will find a test scenario as well as a list of heuristics in the next section. After that, please go to the prototype at [http://climatequest.a-ph.net](http://climatequest.a-ph.net) (requires Chrome) and get familiar with the application.

During the evaluation session, you will go through the prototype again, while keeping both the scenario and the heuristics in the back of your head. Try to view the application through the lens of these heuristics, and please take notes on how well (or bad) the prototype covers them in the corresponding text box in this Google Form.

We encourage you to try and conduct the evaluation mostly on your own – anything you stumble upon can provide insight for us as well. However if you encounter any technical difficulties or feel too unsure about anything during the evaluation, you're welcome to ask us!
at any time. Just remember: We’re not testing your expertise, but ask for your critical feedback in order to improve our product.

Once you’re done, we will take another 10-20 minutes to go through your notes together with you, to make sure we’ve got everything right.

3. Data Usage Agreement

We value your support and your privacy. The data you provide us with will be used within this project in order to improve the prototype. Insights as well as quotes from your notes or the following discussion may be used for describing findings or changes in the prototype in the written thesis document. However please be assured that the raw data will remain available only to the members of this project and everything we publish will be fully anonymized. Furthermore, if you decide that you don’t want to continue with the evaluation or want us to remove specific parts of your notes or what you said afterwards, just tell us and we will meet your wish.

4. Test Scenario

Imagine you’re remotely familiar with how other people use their smartphones and similar devices to track fitness information such as exercise statistics and step count about themselves. While you’re not the biggest couch potato, you’re also not very regular or structured in your exercise routine, and you don’t really see how all the graphs and data sets should help you improve your motivation.

In that situation, you discover the application ClimateQuest, which connects self-tracking with challenges rooted in the real world in a playful way. In our example, we focus on being active in your everyday life rather than exercising specifically. You decide to give it a try and accept the challenge given to you by the application.

5. Heuristics

The following heuristics are criteria or rules of thumb that should be addressed by the design of the system.

System Consistency – The application fits into the practice. It is an embedded part of that practice which is transformed, extended or restricted by the application. Thus, the application serves as a tool within the practice rather than being an end in itself.

Visibility – The meaning of actions is clearly understandable, even before the action is taken by the user. The impact on the whole system is visible, e.g. through constant feedback. The responses of the application to the actions are engaging for the user.

Encoding and Decoding of Knowledge – Knowledge (e.g. information, data) is not simply displayed, but modulated. Modulation means that the raw information is changed in a way that benefits a specific understanding by the user. This could include putting it into a specific context. Users are able to understand the meaning carried by the modulation as well as the context. Thus, they can transfer the knowledge to their behavior in the real world.
**Communication** – Meaningful interpretation of the knowledge through the user is supported by the system rather than relying on the user alone. The relations between the different parts of the application itself and the system are understandable.

**Error** – The application helps to prevent, recognize, diagnose and recover from errors not only within the application, but also in the practice that is supported by the application.
# Appendix 2: User Testing Worksheet

## Information

<table>
<thead>
<tr>
<th>Participants</th>
<th>Name 1:</th>
<th>Name 2:</th>
</tr>
</thead>
</table>

## Part 1: Preamble

<table>
<thead>
<tr>
<th>What's about to happen</th>
<th>Short introduction and some questions on your background</th>
<th>Prototype testing</th>
<th>Questions on your experience using the prototype</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethics</td>
<td>Thank you!</td>
<td>You are free to cancel this testing at any time, if you wish to.</td>
<td>All data will be anonymous and it will only be processed by the people in our team.</td>
</tr>
</tbody>
</table>

| Anything you would like to mention before we start? | |

## Part 2: Pre-Interview

| Can you briefly introduce yourself? | |
| How do you get around? | |
| Please estimate how many steps you take on a normal weekday. | |
| What is your interest in self-tracking? Why? Pros and Cons? | |

## Part 3: Prototype Testing

| Scenario | Open the application on the screen | Try to play around with the app and solve the challenge | Do it together with your partner and talk |
| Notes | |

## Part 4: Interview

<p>| What is your first impression after using the app? | |
| Would you actually use it? Please explain! Why (not)? Under which conditions would you see yourself using it? | |
| How do you imagine your day-to-day use of the application? | |</p>
<table>
<thead>
<tr>
<th>How do you think would such an app change your commuting behavior?</th>
</tr>
</thead>
<tbody>
<tr>
<td>What, in your opinion, is the purpose of the application?</td>
</tr>
<tr>
<td>Please summarize what information the application gave to you, and why.</td>
</tr>
<tr>
<td>Do you expect that knowledge to have any effect on you? If so, which? Why?</td>
</tr>
<tr>
<td>Was there anything confusing?</td>
</tr>
<tr>
<td>Anything else you would like to mention?</td>
</tr>
</tbody>
</table>
Appendix 3: Coding and Code Sorting Examples

This section provides some examples to illustrate how we coded meaning units during the qualitative content analysis. We used direct quotes from the transcripts as meaning units, paraphrased them in a more concise way that also considered the context, and formed a code to easier categorize and talk about them. These were then sorted in sub-categories, categories and themes.

Coding of Meaning Units

<table>
<thead>
<tr>
<th>Meaning Unit</th>
<th>Condensed Meaning Unit</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>S: It would probably be very hard to measure, but it's like ... the morale behind it, you have to try. To change it gradually and hopefully it catches on and makes everyone walk just a bit more.</td>
<td>There is usefulness in the morale of the application, even if the actual effect is hard to measure.</td>
<td>Useful morale</td>
</tr>
<tr>
<td>S: I would say that I was aware of how we can affect the climate before, ehm ... but yeah, maybe it could motivate me to actually stick to it and not just ... F: Yeah, I would say the same as [name removed], I think.</td>
<td>They are aware of the information, but the application could help putting it to practice.</td>
<td>Put knowledge to practice</td>
</tr>
<tr>
<td>L: Like, I really like, for me ... it's a step by step, and that works from a kind of mental model. Like, okay, this is the background, and by doing this, you will solve ... this is like the reward of what you're going to to.</td>
<td>The app gradually builds up knowledge of the background, the task and the result.</td>
<td>Gradual communication of knowledge</td>
</tr>
<tr>
<td>M: Well, I mean it's ... it would be if it had a bit more like, explicit fun. It's very, oh, it's informative, bigger picture and all that, but maybe it should be a bit more fun also, the world ... just general fluff. This sort of thing. L: Mhh. [confirming] M: Put some humor in it. It's a happy polar bear.</td>
<td>The application is informative and shows context, but could present it in a more fun way.</td>
<td>More fun presentation of information</td>
</tr>
<tr>
<td>K: But I want to like, keep track like, on how much have I been saving the world? Like, am I better than... Yeah this is cool but I can't show it to anyone, I have to show my phone. &quot;Hey, CHECK THIS OUT!&quot;</td>
<td>An important part can be to not only know about your successes, but also share them with others.</td>
<td>Share success</td>
</tr>
</tbody>
</table>
### Code Sorting

<table>
<thead>
<tr>
<th>Themes</th>
<th>Embodied Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Category</strong></td>
<td>Knowledge</td>
</tr>
<tr>
<td><strong>Sub-Category</strong></td>
<td><strong>Perspective of Information</strong></td>
</tr>
<tr>
<td>Codes</td>
<td>Benefits everyone</td>
</tr>
<tr>
<td>Relevant perspective</td>
<td>Remember polar bears</td>
</tr>
<tr>
<td>Extreme issues</td>
<td>Polar bears are remembered</td>
</tr>
<tr>
<td>Local perspective</td>
<td>Polar bear mascot is nice</td>
</tr>
<tr>
<td>Encouragement through perspective</td>
<td>More fun presentation of information</td>
</tr>
<tr>
<td>Puts efforts into perspective</td>
<td>Time reference on graphs</td>
</tr>
<tr>
<td>Motivating context</td>
<td>More entertaining descriptions</td>
</tr>
<tr>
<td>Bigger effect</td>
<td></td>
</tr>
<tr>
<td>Hard to form personal connection</td>
<td>Color Coding</td>
</tr>
<tr>
<td>Relevant to everyone</td>
<td>Short and concise information</td>
</tr>
<tr>
<td>Local information is remembered</td>
<td>Boring information</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
References


List of References


