TD

Tracking in the Wild

Exploring the everyday use of physical activity trackers

DOCTORAL THESIS

Rúben Hugo de Freitas Gouveia
DOCTORATE IN INFORMATICS ENGINEERING
SPECIALTY IN HUMAN-COMPUTER INTERACTION

UNIVERSIDADE da MADEIRA
A Nossa Universidade
www.uma.pt

May | 2019

Tracking in the Wild
Exploring the everyday use of physical activity trackers

DOCTORAL THESIS

Rúben Hugo de Freitas Gouveia
DOCTORATE IN INFORMATICS ENGINEERING
SPECIALTY IN HUMAN-COMPUTER INTERACTION

SUPERVISOR
Evangelos Karapanos

CO-SUPERVISOR
Pedro Filipe Pereira Campos

Tracking in the wild: Exploring the Everyday Use of Physical Activity Trackers

Rúben Hugo de Freitas Gouveia

Submitted in partial fulfillment of the requirements for the degree of Doctor of Computer Science, with specialty in Human Computer Interaction

Universidade da Madeira

2019

Supervisors:

Evangelos Karapanos (supervisor)

Pedro Campos (co-supervisor)

This work was supported by the by the Portuguese Foundation of Science and Technology grant SFRH/BD/89796/2012. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author and do not necessarily reflect those of the funding agencies.

Abstract

As the rates of chronical diseases, such as obesity, cardiovascular disease and diabetes continue to increase, the development of tools that support people in achieving healthier habits is becoming ever more important. Personal tracking systems, such as activity trackers, have emerged as a promising class of tools to support people in managing their everyday health. However, for this promise to be fulfilled, these systems need to be well designed, not only in terms of how they implement specific behavior change techniques, but also in how they integrate into people's daily lives and address their daily needs. My dissertations provides evidence that accounting for people's daily practices and needs can help to design activity tracking systems that help people get more value from their tracking practices.

To understand how people derive value from their activity tracking practices, I have conducted two inquiries into people's daily uses of activity tracking systems. In a fist attempt, I led a 10-month study of the adoption of *Habito,* our own activity tracking mobile app. *Habito* logged not only users' physical activity, but also their interactions with the app. This data was used to acquire an estimate of the adoption rate of *Habito*, and understanding of how adoption is affected by users' 'readiness', i.e., their attitude towards behavior change. In a follow-up study, I turned to the use of video methods and direct, in-situ observations of users' interactions to understand what motivates people to engage with these tools in their everyday life, and how the surrounding environment shapes their use. These studies revealed some of the complexities of tracking, while extending some of the underlying ideas of behavior change. Among key results: (1) people's use of activity trackers was found to be predominantly impulsive, where they simultaneously reflect, learn and change their behaviors as they collect data; (2) people's use of trackers is deeply entangled with their daily routines and practices, and; (3) people use of trackers often is not in line with the traditional vision of these tools as mediators of change – trackers are also commonly used to simply learn about behaviors and engage in moments of self-discovery.

Examining how to design activity tracking interfaces that best support people's different needs , my dissertation further describes an inquiry into the design space of behavioral feedback interfaces. Through a iterative process of synthesis and analysis of research on activity tracking, I devise six design qualities for creating feedback that supports people in their interactions with physical activity data. Through the development and field deployment of four concepts in a field study, I show the potential of these displays for highlighting opportunities for action and learning.

Keywords: Physical activity tracking; Personal informatics; Behavior change technologies; Human computer interaction.

Resumo

À medida que a prevalência de doenças crónicas como a obesidade, doenças cardiovasculares e diabetes continua a aumentar, o desenvolvimento de ferramentas que suportam pessoas a atingir mudanças de comportamento tem-se tornado essencial. Ferramentas de monitorização de comportamentos, tais como monitores de atividade física, têm surgido com a promessa de encorajar um dia a dia mais saudável. Contudo, para que essa promessa seja cumprida, torna-se essencial que estas ferramentas sejam bem concebidas, não só na forma como implementam determinadas estratégias de mudança de comportamento, mas também na forma como são integradas no dia-a-dia das pessoas. A minha dissertação demonstra a importância de considerar as necessidades e práticas diárias dos utilizadores destas ferramentas, de forma a ajudá-las a tirar melhor proveito da sua monitorização de atividade física.

De modo a entender como é que os utilizadores destas ferramentas derivam valor das suas práticas de monitorização, a minha dissertação começa por explorar as práticas diárias associadas ao uso de monitores de atividade física. A minha dissertação contribui com duas investigações ao uso diário destas ferramentas. Primeiro, é apresentada uma investigação da adoção de *Habito*, uma aplicação para monitorização de atividade física. *Habito* não só registou as instâncias de atividade física dos seus utilizadores, mas também as suas interações com a própria aplicação. Estes dados foram utilizados para adquirir uma taxa de adopção de *Habito* e entender como é que essa adopção é afetada pela "prontidão" dos utilizadores, i.e., a sua atitude em relação à mudança de comportamento. Num segundo estudo, recorrendo a métodos de vídeo e observações diretas e *in-situ* da utilização de monitores de atividade física, explorei as motivações associadas ao uso diário destas ferramentas. Estes estudos expandiram algumas das ideias subjacentes ao uso das ferramentas para mudanças de comportamento. Entre resultados principais: (1) o uso de monitores de atividade física é predominantemente impulsivo, onde pessoas refletem, aprendem e alteram os seus comportamentos à medida que recolhem dados sobe estes mesmos comportamentos; (2) o uso de monitores de atividade física está profundamente interligado com as rotinas e práticas dos seus utilizadores, e; (3) o uso de monitores de atividade física nem sempre está ligado a mudanças de comportamento – estas ferramentas também são utilizadas para divertimento e aprendizagem.

A minha dissertação contribui ainda com uma exploração do design de interfaces para a monitorização de atividade física. Através de um processo iterativo de síntese e análise de literatura, seis qualidades para a criação de interfaces são derivadas. Através de um estudo de campo, a minha dissertação demonstro o potencial dessas interfaces para ajudar pessoas a aprender e gerir a sua saúde diária.

Palavras-Chave: Monitorização de atividade física; Tecnologia pessoal; Tecnologias para mudança de comportamento; Interação Humano Computador.

Statement of Attribution

This doctoral dissertation is based on material from the following conference publications:

a. **Rúben Gouveia**, Evangelos Karapanos, and Marc Hassenzahl. (2015). How do we engage with activity trackers?: a longitudinal study of Habito. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing* (UbiComp'15), pp. 1305-1316.

b. **Rúben Gouveia**, Fábio Pereira, Evangelos Karapanos, Sean A. Munson, and Marc Hassenzahl. (2016). Exploring the design space of glanceable feedback for physical activity trackers. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing* (UbiComp'16), pp. 144-155.

c. **Rúben Gouveia**, Evangelos Karapanos, and Marc Hassenzahl. (2018). Activity Tracking *in vivo*. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*, pp. 362–13.

All articles are reprinted with permission from Association for Computing Machinery, Inc.

Besides the publications included in this dissertation, my PhD contributed to a number of additional conference (d,e), journal (f) and workshop (g, h, i, j) publications. I further co-organized two workshops on personal informatics/quantified-self technologies (k, l) during my time as a PhD student:

d. **Rúben Gouveia** and Evangelos Karapanos. (2013). Footprint tracker: supporting diary studies with lifelogging. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '13). pp. 2921-2930.

e. Vítor Belim, Olga Lyra, Pedro Teixeira, Ana Caraban, Maria José Ferreira, **Rúben Gouveia**, Andrés Lucero, and Evangelos Karapanos. (2014). Beyond gamification: sociometric technologies that encourage reflection before behavior change. In *Proceedings of the 11th Conference on Advances in Computer Entertainment Technology* (ACE '14).pp. 27:1 – 27:6.

f. Evangelos Karapanos, **Rúben Gouveia**, Marc Hassenzahl, and Jodi Forlizzi. (2016). Wellbeing in the making: peoples' experiences with wearable activity trackers. *Psychology of well-being*, *6*(1), pp. 1-17.

g. Davide Neves, Donovan Costa, Marcio Oliveira, Ruben Jardim, **Ruben Gouveia**, and Evangelos Karapanos. 2016. Motivating Healthy Water Intake through Prompting, Historical Information, and Implicit Feedback. *arXiv:1603.01367*.

h. **Rúben Gouveia,** Fábio Pereira, Ana Caraban, Sean A. Munson, and Evangelos Karapanos. 2015. You have 5 seconds: designing glanceable feedback for physical activity trackers. In *Adjunct Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing* (UbiComp'15), pp. 643-647.

i. Tiago Ornelas, Ana Caraban, **Rúben Gouveia**, and Evangelos Karapanos. 2015. CrowdWalk: leveraging the wisdom of the crowd to inspire walking activities. In *Adjunct Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing* (UbiComp'15), pp. 213-216.

j. Ana Caraban, Maria José Ferreira, **Rúben Gouveia**, and Evangelos Karapanos. 2015. Social toothbrush: fostering family nudging around tooth brushing habits. In In *Adjunct Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing* (UbiComp'15), pp. 649-653.

k. Amon Rapp, Federica Cena, Judy Kay, Bob Kummerfeld, Frank Hopfgartner, Till Plumbaum, Jakob Eg Larsen, Daniel A. Epstein, and **Rúben Gouveia**. (2017). New frontiers of quantified self 3: Exploring understudied categories of users. In *Adjunct Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing:* (Ubicomp'17), pp. 861-864.

l. Amon Rapp, Federica Cena, Judy Kay, Bob Kummerfeld, Frank Hopfgartner, Till Plumbaum, Jakob Eg Larsen, Daniel A. Epstein, and **Rúben Gouveia**. (2016). New frontiers of quantified self 3: Going Beyond Numbers. In *Adjunct Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Ubicomp'16), pp. 506-509.

Acknowledgments

First and foremost, I am grateful to **Evangelos Karapanos** for always challenging me to be more critical, to ask better questions and for always being present. I have learnt a great deal of how to conduct research from you.

I am fortunate to have collaborated with a number of amazing researchers. I would like to thank **Marc Hassenzahl** for the valuable feedback and discussions. Thank you for your refreshing frontality. I would also thank **Sean Munson** for the valuable discussions and opportunity to spend the summer of 2015 in the department of HCDE. Your research has inspired my own, and set a standard for what I consider to be great research. I thank **Ana Caraban** for the endless discussions about the good, the bad and ugly about research.

During my time as a PhD student, I also had the opportunity to co-supervise a group of amazing students. Thank you **Donovan Costa**, **José Rodrigues**, **Marco Leão**, **Pedro Teixeira**, **Tiago Ornelas**, **Vitor Belim** and **Vitor Teixeira.** A special thank you to **Fábio Pereira** and **Sérgio Barros** for your dedication and enthusiasm.

I would also like to express a special thank you to the amazing staff at **M-ITI**. Thank you for all the support, and for fostering such an inspiring place to conduct research.

Beyond the amazing support from all my collaborators and colleagues in the acadmic world, I would also like to thank some important friends outside of the academic world.

To my parents **Renato Gouveia** and **Sandra Gouveia** and my **siblings Alexandre Gouveia** and **Jessica Gouveia**, my deepest gratitude. Thank you for your encouragement, for your support and for your love. To my dear **Jennifer Linkenbach,** thank you for all the **encouragement**, **understanding** and **love**. Thank you for dragging me out on the long days, and for showing me the importance of having a balanced life. I look forward to what the future has in hold for us.

The Beatles said it best: “I get by with a little help from my friends”. To all the friends which have helped me during this path: Thank you.

Table of Contents

Chapter 1.

Introduction

Globally, more than one fourth of adults over 15 years of age are insufficiently physically active [36] . This puts more than 1.4 billion adults at risk of developing chronic diseases related to sedentary lifestyles such as obesity, cardiovascular diseases and diabetes [87]. These conditions are a major threat to today's healthcare systems. In the United States, they account for over two-thirds of healthcare expenditures, and recent reports have predicted this number to rise both in the United States and also in other parts of the world [39]. Policy makers and health leaders have stressed the importance of leading an active lifestyle [37],[78]. However, many of us know how difficult it may be to lead active lifestyles. Even those of us who do come around with a plan to increase our levels of physical activity (e.g. by setting physical activity goals), typically have a hard time fulfilling them. Too many of us end up relapsing into our old habits after several weeks or months of committing to changing our behaviors [17].

Behavior change tools, such as activity trackers, hold great promise for transitioning to a new healthcare landscape that provides individuals' with effective support for managing their everyday health [68]. Numerous research projects have reported the potential health benefits gained from the use of these tools - individuals have been found to walk more [11], lose weight [4],[9] and feel more in control of their behaviors [17] while consistently monitoring their step count with an activity tracking device. These benefits have been found to prevail over months or years of continued monitoring of physical activity [82].

Much of the research on activity tracking has focused on investigating the effectiveness of these systems towards influencing human behavior [44]. Commonly, these studies attempt to answer questions such as: did the use of an activity tracking prototype lead to an increase in users' step count? Were users successfully persuaded towards healthier lifestyles? Often lost in this framing are the actual experiences and practices arising from owning and using these tools in people's everyday lives.

Commonly overlooked, are questions such as: how are activity tracking devices integrated into people's everyday practices? How do they influence people's behaviors while being used in their daily lives?

More than focusing on the effectiveness of these devices, these questions call for investigations on how users appropriate activity trackers in their daily lives, and how adoption, or non-adoption, is shaped by the context of use.

Investigating these points is an important step towards designing tools that best address people's needs. Despite the potential benefits of tracking, people often stop using their activity trackers within early weeks or months of purchase [14],[50]. For instance, Shih and colleagues [77] found that over 50% of activity tracker owners abandon the use of their devices within two weeks of purchase. Similarly, a 2016 survey from Gartner Market Research [90] suggests that over a third of owners of commercially available activity trackers discard them within three months of purchase. While for some, the profound disengagement with trackers may indicate success - namely, when activity trackers instill new practices to the point they are no longer required to motivate physical activity; activity trackers are actually frequently abandoned because they simply fail to address people's needs. Activity trackers have been found to be discarded due to a lack of interest in the level of information that they provide, boredom, and not fitting users' conceptions of themselves [50],[90].

In my dissertation, I contribute to an exploration of how activity trackers can be better designed towards addressing people's everyday uses and needs. My dissertation draws on the *lived informatics* perspective of activity tracking [76], which considers that people's use of self-tracking devices is embedded in their everyday lives and that current tools should be designed, and evaluated, with this reality in mind. This calls for more complex models of activity tracking than the widely accepted *reflective* model [51], which assumes that people collect, then carefully explore and review their data in retrospect to identify patterns in their behaviors and plan future courses of action. I argue instead, that tracking is predominantly *impulsive*, where users simultaneously reflect, learn and take action upon their behaviors as they unfold within their everyday lives [24],[33],[76].

I argue that today's commercially available activity trackers are primarily tailored to support *reflective* modes of tracking through, for instance, the high-level aggregation of data over time (e.g. step count over a week or a month), and that considerably less effort has been placed in designing activity trackers that identify, and support opportunities for learning and action, as they arise in people's everyday lives.

1.1. Research Questions

Within my dissertation, I have explored three particular aspects of the everyday use of activity trackers, which I adopt as the underlying research questions of my work:

RQ1: How do people interact with physical activity trackers in their daily lives? Prior work has found users to abandon their trackers within initial days of use and systematically explored the reasons leading to these abandonments. However, little is known as to how trackers are used by those which manage to successfully adopt these tools into their daily lives. To this end, my dissertation investigates the long-term adoption of activity trackers. Insights into the long-term, daily use of trackers can shed light into the effectiveness of different design strategies and the reasons that underlie any success or failure. They can also uncover discrepancies between expected and the actual adoption of these tools [33].

RQ2: How is the use of activity tracking devices integrated into the fabric of people's daily lives? Recent HCI research has called for wider discourses on the social relationships, practices, places and spaces under which self-tracking data is collected, and interacted with. As suggested by Attfield et al. [3], technology use is not just about how interactions unfold, but also about the motivations that lead people to use technology and the practices they have with these tools in their daily lives. Through the use of video methods, my dissertation looks at a number of practices that surround the use of activity trackers within everyday life, and investigates the interplay between the use of these devices and the surrounding environment.

RQ3: How can we design physical activity feedback that help people learn and take action upon their behaviors? Activity tracking literature has long assumed that people change their behaviors as the result of deep and careful reflections on their behaviors

[51]. As such, considerable amount of efforts have focused on the design of personal visualizations to support reflection, interpretation and reminiscing (e.g. [25]). In my dissertation, I point towards an alternative way of using these tools. One in which learning and action are the result of quick moments of interaction with activity trackers. I explore how to design activity tracking devices that support this type of interaction, and the effects of this form of feedback on users' behaviors.

1.2. Thesis Outline

This thesis presents the findings from three studies that investigate real life practices emerging from owning and using activity tracking devices – i.e., how these devices fit into the everyday life and routines of people that use them, and examine how these devices can be designed to be enmeshed in the people's everyday lives and outlook on their future.

Chapter 2 reviews recent literature on activity tracking. I start by introducing early models on personal informatics systems for physical activity. I describe how activity tracking has been considered a stage-based process, in which people take action on their behaviors as the result of careful planning and reflection. I then highlight a recent shift that has been placed towards a *lived informatics* perspective of tracking, characterized by the integration of tracking into everyday life by people with varying goals.

Chapters 3 and 4 present two inquiries into the everyday use of activity tracking devices. ***Chapter 3*** reports on a 10-month, in-the-wild study of the adoption, usage and discontinuation of a mobile activity tracker called *Habito*, by a sample of 256 users who installed the tracker on their own volition. *Habito* was specially designed and built to study how users interact with activity trackers. *Habito* logged not only users' physical activity, but also their interactions with the app. This data was used to acquire an estimate of the adoption rate of *Habito*, and understanding of how adoption is affected by users' 'readiness', i.e., their attitude towards behavior change. A closer look was also taken at the frequency, duration, and nature of users' interaction with *Habito* itself, the way interaction changes over time, and its impact on physical activity level. ***Chapter 4*** reports on a study where I resort to the use of video methods and direct, in-situ observations of users' interactions, to inquire into the long-term adoption of and experiences with activity trackers, and how usage is shaped by the surrounding context of use. 12 individuals were provided with wearable cameras and monitored two days in their lives, collecting in total, 244 incidents where trackers were used. These recordings were combined with behavioral data from trackers as well as interviews, providing a detailed view on a number of practices that surround tracker use in daily life.

Chapter 5 pivots to designing to help people find quick moments of self-regulation and learning as they arise within their everyday lives. I present an exploration into the design space of *glanceable feedback* for physical activity – i.e., feedback that is presented in an abstract and easily consumable form, at locations where individuals are likely to gaze frequently (e.g. background of a smartwatch), which resulted in 21 concepts and six design qualities for glanceable feedback interfaces. Four of these concepts were developed and deployed in a field study, providing an overview of how different types of glanceable feedback for activity tracking motivate, and help identify opportunities for physical activity.

Chapter 6 discusses the overall contribution of the work described in this thesis

Chapter 2.

Related Work

Over recent years, the *quantified-self* movement has become increasingly popular [79]. The main premise of this movement is to collect personally relevant information for the purpose of self-knowledge and self-improvement [59]. Among popular examples, people have been found to collect data from their physical activity, food, and finances across diverse goals that include improved health (e.g. curing or managing a pre-existing condition), mindfulness, or simply to satisfy curiosity and have fun [12].

2.1 - Personal Informatics Systems

Personal informatics, also commonly known as Quantified-Self, are a set of systems that enable people to “collect personally relevant information for the purpose of self-reflection and gaining self-knowledge” [51](p. 558). These devices are increasingly prevalent in people’s daily lives. The Quantified-Self website has identified several hundreds of health tracking applications, and many other for tracking personal data such as finances, and mood [91].

Most of these devices are built on top of three common functions: First, they collect data from users’ behaviors (e.g., step count, heart rate). This data is then used to provide feedback, and help users set and track progress towards goals. In accordance, most of the research on these tools has focused on how behavioral data can be collected (e.g. what sensors should be used and how accurate they are; where the data will be stored), and how this data will be presented to users.

In a first attempt to understand and conceptualize the use of these devices, Li et al proposed a five stage-based model of personal informatics systems [51] (see Fig. 1). Informed by Prochaska & Velicer’s Transtheoretical Model of Health Behavior Change [72], the use of these systems was described as a process consisting of five stages: preparation, collection, integration, reflection and action. In the preparation stage, people decide what they will be tracking, and how. This is followed by the collection of data (sometimes from multiple tools), and preparation of this data for reflection (e.g. through graphs and figures). This act of observing and recording one's behavior is commonly called *self-monitoring*. Previous research has suggested that self-monitoring is effective in changing behaviors as it heightens the chances of identifying undesirable behaviors [21], as well as taking advantage of opportunities for goal-directed actions [85]. Individuals have been found to walk more [11], loose more weight [2] or improve their eating habits [9] while consistently monitoring their behaviors, leading researchers to suggest the need of *"obsessive-compulsive self-monitoring"* for successful regulations in behaviors [4].

The reflection stage is where users retrospect on their behaviors, to identify patterns and plan future courses of action. This mode of use of PI systems was long assumed to be the dominant mode of interaction: Knowledge of existing behavioral patterns occurred as the result of careful planning and reflection on behaviors. As such, most of today's personal informatics tools support this process through the high-level aggregation of data over time (e.g. step count over a week or a month) and considerable amount of efforts have focused on the design of personal visualizations to support reflection, interpretation and reminiscing (e.g. [25]).

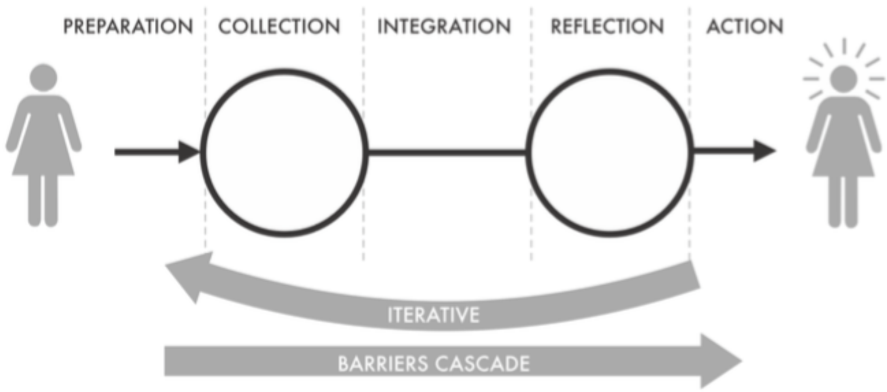

Figure 1 – Li's five-stage model of personal informatics systems

2.1.1 – Personal Informatics for Behavior Change

Much of the research around personal informatics focuses on self-improvement, persuasion and behavior change. Personal informatics devices are commonly designed with a range of behavior change techniques (e.g., self-monitoring, goal-setting, social influence), and evaluated regarding their ability to shape, and promote changes in behaviors [65].

To explain how personal informatics can potentially alter behavior, I present an overview of two widely acknowledged behavioral models and explain some of the connections of these models to PI tools: Fogg's behavior model [28], and the transtheoretical model of behavior change [72] .

Fogg's Behavior Model

BJ Fogg's behavior model shows that three elements must coincide for behaviors to occur: motivation, ability, and triggers. People need to have a certain level of motivation towards a behavior, feel they have the ability to perform it, and be prompted to perform it. The Apple watch, for instance, attempts to engage its users in periodic physical activity by prompting users to engage in short moments of physical activity (e.g., standing up) for every 50 minutes of sedentary behaviors (see Fig 2). Part of the success of these reminders (as found in [26]) can be described by Fogg's model: users are notified to engage in physical activity (i.e., behavioral trigger), and the nature of this activity is low demanding in nature (i.e. users are likely to believe they are able to achieve it, even with low motivation).

Figure 2 - The apple watch attempts to promote physical activity by prompting users to engage in "low demanding " instances of physical activity throughout their day

Further, according to Fogg, people's motivations towards a behavior are continuously changing between high and low peaks. When high, opportunities to do hard behaviors arise; when low, only the simplest behaviors are achieved. Some personal informatics researchers have, accordingly, suggested that feedback should be contextualized according to people's moods, or current motivational state [52]. Users could be prompted to take short walks on an uninspiring day, and pushed towards higher performances when more motivated.

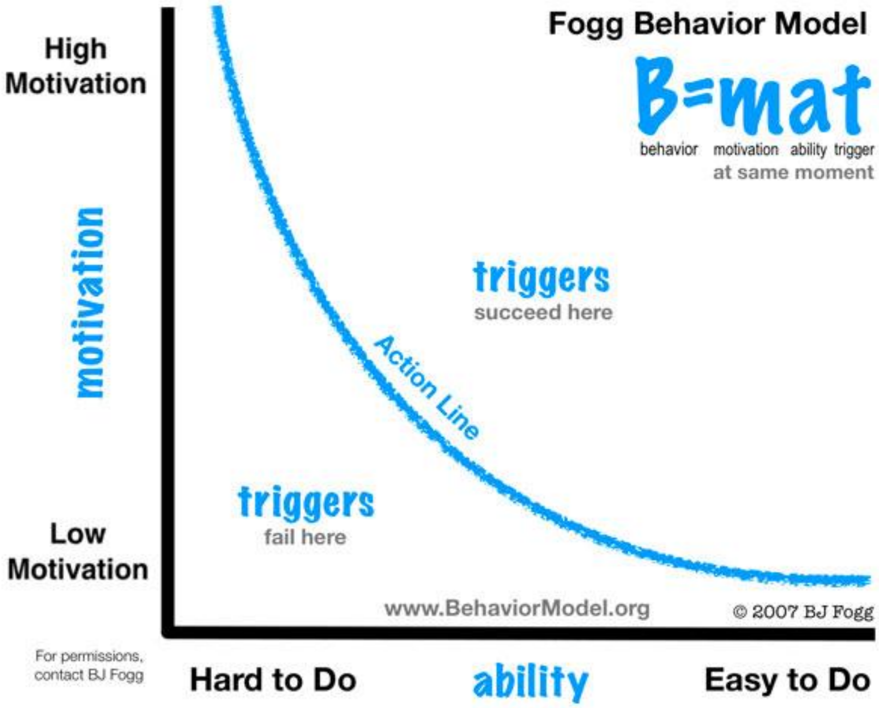

Figure 3 - Fogg's Behavior Model

Transtheoretical Model of Behavior Change

Prochaska & DiClemente's Transtheoretical Model of Behavior Change (or TTM) [72] describes five stages that people go through when intentionally trying to change a problematic behavior, such as smoking and obesity.

- *precontemplation* is the stage in which there is no intention to take change a behavior in a near future. Individuals in this stage are often unaware of the extend of a problem (e.g. being sedentary), and thus, unwilling to change their behavior;
- *contemplation* is the stage in which problems are acknowledged and people are thinking of changing their behaviors, yet have no concrete plan as to how to do so;
- *preparation* is the stage in which people have a plan of how to change their behavior (e.g. purchase a gym membership to help become more active), and plan to do so in a near future.

- *action* is the stage in which people "*modify their behavior, experiences or environment in order to overcome their problems*" (p.4, [71]). People in this stage have taken recent, visible actions upon their behaviors, but not always enough to completely abstain from it (e.g., smoking a reduced number of cigarettes);
- *maintenance* is the stage in which people have successfully modified their behavior for a period of at least six months, and are trying to avoid relapses to any of the previous stages of behavior change.

People progress through these stages cyclically, relapsing and jumping between stages before a behavior is discontinued. Within personal informatics research, the TTM has been used to evaluate a device's ability to promote behavior change (e.g., by seeing how a user is progressing in the stages [53]), and to target specific sets of users with already appropriate level of ´readiness´ for change (e.g., studying the needs of participants in a specific stage of behavior change, as in [53]). More recent research has highlighted the importance of tailoring the feedback, and behavior change strategies provided by PI tools, to the different stages of behavior change [41]. For example, users in intermediate stages of behavior change (i.e.,contemplation and preparation) - where individuals have the intention but not the strategies to change behaviors, could be presented with different design strategies to those in initial phases of behavior change – which are often unaware, and unwilling to change their behaviors.

2.2 – Research in Activity Tracking

Interest in personal informatics technologies for encouraging physical activity has increased over the last decade. Millions of people own a phone with preinstalled mobile application for health and activity tracking (e.g. Apple's HealthKit, and Google's *Google Fit*), and the market of wearable activity tracking (e.g. Fitbit and Jawbone activity trackers) has reached billions in volume over recent years [92]. In this section, I discuss two lines of research for inquiring into the use of these devices: activity tracking in the wild, and in everyday life.

2.1.1 – Activity Tracking in the wild

Much of the early research on activity tracking focused on understanding the impact of these devices for motivating, and persuading, physical activity. Tudor-Locke and colleagues [83], for instance, found individuals to walk more and loose more weight while simply tracking their steps/day with a pedometer. Bravata et al. [8] found these effects to be increased when combining simple pedometers with a daily step goal. Both studies suggested that simply recording the number of steps that one takes can reinforce physical activity behaviors. Users were able to assess their progress, record cases of success and understand how small variations in daily routines (e.g., non-scheduled sports and exercise, commuting behaviors) impacted their daily activity levels [84].

More recently, HCI researchers have produced more sophisticated activity tracking prototypes, leveraging on novel visualizations for displaying physical activity levels and behavior change techniques – such as goal completion or social influence [20] [53] [62]. Driven by a theoretical concern, the goal of these studies was to assess the efficacy and user acceptance of the prototypes and their underlying design strategies through in the wild field studies. Consolvo and collegues, for instance, explored how technology can be designed to motivate physical activity through glanceable, stylized displays. They designed *UbiFit* [20], a mobile application that provides feedback on one's physical activity through a glanceable display on the background of a mobile phone. *Ubifit* followed the metaphor of a garden, blossoming as users increase exercise levels – butterflies appear when goals are achieved and different flowers grow as a consequence of different types of exercise performed. In a 3-week comparative study with 12 participants (all with a pre-existing motivation to increase their daily levels of physically active), participants using *UbiFit Garden* had higher activity levels than those without the persistent feedback. The always-available information about their activity levels acted as a reminder to stay engaged and committed to the goal of increasing physical activity.

Researchers have resorted to field studies to explore the impact of different design strategies to motivate physical activity – from the use of novel textual messages [6], to stylized representations [49] and multi-location feedback [6]. Lin et al. [53], for instance, created *Fish'n'Steps*, a social display designed to motivate individuals to stay physically active while at work. *Fish'n'Steps* linked an individual's step count to the emotional state, growth, and activity of a virtual fish in a shared virtual fish tank. A fourteen-week study revealed increased user interest and intention to exercise when presented with appealing designs of physical activity. Tollmar et al. [6] built *Health Mashups*, a mobile application aimed at promoting reflection on one's data by identifying significant correlations between several tracked metrics - weight, sleep, step count, calendar data, location, weather, pain, food intake, and mood. If a correlation was found, the user was presented with textual messages explaining it (e.g., you walk more on days when you get more sleep). Rather than "boring users to death with numbers and graphs" (p. 48), textual feedback was found to tell users a story, and help them in drawing immediate attention towards important information and instigate action.

In the wild studies were, and are still, instrumental to the development of the field of activity tracking. They moved evaluation out of the lab, enabling the evaluation of novel technological systems in natural settings. Through quasi-experimental setups, and mixed-method approaches, they showed that decisions such as the type of feedback (e.g., stylized representations vs. narrative information), or the location in which it is presented (e.g., presented within applications vs. frequently accessed locations, such a phone's background) substantially impact user experience, and in turn, ability to motivate physical activity [17].

However, they are limited in a number of ways. First, much of this early work focuses on how activity tracking systems can support people in improving themselves. Second, participants are typically given a prototype, rather than purchasing a product on their own, and are incentivized to use the prototype for the duration of the study, which reduces the ecological validity of the study. Third, they feature a specific set of users with an already appropriate level of 'readiness' for change [53]. As a result, while these studies are extremely useful as efficacy evaluation of different design strategies [44], they do not necessarily represent real-life use.

2.1.2 – Activity Tracking in everyday life

With the widespread adoption of commercial activity trackers, researchers have increasingly shifted attention to the study of peoples' real-life practices emerging from owning and using an activity tracker. The focus of respective studies is not on the effectiveness and the efficiency of these tools, but on how users appropriate them in their daily lives, and how adoption, or non-adoption, is shaped by the context of use.

Lived perspective of activity tracking

Rooksby et al.'s [76] notion of personal tracking as lived informatics is commonly cited as a representation of the importance of considering how tracking unfolds within people's every lives. A key notion of their research is that tracking is enmeshed in the everyday life, and strongly connected to the activities and experiences that occur alongside tracker use. Tracking often involves other people (e.g. friends, family), and activities (e.g. work, jogging) [81] and is connected to a high level of emotionality, deriving from people's lives, worries, hopes and interests.

Drawing inspiration from this notion, Epstein et al. [24] recently argued that the use of personal tracking systems does not adhere strictly to the division of stages described in Li's stage-based model of personal tracking (see Chapter 2, subchapter 2.1). They proposed an alternative model for using personal tracking devices, based on the perspective of lived informatics. Tracking, under this model, is a messy reality in which people collect, integrate and reflect on their data simultaneously, within their everyday lives.

Inquiries into the everyday use of activity trackers

Rooksby et al.'s [76] and more recently, Epstein et al.'s [24] studies, have highlighted the need to explore the experiences associated to the everyday use of trackers. Researchers, in line with this call, have produced models of how people use self-tracking tools informed by qualitative inquiry [24] [51]; explored the ability of trackers to align with users' motivations and desires [38] [42]; and investigated tracking practices in everyday life [12] [29] and in specific contexts, such as the workplace [13].

Most of these studies have been qualitative in nature, relying on users' self-reports, either through interviews or online surveys. For instance, Fritz et al. [29] interviewed

30 participants who purchased a tracker of their own volition and used it from 3 to 54 months. Their study revealed how the practices that surround long-term tracking were different from those of early adoption, and how non- beneficial practices were afforded by the design of the technology, such as “number fishing” which turned exercise from a meaningful intrinsic endeavor into an extrinsically- rewarded activity. Rooksby et al. [76] interviewed 22 people two times, separated by a month, and found five different motives for tracker use, from directive, to documentary and diagnostic tracking.

Others have gathered qualitative insights by analyzing reports of experiences with trackers. For instance, Clawson et al. [14] analyzed 1600 advertisements of personal health tracking technologies on craigslist and found that individuals often abandon these, not due to technologies’ failure, but often because they achieved their goal (e.g., lost weight), they desired an upgrade to a newer model, or because of unanticipated changes in their life (e.g. surgery). Choe et al. [12] analyzed video-recordings of Quantified Self talks and found that individuals often use tracking tools for multiple reasons beyond self-improvement - from finding new life experiences (e.g. to learn something interesting) to simply having fun.

Due to their inexpensive format, qualitative studies based on interviews, surveys and reports of experiences, have inquired into the long-term adoption of and experiences with activity trackers. However, these studies are limited as well. Asking people to retrospect on typical use patterns, or experiences, may suffer from recall bias [30]. Insights on actual behaviors and experiences are likely to be forgotten, overlooked, or avoided.

My dissertation has explored the everyday use of trackers as users engage with their devices and data in their everyday lives. In the next sections, I describe how the use of interaction logs, and direct in-situ observations of users were used to depict realistic pictures of how people actually engage with trackers, and challenge some of the underlying assumptions of how these systems are used to motivate behavior change.

Chapter 3.

How Do We Engage With Activity Trackers? A Longitudinal Study of Habito

3.1 – Introduction

With systems for health and wellbeing – such as activity trackers, playing a pivotal role in health care reform, understanding the *effectiveness* of these systems towards influencing human behavior has never been more important. *Effectiveness*, in the context of HCI research in health and wellbeing, has been commonly defined as the ability of a system to bring about a change in a target behavior [44] – such as an increase in one's levels of physical activity.

Randomized control trials (RCTs) have been considered as the "gold standard" for evaluating the efficacy of health interventions [16]. Originally introduced in medical research over 50 years ago [89], the main purpose of RCTs was to improve the evidence of effectiveness of therapeutic agents. RCTs are characterized by the random assignment of subjects to one of two groups: an experimental group - in which subjects receive an intervention that is being tested, and a control or baseline group which is used for comparison or reference. Classic examples include testing the efficacy of a novel drug, as compared to a placebo, or standard drug, for reducing diastolic blood pressure [15] or treating Alzheimer's disease [48].

RCTs have, and continue to be, an imperative, widely used method for assessing the effectiveness of behavior change systems (BCSs) [7], allowing researchers to produce observable measurements of cause-effect relations between technology-mediated interventions and behaviors. An analytic review conducted by Joseph et al [7] found 64% of internet-based physical activity interventions to use randomized control studies. RCTs have been used for evaluating a variety of technology mediated strategies for behavior change, such as social support [55], self-monitoring [43] and prompts [86]. However, on their own, RCTs provide limited insights into *how*, and *why* the technology under evaluation did, or did not change people's behaviors. Often unanswered are questions such as: how was the system used by participants? Which components of a system impacted people's behaviors (and how)? How did this impact vary across participants?

For example, Poirier and colleagues [70] used a randomized control trial to examine the effectiveness of a system for encouraging physical activity that combined a clip-on activity tracker (Pebble+ Fitlinxx Inc [93]) and *Walkadoo* [94], a web-program that incorporated several strategies for promoting behavior change, such as goal setting, virtual rewards, social support and customized messages. The study examined *Walkadoo*'s effectiveness in increasing daily steps, in comparison with a control intervention. All participants (N=265) took part in a one-week run-in, during which they used the activity tracker to establish baseline measurements of their step count. During this week, participants did not receive any feedback on their physical activity. After this period, participants were randomly assigned to a control (N= 132) and intervention (N= 133) group for six weeks. Participants in the intervention group were asked to use *Walkadoo* and keep using their tracker, which displayed their progress towards a daily goal through 10 LED's – each lighting up as 10% of a goal was achieved. Participants in the control group did not receive any feedback on their physical activity. Participants' step count and frequency of access to *Walkadoo* were logged during the study and used as primary outcome measures.

Difference between both groups was striking, with the intervention group showing an increase of 845 steps/day over the control group. Participants in this group were more likely to achieve an increase of 1000 steps/day, and spend less time in sedentary behaviors as compared to the control group. These results were sustained over the 6 weeks of study.

The study of Poirier et al. was successful in many ways. First, it provided observable differences between both conditions by gathering objective measurements of people's behaviors. Second, it showed that these differences were due to the particularities of each condition – namely, of providing participants with a range of behavioral feedback on a website and activity tracker, as compared to having no feedback on their physical activity.

However, their study also revealed some limitations. For example, it did not investigate *how*, and *why* the individual components of their system impacted people's behaviors. No focus was placed on uncovering the isolated effects of *Walkadoo's* virtual rewards, goal setting, social support and customized messages, as well as the feedback conveyed by their activity trackers, towards motivating physical activity. As such, the effectiveness of each of these individual components, the way In which they interact, and contribution to the highlighted effects, is unknown.

The above scenario exemplifies a common pitfall of many evaluations of behavior change systems within HCI literature. While providing assessments of the effects of these systems as a whole, they commonly provide limited insights towards how and why specific components influence behaviors – which are of essential understanding towards informing the design of systems that effectively support people in achieving healthier lifestyles [45]. HCI researchers have, thus, recently called for evaluations that provide a deeper understanding of how these tools impact people's behaviors. As emphasized by Klasjna and collegues: "*we need answers to questions such as when people choose to use or not use a system (…) what aspects of the system they find most helpful or frustrating and why (…) Answers to these kinds of questions can help us design technology that fits into people's lives and that is likely to be effective for helping them change their habits"* ([44], p. 8).

I suggest that many of these questions may be addressed by adopting an experimental framing on users' interactions with activity trackers. I suggest that users' interactions are important mediators towards understanding behavior change. First, the most commonly employed strategy for behavior change, self-monitoring, requires interaction with feedback. So, while we may design features to provide self-relevant feedback, it can only impact behavior if people interact with these features [69]. Second, the effects of different strategies for behavior change are likely to become more apparent (and confirmed and tested) through the examination of the moments in which users engage with them. For instance, prior work has suggested that activity trackers both serve reflective and persuasive goals [61], however, little is known as to how and when users engage with their trackers to be persuaded and engage in reflections – and how this impacts people's subsequent behaviors.

In this chapter, I describe a study that leverages on data from users' usage sessions with an activity tracker, towards deriving insights on how these tools are used in people's daily lives, and how they impact behaviors. I report on a 10-month field study of *Habito*, our own mobile activity tracker, by a sample of 256 users. Designing our own application had a number of benefits. First, it allowed us to test different approaches to activity tracking through manipulating the type of feedback given. To this end, *Habito* incorporated three strategies for behavior change: *goal setting*, *textual feedback that keeps updating* and *contextualizing physical activity through location.* Second, it enabled us to log a range of behavioral data, such as the distance of walking activities and their start and end time, as well as usage data, such as the duration and time a usage session took place – and use this data to derive insights of how Habito's individual components of behavior change impacted users behaviors.

3.2 – How Do We Engage With Activity Trackers? A Longitudinal Study of Habito

This article is organized in four main sections. The first section describes *Habito* and its three design strategies for behavior change: *goal setting*, *contextualizing physical activity* and *textual feedback that keeps updating*. The second section describes the study we conducted to understand the everyday use and effectiveness of *Habito* towards motivating physical activity. The third section presents our findings and discussion of our results. Our findings are split into three main sections: First we discuss the adoption of *Habito*, and how this varies according to participants' attitudes towards physical activity. This is followed by an analysis of the patterns of use of the adopters of *Habito* – i.e., individuals that used *Habito* for at least a week. Our analysis concludes with an analysis of the effectiveness of *Habito*'s different strategies for behavior change – towards motivating physical activity, and usage. The article ends with a discussion of some implications for design.

This article was published in UbiComp 2015 [32] with co-authors Dr. Evangelos Karapanos and Prof. Dr. Marc Hassenzahl and additional contributors of Sergio Barros. I led the design of Habito, analysis of data and writing drafts of the paper, under the supervision of Dr. Evangelos Karapanos. Development of the mobile application was led by Sérgio Barros.

How Do We Engage With Activity Trackers? A Longitudinal Study of Habito

Rúben Gouveia
Madeira Interactive Technologies Institute, Portugal
rubahfgouveia@gmail.com

Evangelos Karapanos
Madeira Interactive Technologies Institute, Portugal
e.karapanos@gmail.com

Marc Hassenzahl
Industrial Design, Folkwang University of Arts, Germany
marc.hassenzahl@folkwang-uni.de

ABSTRACT
We report on a 10-month in-the-wild study of the adoption, engagement and discontinuation of an activity tracker called *Habito*, by a sample of 256 users who installed the tracker on their own volition. We found 'readiness' to behavior change to be a strong predictor of adoption (which ranged from 56% to 20%). Among adopters, only a third updated their daily goal, which in turn impacted their physical activity levels. The use of the tracker was dominated by *glances* – brief, 5-sec sessions where users called the app to check their current activity levels with no further interaction, while users displayed true lack of interest in historical data. Textual feedback proved highly effective in fueling further engagement with the tracker as well as inducing physical activity. We propose three directions for design: designing for *different levels of 'readiness'*, designing for *multilayered and playful goal setting*, and designing for *sustained engagement*.

Author Keywords
Personal informatics; persuasive technologies; behavior change technologies; physical activity trackers.

ACM Classification Keywords
H.5.2. User Interfaces: Evaluation/methodology.

INTRODUCTION
Chronic diseases account for nearly 40% of mortality and 75% of healthcare costs worldwide, with obesity alone being responsible for an estimated 12% of the total health spending growth in the United States [39]. Consequently, policy makers argue for a health care model that stresses patient-driven prevention rather than after the fact cure. This burst of interest in prevention combined with progress in technology has lead to a whole new genre of products: wearable activity trackers. Their market has rapidly grown to a volume of $1.15 billion worldwide in 2014 [33].

UbiComp '15, September 7-11, 2015, Osaka, Japan.

http://dx.doi.org/10.1145/2750858.2804290

Research on activity trackers so far led to a repertoire of theoretically informed design strategies, such as making use of deliberate goal-setting [6,31], increased self-monitoring [5,24], as well as the exploitation of social influence [31]. Yet, despite promising early results, more recent studies raised concerns about activity trackers' long-term efficacy [19]. Shih et al. [40] studied the adoption of *Fitbit* – a wearable activity tracker – by 26 users. They found that 50% quit using the tool after only two weeks. Similarly, a recent survey [22] revealed that over a third of owners of commercially available trackers discarded them within six months after purchase.

While the quick and profound disengagement with trackers seems disheartening, we do not even know whether this is not actually a positive sign. Trackers as currently designed work primarily as "scaffolding" [12]. They provide structure and motivation to people who feel incapable of implementing their intention of exercising without support. In terms of Deci and Ryan's *Self-Determination Theory* ([9], p. 237), motivation to exercise has to be transformed form external to internal, often through steps of introjection (e.g., the tracker embodies exercising as an activity one should do), identification (e.g., exercising is accepted as necessary) and finally integration (e.g., exercising becomes an intrinsically-motivated activity). Thus, disengagement can signify two opposite outcomes: failure to integrate exercising into daily life or a swift adoption of exercising as an intrinsically motivated practice.

In fact, the majority of the studies have focused primarily on the impact of the tracker on behavior rather than, for example, users' intensity of engagement with the tracker. However, we find user engagement to be an important mediator variable for a number of reasons. First, the most commonly employed strategy for behavior change, self-monitoring, requires engagement. So while we may design features to provide self-relevant feedback, it can only impact behavior if people engage with these features [21]. Second, recent studies have revealed rich qualitative findings on the diversity of motives and behavioral practices that surround the use of physical activity trackers. For instance, prior work has suggested that activity trackers

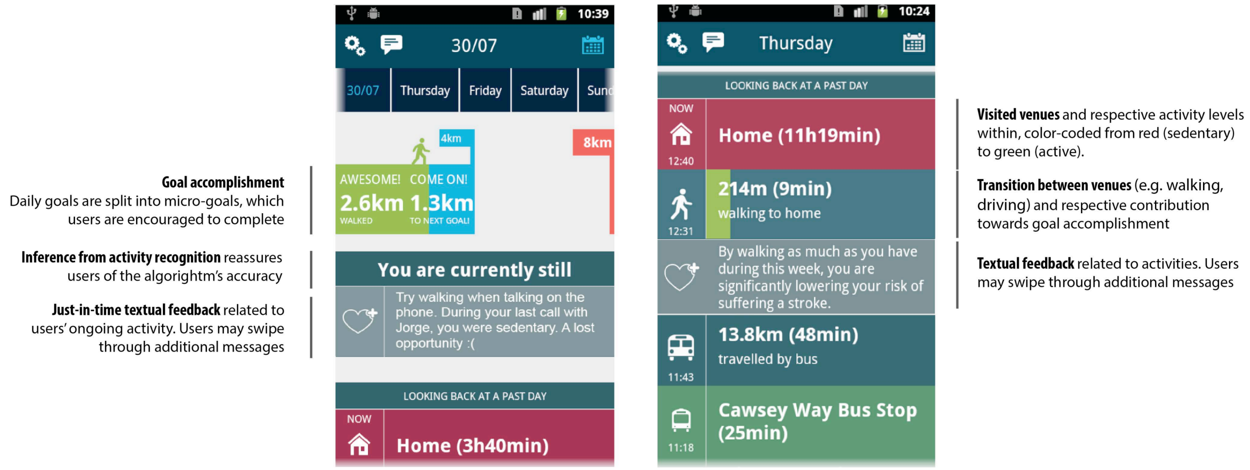

Figure 1. *Habito* employs three design strategies: *goal setting*, *contextualizing physical activity* and *textual feedback that keeps updating*. The in-the-wild deployment of *Habito* aimed at exploring its adoption, how users engage with feedback, and the impact the design strategies had on users' engagement with the tracker and likelihood to engage with physical activity on the short term.

serve both persuasive and reflective goals [12]. However, we have no knowledge as to when and how users engage with trackers to reflect or to be persuaded. The nature of those different interactions is likely to become apparent through the examination of the frequency and duration of users' engagement with the tracker.

This paper reports an in-the-wild study of 256 users who voluntarily installed *Habito*, a specifically designed physical activity tracker that runs on Android OS. We were interested in acquiring an unbiased estimate of the adoption rate of *Habito* and in understanding how adoption is affected by users' 'readiness', i.e., their motivational stage. Subsequently, we took a closer look at the frequency, duration, and nature of users' engagement with *Habito* itself, the way engagement changes over time, and the impact engagement has on physical activity levels.

In the following section, we describe *Habito* in more detail.

HABITO

Habito (see Figure 1) was specially designed and built to study how users engage with activity trackers. Designing our own application had a number of benefits. First, it enabled us to test different approaches to activity tracking through manipulating the type of feedback given. Second, while commercially available trackers (such as *Fitbit* or the *Moves* app) provide access to people's physical activity data, their APIs have a number of limitations, such as granting access only to the past week's data or providing no tracking of users' actual interaction with the tracker. *Habito* allowed us to not only capture users' physical activity, but also their interactions with the app. *Habito* was developed for Android OS, which allowed us to reach an unbiased sample of users through its deployment on Google Play.

The present version of *Habito* employed three design strategies: *goal setting*, *contextualizing physical activity* with cues relating to location and daily commutes, and *textual feedback that keeps updating* with the goal of sustaining users' interest.

Goal setting

Goal setting is one of the most popular, theoretically informed and empirically grounded approaches to instill behavior change. Research has repeatedly shown that setting concrete goals makes individuals more likely to accomplish them [16,26]. However, goals have to be adopted to become active. The availability of the mere functionality is irrelevant, as long it is not used. Accordingly, knowledge of the extent to which individuals change the defaults goals, of how frequently they update goals, and how those interactions affect engagement and physical activity, would help shed light on individuals goal setting practices and relation to behavioral outcomes.

Upon installing *Habito*, users were prompted to define their daily walking goal. However, a default goal was provided, as prior work has shown that many first-time users are uncertain about how much they walk (or should walk) [20]. While a walking distance of 8 km/day is recommended, we chose to set a default of 1 km/day. We did so for a number of reasons. First, prior work [20] has shown that users tend to underperform compared to medical recommendations. Challenging goals induces initial surprise, experienced as a wake-up call for some, but also induces reactance and higher chance of rejecting the tracker for less motivated individuals. Second, a lower default goal would be achieved easier and, thus, possibly motivate users to reflect (and update) towards their own, appropriate and attainable goal.

Habito provides an awareness of users' current activity levels and goal completion at the top of the screen (see Figure 1). To provide positive reinforcement throughout the day, we split their daily walking goal into four and provided interim milestones. For instance, a goal of 8 km would have three sub-goals: 2, 4 and 6 km. Upon completion of a sub-goal, *Habito* would "reward" the user and motivate her towards achieving the next sub-goal (e.g., "Awesome! 2.6 km walked! Come on, 1.4 km to next goal!").

To track users' physical activity, *Habito* makes use of (1) Google's activity detection API [13], which senses users' state of physical activity (e.g., still, walking, driving) over 30 sec intervals, and (2) an open source step counting algorithm that combines data from the phone's accelerometer and gyroscope [1]. Steps were counted only when 'walking' was detected by the activity detection API. This improved the accuracy of the step counter but also reduced battery drain considerably. Distance walked was further inferred from step count and users' height.

Contextualizing physical activity through location

Habito provides a view of user's daily visited locations and commutes between them (see Figure 1). A new location entry is made if a user spends at least 5 minutes within a 50-meter radius [43]. Locations can further be associated with a name, automatically identified in all subsequent visits. A list of suggested names is provided by *Foursquare*, along two additional places – 'home' and 'work'. Locations are color-labeled according to users' activity levels within this place, ranging from red (sedentary), through orange, to green (physically active). Commutes (as sensed through Google's activity recognition API) and walks outside of places are represented through additional entries.

Contextualizing physical activity through location assists the user in a number of ways. First, presenting additional memory cues (such as places and commutes) supports the recall of episodic memories [8,14], enabling users to identify particular instances of walking that contributed to their daily walking goal. Second, associating physical activity to places supports users in identifying patterns and ill habits, such as places where they are particularly inactive. This should prompt the development of strategies to overcome particularities of places (e.g., when watching TV at home, stand up and move within commercial breaks).

The idea of contextualizing information is not new. Li [23] argued that contextual information may enable users to identify the factors that affect their physical activity levels, eventually "increasing users' awareness of opportunities for physical activity" (p. 53) in the different activities of one's life. In fact, several authors [10, 23] have pointed out that enriching behavioral with contextual information – such as places or people – can reveal factors that affect behavior, and help users to make more informed decisions about how to change their behavior.

However, we have an only limited understanding of how users interact with such contextual information. Epstein et al. [10] explored user preference and perceived values of different visualizations of contextual information (e.g. maps with average time spent in different modes of transportation, and graphs with the total minutes of physical activity for a certain week). However, those insights were based on self-reports while no objective data exists on users' consumption of contextualized data or on their impact towards shaping behaviors (e.g. does a certain representation of contextualized information actually lead users to walk more or eat healthier?).

Textual feedback that keeps updating

The potential of textual feedback in inducing behavior change has been repeatedly highlighted [7]. Rather than "boring users to death with numbers and graphs" [17, p. 48], textual feedback is potentially able to tell a story, is less ambiguous and can help in making sense of the data captured by the tracker. Textual feedback can highlight patterns and draw immediate attention towards important information and instigate action [7] or support reflection over extreme behaviors [30].

Perhaps more importantly, textual feedback can take multiple forms, thus strengthening the tracker's capacity to sustain the novelty of feedback. Prior work has found instant information rewards, such as social media updates and incoming emails on smartphones, to have the capacity to form "checking habits: brief, repetitive inspections of dynamic content quickly accessible on the device" [34]. Consequently, one could wonder whether presenting users with novel textual feedback can lead to checking habits, which sustain engagement with the tracker.

Habito provides users with textual feedback based on their present and past activity levels. Following Munson's classification [30], *Habito's* textual feedback was designed to support either *reflection* or *persuasion*. *Persuasive* messages attempt to instigate behavior change by providing explicit recommendations (*e.g.,* "Try walking when talking on the phone. During your call with Bob, you were sedentary", "Last week, you reached your daily walking goal 2 times, try updating it to 8 km*"). Informational* messages, on the other hand, attempt to assist users in gaining better knowledge about their behaviors, avoiding to employ any form of recommendation or nudging (*e.g.,* "You are the second most active person at work", "You just burned 1560 calories, equivalent to 5 cheeseburgers").

Habito contained a total of 91 different messages, displayed to users over time and in certain conditions. Some of these messages aim to support further inferences about the activities performed. For instance, when registering high physical activity at a given place, messages provided further detail, such as "M-ITI has been your most active location of the week. On average, 400m more than any other location," "In your breaks at M-ITI, you walked an average of 50 meters. Other messages provided mere facts such as “Only 13% of children walk to school nowadays compared with 66% in 1970” or “Simple movements such as fidgeting, which includes knee shaking or pen tapping can burn up to 800 calories per day.” Others provided just in time recommendations such as “You have been sitting for 45 minutes. Try taking a break every 30 minutes,” during extended sedentary activities, while others try to create a sense of community, e.g., “M-ITI is the 2nd most physically active community in Madeira. Just 300 meters below the first (University of Madeira)”.

STUDY

Habito was posted on Google Play and voluntarily downloaded by users. Over the 10-month period of the study, a total of 256 users downloaded the application. All users had quit *Habito* by the end of the 10-month period.

Contrary to prior work [5,24], we did not sample for users with specific levels of physical activity or increased motivation for becoming fitter, as we wanted to reach out to a representative population of users. We however tried to understand if users’ commitment to exercise influenced their adoption of *Habito.* Upon installation of *Habito*, users received an e-mail with the *stage of change questionnaire* [28], which maps people’s motivations to change behaviors (i.e. to become more active) to Prochaska’s and Velicer’s [36] stages of behavior change: *precontemplation* – having no plan to become more active, *contemplation* – not being active but intending to become soon, *preparation* – trying but not yet being regularly active, *action* – being regularly active but for a period less than six months, and *maintenance* – being regularly active for the last six months or more. 54% of users completed the questionnaire.

Users were informed that their data would be stored and analyzed for research purposes. Next to monitoring physical activity and context, application usage was logged, including when the app was launched and quit as well as all interactions within, such as clicking on a specific location, commute or physical activity entry, swiping to a new message, or looking at past days.

Most users (42%) were located in Portugal, followed by United States (25%), United Kingdom (22%), India (9%) and China (2%). All users installed the application on their own volition and no financial incentives were provided.

FINDINGS AND DISCUSSION

"Readiness" for use: motivation and adoption

One hundred and sixty nine users (66%) used *Habito* longer than two days, 97 (38%) longer than a week and only 36 (14%) longer than two weeks. To identify adopters and non-adopters, we ran a k-means cluster analysis on the maximum number of days of use with the number of clusters inferred from the sum of squared error (SEE) curve. This revealed two groups: *adopters*, who used *Habito* for more than a week (97 of 256, 38%), and *non-adopters*, who quit within the first week (159 of 256, 62%). The former group used the application for a median of 11 days (IQR: 8-16), while the latter used the application for a median of 2 days (IQR: 1-4).

The resulting adoption rate of 38% is clearly below Shih's [40] conservative estimate of 50% for *Fitbit* purchasers. Of course, the present study involved downloading a free mobile app rather than purchasing a wearable device. App acquisition in general is highly exploratory, with only 69% of all apps being kept for longer than two weeks after downloading [37]. For health-related apps this is even worse: Only 1 out of 100 people keep the app, whereas, for example, *Whatsapp*, is kept by every second (50%).

We expected strength of motives to determine whether people adopt *Habito* or not. In fact, previous work has found adoption of interventions for behavior change to be higher in intermediary stages (contemplation, preparation) compared to all other stages (precontemplation, action, maintenance) [24]. 138 users answered the *stage of change questionnaire.* Of these, 43% were found to be in the intermediary stages of behavior change (contemplation: 19%; preparation: 24%). Table 1 shows the adoption rates per stage and in total, along with 95% confidence intervals.

Table 1. Adoption rates of *Habito* per stage of motivation to exercise.

Stages of behavior change	Adopters	%	95% CI (adj. Wald)
Precontemplation	5 of 36	14%	6%-29%
Contemplation	14 of 26	54%	36%-71%
Preparation	19 of 33	58%	41%-73%
Action	7 of 24	29%	15%-49%
Maintenance	4 of 19	21%	8%-44%
Total	49 of 138	36%	28%-44%

The overall adoption rate of 36% (note that this slightly differs from the 38% reported above due to the fact that not all users responded to the stage questionnaire) is clearly a consequence of the stage the person was in. Among the target group (contemplation, preparation), the combined adoption rate was 56% (33 of 59, adj. Wald 95%-CI: 43%-68%), while among the other stages (precontemplation, action, maintenance), the combined adoption rate dropped to 20% (16 of 79, adj. Wald 95%-CI: 13%-30%). A χ^2-test of independence showed adoption not to be independent from the stage a person was in, $\chi^2(1)=18.8$, $p<0.01$.

In sum, given a certain readiness on behalf of users, the adoption rate of *Habito* resembled that found by Shih [40] in the context of *Fitbit*. Obviously, readiness is a strong predictor of adoption, which must be incorporated into studies of the adoption of health-related apps and devices.

In the remainder of the analysis we focus on the adopters' engagement with *Habito*.

Engagement

The 97 adopters had 2737 individual usage sessions (median usage sessions per adopter=28, IQR: 15-45). A session was defined by the moment a user opens the application, until the phone was locked or the application was closed [34]. First we looked at the sessions themselves (and their duration), then we explored patterns across sessions.

Usage sessions

Usage sessions were brief, with 50% of them not longer than 10 sec and 81% not longer than 30 sec. The median session duration was 10 sec (IQR: 4-24). They were thus on average even briefer than in earlier studies of mobile phone use, which found 50% [44], 54% [2] and 61% [11] of usage sessions to last no longer than 30 sec. These differences are unsurprising, as mobile phones are multi-purposive, offering a large range of applications and features, explored sequentially within individual usage sessions (e.g. opening a social media application, switching to a news application and ending up browsing web content) [37]. In the current case, we only focus on using a single application.

Banovic' et al. [2] qualified a usage session as either *glance*, *review* or *engage* session. Glance sessions are brief interactions, in which users check information on the lock screen and then turn the screen off or let the phone timeout [2]. For *Habito*, we define *glance* sessions as sessions in which users open and close *Habito* with no additional actions or inputs (e.g., activating *Habito* to gain awareness of physical activity levels). In Banovic et al. [2], *review* and *engage* sessions involved access to at least one application. These differed however in terms of duration, with *review* sessions lasting up to 60 seconds and *engage* sessions lasting more than 60 seconds. This time split was determined through a head/tail classification [18]. Following this approach, our analysis revealed a natural break point on 22 seconds. *Review* sessions are thus sessions, which last up to 22 seconds, while *engage* sessions last more than 22 seconds, with both sessions involving at least one action within *Habito* (e.g., scrolling through the past day's performance).

Over half (57%) of all usage sessions were *glance* sessions (median duration=5sec, IQR: 2–11), while *review* and *engage* sessions were evenly distributed (review: 22%; median duration=12sec, IQR: 8–18; engage: 21%, median duration=45sec, IQR: 29–67). These results are similar to those of Banovic et. al. [2], who found 47% *glance*, 25% *review* and 22% *engage* sessions with median durations of 14, 23 and 136 sec, respectively.

Review and *engage* sessions were characterized by a high number of interactions with contextual and textual feedback. In fact, 88% and 89% of all *review* and *engage* sessions, respectively, involved exploring an ongoing days' contextual feedback, while exploring textual messages occurred in 84% and 88% of *review* and *engage* sessions. A χ^2-test of independence revealed a significantly higher frequency of accessing textual messages during *engage* as opposed to *review* sessions, $\chi^2(1)=4.11$, $p<0.05$. Further, users would access past days with a higher frequency within *engage* (18%) as compared to *review* sessions (13%), $\chi^2(1)=5.92$, $p<0.05$.

We found the type of session to be linked with goal accomplishment. *Engage* sessions were more frequent when goal accomplishment was low ($\rho(387) = -0.58$, $p<0.05$, see Figure 2) while *glance* sessions became more frequent as users progressed towards their set walking goals ($\rho(1084) = 0.35$, $p<0.05$). Moreover, the percentage of *glance* sessions would increase over time ($\rho(1053) = 0.41$, $p<0.05$), from 45% during first week of use to 68% and 73% during the sixth and twelfth week of use. Additionally, the percentage of *engage* sessions decreased over time ($\rho(1153) = -0.38$, $p<0.05$), from 28% during first week of use to 13% and 9% during the sixth and twelfth week of use.

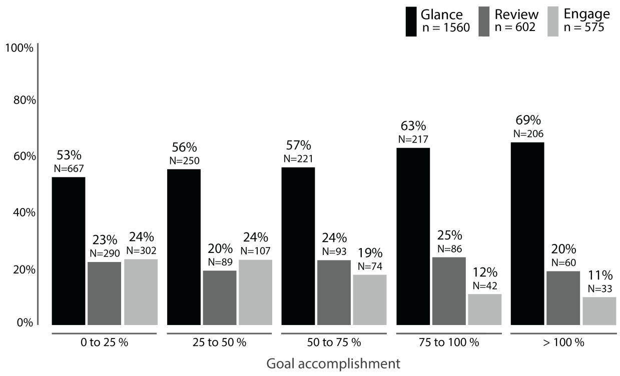

Figure 2. Users' engagement with *Habito* and goal accomplishment (Percentiles and frequencies)

All in all, these results support the notion of activity trackers as "deficit" technologies, to which people turn when they are afraid of failing. During low levels of goal accomplishment individuals exhibit higher dependency on the tracker, as signified by the prominence of engage sessions which focus on feedback. As users progress towards their goal, the prominence of engage sessions decreases and people use the tracker only briefly to acquire an awareness of their current progress towards goal completion (i.e., glance). Over time, when individuals become more self-reliant, use is more and more marked by brief, reassurance-seeking, glance interactions.

Pattern across usage sessions

Approximately a third (29%) of all usage sessions were separated by less than 5 minutes. This resembles the findings of Banovic et al. [2], with 50% of all sessions having been separated by 5 min or less.

We found the time users took to re-engage with the application to increase with their progression towards completion, $\rho(2605) = 0.21$, $p<0.01$, see Table 2. Further, nearly half (46%) of all usage sessions occurred during the first 25% of goal accomplishment, decreasing towards goal accomplishment, $\rho(1067) = -0.32$, $p<0.05$. These results further support the view of trackers as "deficit" technologies with users taking longer times to re-engage as they become more confident and progress towards their goal.

Table 2. Usage sessions and re-engagement time per goal accomplishment

Goal accomplishment	Usage sessions	Minutes to next session (IQR)
0-25%	46% (N=1259)	10 (1-51)
25-50%	16% (N=446)	27 (2-147)
50-75%	14% (N=388)	82 (1-330)
75-100%	13% (N=345)	133 (9-312)
>100%	11% (N=299)	189 (22-405)

Contrary to what we expected, users would take less time to re-engage with *Habito* after an *engage* session compared to a *glance* session (see Figure 3). In fact, transitions between subsequent *engage* sessions had the lowest re-engagement time (median=4min, IQR: 1-22) than compared to subsequent *glance* sessions (median=27min, IQR: 6-143, Mann-Whitney *U*=54418, p<0.01).

With *engage* sessions displaying short re-engagements and high frequency during lower levels of goal accomplishment, these results hint towards heightened dependency on the feedback and support provided by activity trackers during moments of underperformance. These results may also indicate a break of longer tasks, such as exploring historical information, into closely related micro-tasks, defined by shorter sessions, such as exploring past days over multiple sessions.

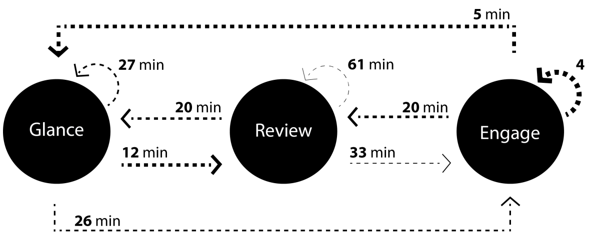

Figure 3. Median transition time between glance, review and engage sessions.

Impact of *Habito's* design strategies

Next we looked at how the three embedded strategies – goal setting, contextualizing physical activity through location and the continuously updating textual feedback – affected users' engagement with the tracker and their levels of physical activity.

Goal-setting

We found only 30 of the adopters (31%) to change their preset goal of 1km per day walking distance. Even more, the large majority of those users (24 of 30, 80%) updated their goal only once: during the first use of *Habito*. During subsequent use, 87% (N=84) of all adopters were recommended, at least once, to update their walking goal (e.g., "last week, you reached your daily walking goal 2 times, try updating it to…"), yet only 5% of them (4 of 84) followed the tracker's recommendation. The median of all updated goals was 8 km (IQR: 5-10).

Figure 4 illustrates the differences among users who updated their goal (self-set goal) and those who didn't (preset goal) in terms of engagement with the tracker as well as physical activity. As one may notice, while both groups had similar patterns of engagement with the tracker, significant differences are found in their levels of physical activity. Users who updated their goal walked more per day (median=6 km, IQR: 3-10) when compared to users that did not update their goal (median=2 km, IQR: 1-4, Mann-Whitney *U*=263, p<0.05).

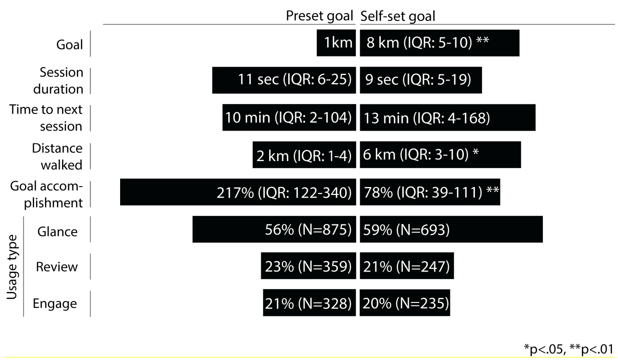

Figure 4. Users' engagement with *Habito* and physical activity levels (Median and IQR values) for those who updated the preset walking goal and those who did not.

This is in line with goal setting theory that argues that setting "difficult goals consistently leads to higher performance than [just] urging people to do their best" [26, p. 706]. However, this did not imply that they were more likely to meet their goal (see Figure 4). In fact, despite a positive correlation between goal and the actual distance walked per day ($\rho(1068)=0.51$, $p<0.05$), we found a negative correlation between goal and goal accomplishment ($\rho(1068)=-0.67$, $p<0.01$), which implies that setting a high goal decreases the chance of achievement, but increases physical activity. Supporting users in finding the optimal goal in terms of challenge and achievability is a relevant challenge for activity trackers.

Contextualizing feedback though location and commutes

Users only accessed contextual information in approximately one third (38%) of all usage sessions. This percentage decreased over time, ($\rho(1068)=-0.41$, $p<0.05$), from 43% during first week of use to 25% and 18% during the sixth and twelfth week of use.

Interactions with contextual feedback concerned in most cases (89%) the ongoing day, in 7% the past day, and only in 4% a day further in the past. In fact, of all sessions where users looked at a past day's feedback, only 29% involved exploring contextualized feedback. In the majority of cases (71%) it only involved glancing at the overall distance walked during that day.

All in all, these results suggest that contrary to conventional wisdom in personal informatics that portrays behavior change as the result of deep knowledge about one's own behaviors, users lack the interest to reflect on past behaviors. Users' interactions with historical feedback concerned only a small fraction of total use, which further decreased over time, and when users' interacted with past behaviors, these interactions mostly concerned the ongoing, rather than past days. One would expect our added contextual cues (such as location visits and commutes) to strengthen users' capacity to reconstruct past days, which should make past days' history more meaningful and interaction more likely. This was not supported.

One should, however, note that the chosen representation of context may not have been ideal for supporting learning and sustaining users' interest on the long-term. Our representation of context followed the line of the commercial application *Moves*, in which location and physical activity information are presented as unprocessed streams of data in the course of a day. Such low-level representations of context have been found to be less valuable in uncovering the factors that influence behaviors when compared to high-level representations of physical activity (e.g., tables with overall exercise performed during a week at work), as they require "paging days of detailed data to attempt to find trends, correlations, or opportunities for change" [10, p. 2].

Novelty in textual feedback

Users were presented with textual messages (from a pool of 91 different messages) that provided further insights into their physical activity levels. In approximately one third (32%) of all usage sessions, users were presented with a novel message (i.e., one they had not seen before).

When presented with a novel message, users were more likely to swipe to additional messages (median additional messages explored=2, IQR=0-4) as opposed to when presented with a familiar message (median additional messages explored=0, IQR =0-2, Mann-Whitney U=164553, $p<0.05$). Altogether, after a novel message, users interacted longer with *Habito* (median duration=15sec, IQR: 10-33) than when presented with a familiar message (median duration=7sec, IQR: 2-15, Mann-Whitney U=192711, $p<0.05$, see Figure 5). A χ^2-test of independence further revealed a significantly higher likelihood of resulting to an *engage* session, when novel messages were presented (219 of 876) as opposed to familiar messages (354 of 1864, $\chi^2(1)=13.01$, $p<0.01$), but a significant lower likelihood of resulting to a *glance* session (464 of 876), when novel messages were presented as compared to familiar messages (1100 of 1864, $\chi^2(1) = 185.5$, $p<0.01$).

Besides engaging longer, novel message made users return to the application in a shorter period of time (median=13 min, IQR: 2-238), as compared to when a familiar message was presented (median=21 min, IQR: 3-318, Mann-Whitney U=212971, $p<0.05$).

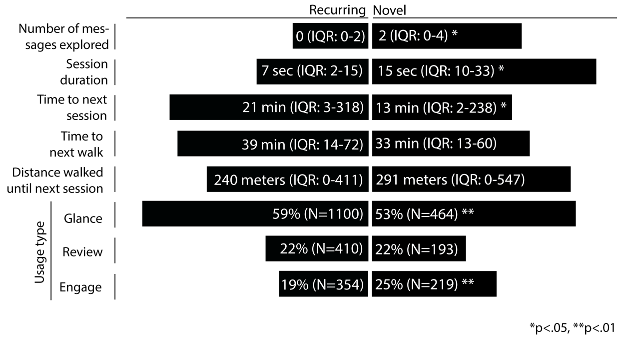

Figure 5. Impact of novel messages on users' engagement with *Habito* and physical activity (Median and IQR values).

Did these bursts of interest that novel content brought inspire users to walk more? Unfortunately not, since no significant differences were found in the time users waited before taking the next walk, or in the distance walked after interacting with a novel or a familiar message (see Figure 5).

All in all, these findings highlight the impact novel content can have on users' engagement with the tracker, both on a single session level (e.g., duration) and in terms of overall patterns of interaction (e.g., time to next usage). However, novelty per se – while intensifying engagement with the tracker – does not translate directly into the target behavior.

Persuasion in textual feedback

We employed two different types of messages in *Habito*: *persuasive* – messages that suggest activities such as "Try walking when talking on the phone. During your call with Jorge, you were sedentary" – and *informational* – messages that provide summative feedback, such as "You just burned 1560 calories, that is equivalent to 5 cheeseburgers"). Prior work has shown that while persuasive messages hold significant motivational power, they can lead to aversion and reactance [3]. Our interest is to understand the impact of both types of messages on engagement with the tracker, and to assess the overall value of persuasive messages with respect to users' level of physical activity.

Approximately two thirds of usage sessions presented exclusively either *persuasive* messages (30%) or *informational* messages (35%). *Persuasive* messages led to briefer engagement in the respective session (median=7sec, IQR: 3-11) compared to *informational* messages (median=13sec, IQR: 5-21; Mann-Whitney U=140825, $p<0.05$). Moreover, users would take significantly more time to re-engage with *Habito* following *persuasive* messages (median=22min, IQR: 5-293) compared to *informational* messages (median=13min, IQR: 2–248, Mann-Whitney U=135533, $p<0.05$).

However, while persuasive messages led to greater time until re-engagement, users would take less time to start walking and walk for longer distances when presented exclusively with *persuasive* messages ($median_{timewalk}$=29 min, IQR: 16–65, $median_{distancewalk}$=359m, IQR: 0–714) as opposed to *informational* messages ($median_{timewalk}$=38 min, IQR: 21–80, Mann-Whitney U=133181, $p<0.05$, $median_{distancewalk}$=203m, IQR: 0–493, Mann-Whitney U=113211, $p<0.05$).

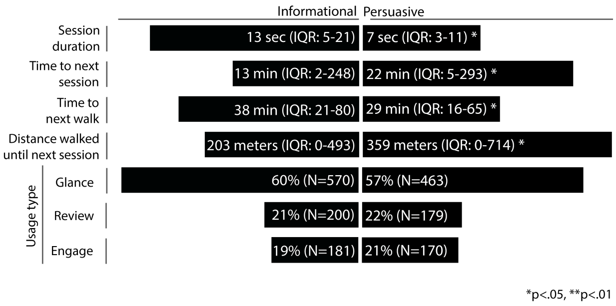

Figure 6. Users' engagement with *Habito* and physical activity (Median and IQR values) when interacting with exclusively informational or persuasive messages.

All in all, our findings seem to support previous research on the dual nature of persuasive messages [3]: while instigating action in the short-term, aversion and reactance also arise, potentially constraining long-term engagements. Further research should employ in-situ methodologies such as Experience Sampling to further inquire into how these effects are mediated through users' subjective experience, such as a momentary decrease in users' perceived autonomy. Next, building upon Munson's [30] guideline for context sensitive messages, research should further estimate the impact context sensing can bring to persuasive messages on increasing the likelihood of opportunistic behavior change and diminishing negative feelings.

IMPLICATIONS FOR DESIGN

All in all, our findings highlight the complexity of the adoption of activity trackers. In the remainder of the text, we discuss some implications for design.

Designing for different levels of 'readiness'

Similar to, but even more than Lin et al. [24], we found 'readiness' for change to be a strong predictor of adoption. Individuals in the contemplation and preparation stages had an adoption rate of 56%, whereas individuals in precontemplation, action or maintenance stages had an adoption rate of only 20%. This has a number of implications for the design and evaluation of physical activity trackers.

First, it reminds us to take into account people's current motivational stage when evaluating the efficacy of behavior change technologies. Without acknowledging their readiness to behavior change, comparisons of adoption rates and behavior change across studies may not be meaningful.

Second, available findings suggests that current trackers are most likely to work at intermediate stages of behavior change, where individuals have the intention but not yet the means (i.e. motivation, strategies) to change. This leaves out about 57% (in our sample) of the total of potential adopters. Consequently, opportunities to support individuals in the remaining stages is a worthwhile question for the design of activity trackers. For instance, considering the precontemplation stage, an opportunity could be to instill a desire for change rather than merely supporting the process of change. Individuals in the precontemplation stage are often unaware of the extent of their inactivity and are, thus, unwilling to change their behaviors. While existing trackers, if used, just confront them with this "truth" – unblinkingly, in the disguise of a seemingly neutral number – this may turn their initial experiences into something negative, marked by dismay [20]. Engaging users' in the precontemplation stage requires an experiential focus (see [29] for a range of techniques applied by doctors in the precontemplation stage) – one that asks how to increase users' perceptions of self-efficacy and competence.

Designing for multilayered and playful goal setting

Only a third of adopters changed the default daily walking goal. The fact that those who changed their goal walked more, raises the question of how to motivate individuals to reflect upon potential goals and to deliberately set one that is challenging, but achievable. One approach might be enforcing goal-setting. For instance, the commercial tracker 'Basis' asks the user to update their goal once per week. Our results however, seem to indicate this may not be an ideal strategy, as users did not update their goals when recommended to do so, even though achieving their daily goal on a regular basis.

One may attribute this low adherence to more demanding goals as a consequence of low levels of self-efficacy as goal achievement becomes more uncertain. In fact, previous studies have shown that individuals pursue goals more effectively when believing they can be achieved [25]. Activity trackers could allow multiple, simultaneous goals to be set, thus better accounting for the complexity of daily life. This would motivate users to pursue challenging goals in some days, while also guaranteeing a fallback towards more moderate, achievable goals in days that provide limited opportunities for physical activity [25]. Finally, trackers could build upon persuasive strategies to raise users' perception of the importance of maintaining challenging goals, while also providing reassurance and support regarding their attainability. For instance, opportunities to update goals could be highlighted (e.g. an upcoming weekend or hiking trip), as well as motives towards why goals should be updated (e.g. "setting challenging goals can potentially lead to higher walking distances!").

Further, while setting a challenging, explicit goal, would be beneficial, prior work has shown that users may have no idea how much they should walk and how they compare to others with similar lifestyles [20]. Forcing goal-setting in such a situation is likely to be aversive, as goals may be set above one's real capabilities. One solution would be to devise more playful strategies to goal setting. For example, one's daily walking goal could be set randomly (in sensible bounds of course) like a daily "lottery". It could depend on the weather, drift as users become more skilled with time to make the user more curious about the goal, or it could be imposed by his or her social network. All these would be playful techniques to involve users into a more deliberate reflection and goal-setting.

Designing for sustained engagement

We found 57% of usage sessions to be *glances*, increasing over time to 73% by the twelfth week. *Glances* are brief – with a median duration of 5 sec, spread throughout the day, and mainly provide immediate awareness of one's physical activity. On the contrary, *engage* sessions, where users would spend more time reflecting on the contextual and the textual feedback, were rarer and more frequent during moments of underachievement. As users progress towards meeting their goals, and over time, *engage* sessions become less and less frequent.

On one hand, this supports the dual nature of trackers as "deficit" technologies that "scaffold" behavior during particular problematic moments in time, and as "transformational" technologies that instill and routinize new practices to the point that the technology is no longer necessary [12,17]. On the other hand, it highlights the importance of very brief interactions to maintain engagement.

While glances were frequent and became more frequent with time compared to other types of use, users still became gradually disengaged with the tracker. Note that this decrease in engagement did not lead to reduced physical activity. In fact, user engagement was negatively correlated both with the daily distance walked and the ratio of days in which one's walking goal was met. Similarly to Fritz et al. [12], our findings suggest that users come to disengage with the tracker as they become more likely to meet their daily walking goals. The tracker in its current form is no longer needed. Note however the limitation of our study. We have no knowledge as to what happened after users completely ceased to use and stopped logging their activity with *Habito*. Research has repeatedly found that once the intervention ceases to exist, individuals may relapse to prior stages of behavior change (see [35] for a review). These findings are likely to replicate in the context of activity trackers. Keeping up a minimum of engagement with trackers even in later stages is thus important.

But how could trackers sustain users' engagement? We outline below two plausible directions: *creating checking habits*, and *transitioning glances to moments for reflective engagement*.

Creating checking habits

How can we design behavior change technologies so that they entice users to keep checking their data? Such brief, but frequent interactions (i.e., glances), as we found, can drive much of the usage of the tracker and contribute towards sustained engagement.

Building upon recent work that highlighted smartphones' capacity to create strong *checking habits* [34] – "brief, repetitive inspection[s] of dynamic content" such as social media updates and incoming emails, we asked: *what if the feedback provided by an activity tracker is constantly updating?* In doing so, we would sustain the informational reward people attain from checking feedback, which is assumed to be the primary cause of the formation of checking habits. After all, *novelty* has been a well-established strategy for sustaining engagement in a variety of industries. Consider for instance, the computer gaming industry as well as the airline services, which regularly update their content to sustain user interest.

Our study highlighted that updating the tracker with novel textual messages has the potential to sustain engagement through getting users back to the application faster than when presenting familiar textual messages, possibly hinting towards the formation of checking habits [34]. While the introduction of novel messages per se did not lead to an increase in physical activity, our data suggests that when coupled with persuasive strategies, textual feedback has the potential to lead to an increase in physical activity.

Of course, novelty is not the only means to sustain engagement. Consider for instance the first two concepts illustrated in Figure 7. *TickTock* (left) visualizes when one was physically active, but only over the past half hour. As a result, feedback on one's physical activity is becoming a *scarce* resource [4], and checking habits might be created as a result. As another concept, *Catchup* (middle) enables just-in-time competition with one's past self by contrasting current goal completion (outer ring) to that attained at the same time yesterday (inner ring). Contrary to traditional ways of visualizing goal completion that require projection to establish whether one will meet his or her goal by the end of the day, *Catchup* provides simple, normative feedback (e.g., "you are 200m ahead of the distance you had walked yesterday by the same time"). As such, the user is likely to frequently reengage with the tracker in order to reassure herself that she is keeping up and ahead of yesterday's performance (see [15] for an elaboration).

Figure 7. *TickTock* (left) visualizes physical activity of only the past half an hour, thus rendering feedback a scarce resource. C*atchup* (middle) enables competition with one's past self through contrasting current goal completion (outer ring) to that attained at the same time yesterday (inner ring). *Predicto* (right) fosters surprise by predicting upcoming day's walking distance from one's sleep pattern, weekday physical activity patterns and recent walking tendencies.

Transitioning glances to moments for reflective engagement

While *engage* sessions are important for reflection and, thus, crucial to sustained behavior change, we found them to represent only a small fraction of use and their frequency to further decrease over time. While this could be understood as the "natural" use cycle of a "scaffolding" technology, one may still attempt to sustain engagement to prevent relapse. Turning glances, which are low on information and reflection, into engage sessions would, thus, be a potential objective. We understand glances as "portals" to deeper engagement with the tracker. One possibility is to avoid providing too flexible, all-embracing and customizable displays, but to provide well-crafted narrative content, a single, tailored and well-crafted story. Or better: The glanceable beginning of a story, which will further unravel when people chose to engage further. For instance, a smart watch may notify a user about his or her high sedentary levels and only through further interaction this story becomes more telling, by for example, providing physical activity levels over the past 30 minutes, creating more opportunities to reflect about reasons for the current lack of physical activity (e.g., place, time, habits).

Such displays could also be sensitive to users' current motivational state, by providing "stories" that match the state. For instance, we found users' informational needs, and consequently their interactions with the tracker, to evolve as they got closer to meeting their goals. While early on, during moments of underachievement, users may need reflection and rich information, later, it may be more about collecting extraordinary accomplishments. When repeated days of inactivity pass by, users may need a wake up call. Consider for instance, *Predicto* (fig. 7 right), a concept of a tracker that, with its ambiguity, aims at fostering surprise by predicting his or her walking distance from her sleep pattern, general physical activity patterns and recent walking. *Predicto* leverages on the unexpectedness of predictions (overestimation or underestimation) to capture users' attention and make sense of the data.

CONCLUSION

Our study explored the real-world use of an activity tracker. While most studies so far focused on the effectiveness of trackers, that is, their impact on users' levels of physical activity, we sought to better understand the subtleties of how users engage with trackers and how this in turn affects their physical activity.

We found adoption to be strongly influenced by users' 'readiness' concerning the required behavior change. This is certainly one of the reasons why our adoption rate is lower than that of early work [22,40]. The majority of prior studies had systematically biased samples. They selected a specific set of users that already had an appropriate level of 'readiness' for change [24] or provided financial incentives as rewards for participating in the study [5,24,31]. While such studies are useful for tailored efficacy evaluations and greatly advanced our understanding of the effectiveness of different design strategies, they have limited predictive power concerning the adoption and use of a tracker in 'real-life' [21]. Our study revealed the complexities of activity tracking in everyday life, with, for example, users lacking the motivation to set goals, interacting only very briefly with the tracker, and revealing a profound lack of interest in their own historical data. It further showed the dual nature of trackers as a "deficit" technology that "scaffold" change during particular moments in time, and as a "transformational" technology that instills and routinize new practices to the point that the tracker is no longer necessary [12].

We believe activity trackers to be an example for a whole new genre of interactive "transformational technologies" [12], only poorly understood, yet. Detailed and naturalistic inquiries into adoption, use, and disengagement increase our understanding of how current versions of such technologies are actually used. Through this deeper understanding, we may become able to evolve activity trackers from a "deficit" technology, providing the already motivated with some information, into a fully-fledged supportive technology.

ACKNOWLEDGMENTS

The authors wish to thank Sérgio Barros for his work on the development of Habito. This research was supported by the Portuguese Foundation of Science and Technology grants CMUP-EPB/TIC/0063/2013 and SFRH/BD/89796/2012.

REFERENCES

1. Bagi, L. (2013). android-pedometer (Version 1.4). Retrieved from https://github.com/bagilevi/android-pedometer, April 2014.
2. Banovic, N., Brant, C., Mankoff, J., & Dey, A. (2014). ProactiveTasks: the short of mobile device use sessions. In *Proceedings of MobileHCI'14,* pp. 243-252.
3. Byrne, S., & Hart, P. S. (2009). The boomerang effect: A synthesis of findings and a preliminary theoretical framework. *Communication yearbook*, *33*, pp. 3-37.
4. Cialdini, R. B., & James, L. (2009). *Influence: Science and practice*, 4. Boston, MA: Pearson education.
5. Consolvo, S., Klasnja, P., McDonald, D. W., Avrahami, D., Froehlich, J., LeGrand, L., Libby, R., Mosher, K., & Landay, J. A. (2008). Flowers or a robot army?: encouraging awareness & activity with personal, mobile displays. In *Proceedings UbiComp'08,* pp. 54-63.
6. Consolvo, S., Klasnja, P., McDonald, D. W., & Landay, J. A. (2009). Goal-setting considerations for persuasive technologies that encourage physical activity. In *Proceedings Persuasive'09,* pp. 1-8.
7. Consolvo, S., Klasnja, P., McDonald, D. W., & Landay, J. A. (2014). Designing for Healthy Lifestyles: Design Considerations for Mobile Technologies to Encourage Consumer Health and Wellness. *Human–Computer Interaction*, *6*(3-4), pp. 167-315.
8. Conway, M.A. (2009). Episodic memories. *Neuropsychologia*, *47*(11), pp. 2305-2313.
9. Deci, E. L., & Ryan, R. (2000). The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior. *Psychological Inquiry*, *11*(4), pp. 227–268.
10. Epstein, D., Cordeiro, F., Bales, E., Fogarty, J., & Munson, S. (2014). Taming data complexity in lifelogs: exploring visual cuts of personal informatics data. In *Proceedings of Ubicomp'14,* pp. 667-676.
11. Ferreira, D., Goncalves, J., Kostakos, V., Barkhuus, L., & Dey, A. K. (2014). Contextual experience sampling of mobile application micro-usage. In *Proceedings of MobileHCI'14,* pp. 563-572.
12. Fritz, T., Huang, E. M., Murphy, G. C., & Zimmermann, T. (2014). Persuasive technology in the real world: a study of long-term use of activity sensing devices for fitness. In *Proceedings of CHI'14,* pp. 487-496.
13. Google Developers. (2015). ActivityRecognitionApi. Retrieved from https://developers.google.com/android/reference/com/google/android/gms/location/ActivityRecognitionApi, April 2014.
14. Gouveia, R. & Karapanos, E. (2013). Footprint tracker: supporting diary studies with lifelogging. In *Proceedings of CHI'13*, pp. 2921-2930.
15. Gouveia, R., Perreira, F., Caraban, A., Munson, S., & Karapanos, E. (2015) You have 5 seconds: Designing Glanceable Feedback for Physical Activity Trackers. In adjunct proceedings of Ubicomp'15.
16. Gollwitzer, P. M., & Sheeran, P. (2006). Implementation intentions and goal achievement: A meta‐analysis of effects and processes. *Advances in experimental social psychology*, *38*, pp. 69-119.
17. Hassenzahl, M., & Laschke, M. (2015). Pleasurable Troublemakers. In S. Walz and S. Deterding, eds., *The Gameful World: Approaches, Issues, Applications*, pp. 167-195.
18. Jiang, B. (2013). Head/tail breaks: A new classification scheme for data with a heavy-tailed distribution. *The Professional Geographer*, *65*(3), pp. 482-494.
19. Karapanos, E. (2015) Sustaining User Engagement with Behavior Change Tools, Interactions 22(4), pp. 48-52.
20. Karapanos, E., Gouveia, R., Hassenzahl, M., & Forlizzi, J. (2015) "It's not that hard to walk more": Peoples' experiences with wearable activity trackers. M-ITI Technical Report.
21. Klasnja, P., Consolvo, S., & Pratt, W. (2011). How to evaluate technologies for health behavior change in HCI research. In *Proceedings of CHI'11,* pp. 3063-3072.
22. Ledger, D. and McCaffrey, D. (2014) How the Science of Human Behavior Change Offers the Secret to Long-Term Engagement. Retrieved from http://endeavourpartners.net/white-papers/, December 2014.
23. Li, I. (2012). Personal informatics & context: Using context to reveal factors that affect behavior. *Journal of Ambient Intelligence and Smart Environments*, *4*(1), pp. 71-72.
24. Lin, J. J., Mamykina, L., Lindtner, S., Delajoux, G., & Strub, H. B. (2006). Fish'n'Steps: Encouraging physical activity with an interactive computer game. In *Proceedings of UbiComp'06*, pp. 261-278.
25. Locke, E. A. (1996). Motivation through conscious goal setting. *Applied and Preventive Psychology*, *5*(2), pp. 117-124.
26. Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American psychologist*, *57*(9), pp. 705-715.
27. Lucero, A., Karapanos, E., Arrasvuori, J., & Korhonen, H. (2014). Playful or Gameful?: creating delightful user experiences. *interactions*, *21*(3), pp. 34-39.
28. Marcus, B. H., Selby, V. C., Niaura, R. S., & Rossi, J. S. (1992). Self-efficacy and the stages of exercise behavior change. *Research quarterly for exercise and sport*, *63*(1), pp. 60-66.

29. Miller, W. R., & Rollnick, S. (2012). *Motivational interviewing: Helping people change*. Guilford press.

30. Munson, S. (2012). Mindfulness, reflection, and persuasion in personal informatics. In *Extended Abstracts of CHI'12*.

31. Munson, S. A., & Consolvo, S. (2012). Exploring goal-setting, rewards, self-monitoring, and sharing to motivate physical activity. In *Proceedings of PervasiveHealth'12,* pp. 25-32.

32. Mutsuddi, A. U., & Connelly, K. (2012). Text messages for encouraging physical activity Are they effective after the novelty effect wears off?. In *Proceedings of PervasiveHealth'12,* pp. 33-40.

33. NPD (2013) DigPital Fitness and Wearable Technology. NPD Group's consumer technology report, September 2013

34. Oulasvirta, A., Rattenbury, T., Ma, L., & Raita, E. (2012). Habits make smartphone use more pervasive. *Personal and Ubiquitous Computing*, *16*(1), pp. 105-114.

35. Prochaska, J. O., DiClemente, C. C., & Norcross, J. C. (1992). In search of how people change: applications to addictive behaviors. *American psychologist*, *47*(9), pp. 1102-1114.

36. Prochaska, J. O., & Velicer, W. F. (1997). The transtheoretical model of health behavior change. *American journal of health promotion*, *12*(1), pp. 38-48.

37. Rahmati, A., Tossell, C., Shepard, C., Kortum, P., & Zhong, L. (2012). Exploring iPhone usage: the influence of socioeconomic differences on smartphone adoption, usage and usability. In *Proceedings of MobileHCI'12,* pp. 11-20.

38. Rooksby, J., Rost, M., Morrison, A., & Chalmers, M. C. (2014). Personal tracking as lived informatics. In *Proceedings of CHI'14,* pp. 1163-1172.

39. RWJF Health Policy Snapshot (2011) "What are the biggest drivers of cost in U.S. health care?"

40. Shih, P. C., Han, K., Poole, E. S., Rosson, M. B., & Carroll, J. M. (2015). Use and Adoption Challenges of Wearable Activity Trackers. In *Proceedings of iConference'15*.

41. Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *63*(2), pp. 411-423.

42. Trope, Y., & Liberman, N. (2003). Temporal construal. *Psychological review*, *110*(3), 403.

43. Venkatanathan, J., Ferreira, D., Benisch, M., Lin, J., Karapanos, E., Kostakos, V., Sadeh, N., & Toch, E. (2011). Improving users' consistency when recalling location sharing preferences. In *Proceedings of Interact*, pp. 380-387.

44. Yan, T., Chu, D., Ganesan, D., Kansal, A., & Liu, J. (2012). Fast app launching for mobile devices using predictive user context. In *Proceedings of MobiSys'12,* pp. 113-126.

3.3 – Critical Review of the Work and Conclusions

Our study inquired into the real life adoption and use of a mobile activity tracker, and its effectiveness for promoting physical activity. Within this section, I discuss the relevance of initial interactions for promoting the adoption of behavior change tools, and briefly suggest ways of motivating the adoption of these tools. Further, I critically reflect on the approach used within our study to assess the effectiveness of *Habito*.

Predicting and Motivating adoption from early interactions

Our work contributed to an understanding of how the adoption of a mobile activity tracker is shaped by users' motivations towards behavior change. We found adoption rates to be highly influenced by the stage of behavior change in which a user was. Users in intermediate stages of behavior change – characterized by having the intention but not yet the means to become physically active, had an adoption rate of 56%, whereas those in initial and advanced phases of behavior change – characterized by being unwilling to become physically active or already having physical activity incorporated as an intrinsically motivated practice, had adoption rates of approximately 20%.

On one hand, these results highlight the need for taking users' motivations in account when evaluating the adoption of activity trackers. Without acknowledging users' motivations towards physical activity, comparisons of adoption rates and behavior change across studies may not be meaningful. One challenge, here, is detecting users' motivations towards a behavior during their use of these tools (as noted in [41]). Previous work has found that individuals have frequent fluctuations in their motivations towards behavior change, while attempting to change a behavior. They progress through stages of behavior change not only cyclically, but also relapsing and jumping between stages before a behavior is discontinued [72].

The effectiveness of a tracker, and its respective design strategies for behavior change, is likely to change in accordance to these fluctuations. Strategies based on self-efficacy are more likely to be effective on days where one lacks motivation for physical activity; yet, less likely to be effective on days where physical activity is a highly intrinsically motivated practice. Future work should focus on analysing these changes, and investigating how to accordingly tailor the strategies employed by these tools.

Perhaps more interestingly, these results show that *Habito* was only adopted by a minority of users, with many abandoning use within early interactions and days of use. This seems to indicate a clear message: initial interactions play a key role towards determining the successful adoption of trackers for a significant population of users. These insights, in a way, are not a surprise. Faust and colleagues [27], for instance, found that users which frequently use their activity trackers within the first 30 days of use were twice as likely to continue using their device, and had on average, 11% higher chances of adoption during subsequent months. Similarly, Meyer et al. [60] found the long-term adoption of activity trackers to be strongly defined by users' initial patterns of use. In both cases, initial patterns of use were strong predictors of the adoption of trackers.

These results seem to highlight the need for strategies to predict and promote adoption within early interactions with behavior change tools. On one hand, a richer understanding is needed of the factors that affect the adoption (and abandonment) of these tools, and how they interplay with users' use of activity trackers. Future research could, for instance, attempt to combine data gathered from early interactions (e.g. usage patterns, accessed features) with experiential data (e.g. users attitudes towards behavior change), for developing a richer understanding of how different factors impact the abandonment (or adoption) of these tools for different people.

These insights could then, be used to design tailored feedback towards gauging long-term adoption. Taking users' attitudes towards behavior change as an example: when detecting less motivation towards physical activity, trackers could present it's user with strategies focused on self-efficacy and competence (e.g., encouraging messages from peers or family members); when higher motivation is detected, a focus could be placed on strategies targeted to promote fulfillment (e.g. pursuing challenging goals).

Our own study overlooked the significance of users' initial interactions towards the adoption of *Habito*. Given the goal of our study was to assess the long-term effectiveness of *Habito,* our analysis placed minimal focus on how initial interactions unfolded, and their role towards a longer-term adoption of *Habito*. For instance, our analysis focused exclusively on data from the adopters of *Habito* (i.e., individuals which used *Habito* at least one week). This left out a significant amount of users for which adoption was not as successful (i.e., which used them for less than one week) – mostly those in initial and advances stages of behavior change.

A broader analysis, focused on the initial interactions of all users could have provided deeper insights into *Habito*'s adoption rates – such as why adoption varied across individuals in different stages of behavior change, and how *Habito*'s design strategies contributed to these differences. *Habito*'s persuasive messages, for instance, were likely to have different motivational effects depending on one's state of behavior change. Individuals in initial stages of behavior change are often unwilling to adopt a target behavior. Confronting them, within early interactions, with messages that explicitly tried to persuade them towards a certain behavior may have instigated aversion, and shaped their abandonment.

Assessing the proximal effects of behavior change tools

One of the main goals of our study was to investigate *how* Habito's individual strategies for behavior change influenced its users behaviors. To this end, we collected and analyzed a rich data set arising from users' interactions with *Habito* - such as the duration and type of their interactions, and their instances of physical activity. We adopted an experimental framing on users' interactions with *Habito* towards understanding how our device was used, and how physical activity was brought about as the result of interacting with the different components of its system.

This approach provided at least two key advantages for assessing the effectiveness of *Habito*. First, it allowed us to inquire into the proximal impact of *Habito* on users' behaviors. By collecting data from users' interactions – such as the feedback conveyed to users within their individual usage sessions and their instances of their physical activity, we were able to gather hundreds of daily data points from each participant, and use this data to inquire into the short-term effects of interacting with *Habito.*

While traditional methods for evaluating behavior change tools (such as RCTs) have focused on evaluating the distal effects of these tools (e.g. did a user increase their daily levels of physical activity after using a tracker for four weeks?), the short-term nature adopted in our analysis provided more immediate insights into the effects of behavior change tools (e.g. how much time did a user take to perform physical activity after being presented with a persuasive message?). Second, because *Habito* combined a variety of different behavior change strategies, this approach provided a more granular view of how each of its strategies impacted users' behaviors.

Recent research has highlighted the advantages of using *proximal outcomes* for assessing the individual components of behavior change systems – as opposed to the distal effects of systems as a whole [45]. The goal of such studies is to build stronger evidence of how and why BCSs impact individuals' behaviors by investigating the most immediate effects of these systems on individuals' behaviors. For example, a proximal outcome for a reminder system in a physical activity intervention might be the time taken by an individual to perform physical activity, and the number of steps taken within this activity, after receiving a notification. This information can then be used to gain an clearer understanding of *if*, and *how* individual components of BCSs relate to distal health outcomes (e.g. the relation between a notification system, and an observed increase in daily step count of 500 steps). This not only allows behavior change tools to be evaluated in relatively short studies, lasting days or weeks, but also allow researchers to assess the how these effects change over time (as suggested in [45]).

Mixed-method approaches for inquiring into the effectiveness of activity tracking devices

One should note a number of limitations in our study. First, our approach inferred user intent from people's actions. For instance, while we were assessing the proximal effects of an interaction – we did not control for potentially confounding factors, such as the physical or social context of the individual. Users might have taken action upon their behaviors due to external factors – and not necessarily due to their interactions with their data. To minimize these effects, we suggest complementing log-data driven studies with approaches such as micro randomized control trials. Micro- randomized trials [46] are trials where participants are randomly assigned treatments from a set of possible treatment actions at several times throughout the day. Micro randomized trials might thus inquire into the proximal effect of treatments while accounting for potentially confounding factors.

Second, our study provided a limited understanding of how users' surrounding environment impacted their use of Habito, and accounted for the impact of Habito's strategies for behavior change. For instance, while our study found persuasive messages to lead to aversion and reactance, we do not know if this happened because users had limited availability to perform physical activity when suggested to do so (e.g. being busy at work), or because they simply felt they had already walked enough for a current day. However, as HCI researchers move increasingly out of the lab, and commercially available activity trackers become increasingly in people's possessions, understanding the context of use of these devices has never become so important. An inquiry into the contextual factors surrounding the use of *Habito* could have provided a richer understanding of the conditions under which *Habito*'s strategies were most likely to be effective [63], and reduce the burden from having to engage with these strategies when assistance is not needed or wanted.

Third, due to the purely quantitative, data-driven approach of our study, we lacked insights into the motives that drove the use of *Habito*. For instance, while our study classified users' engagements with trackers into *glance*, *review* and *engage* sessions, based on the duration of the usage session and the actions taken (e.g., whether users checked historical data), it provided limited understanding of users' goals within each usage session and the reasons that led them to check the tracker.

A qualitative inquiry into the use of *Habito* could have provided a richer understanding of the motivations that drove the use of *Habito*, and even its non-use. For instance, while we found users' interactions with historical feedback to concern only a small fraction of total use – which even decreased over time, we did not investigate how the chosen representation of historical feedback may have accounted for these results. Future studies should focus on mixed approaches for inquiring the use and effectiveness of these tools. By combining quantitative (e.g., usage data), with qualitative (self-reports) and direct, in-situ observations of users' interactions with these devices, this approach allows HCI researchers to create robust evidence about whether and how different strategies for behavior change work.

Chapter 4.

Activity Tracking *in vivo*

4.1 – Introduction

Our study with *Habito* provided rich, quantitative-driven insights into how users interact with their activity trackers in their daily lives – namely, the frequency under which activity trackers are interacted with, the nature of interactions, and the proximal effects of these interactions on users' behaviors. For instance, our study found over 70% of usage sessions to be related to *glances* – brief, up to 5 second interactions with trackers. Glances were assumed to support the frequent regulation of behaviors; people were assumed to glance to acquire frequent awareness of how they were progressing towards their goals, and introduce new actions when needed. These assumptions, however, were purely derived from users' usage data. We often lacked qualitative insights, from (or deriving from) participants, in order to confirm such assumptions.

Further, our study captured limited insights of the interplay between the use of these devices and the surrounding environment. For instance, while we classified users engagements with trackers into *glance*, *review* and *engage* sessions, based on the duration of their usage session and actions taken, we gained limited understanding of how these uses were influenced by users' settings and daily practices. Activity tracking, however, as a practice, is strongly influenced by the context in which takes place [57]. For example, Deborah Lupton's concept of lively data acknowledges that self-tracking is entangled with people's daily lives in an evolving, and hybrid way [56]. This concept assumes that personal data takes on new meanings and values as people purpose, and re-purpose, their data within their everyday lives. For example, when investigating how cyclists reviewed data collected during their rides, Lupton and colleagues found cyclists to attribute different values to their data, depending on the context under which their trips had taken place - such as the spatial conditions (e.g. how cold or windy it was, or the traffic conditions), or how their bodies felt at the time a trip took place (e.g. whether they were overcoming an injury or getting over a cold)

[58]. Sometimes, data was deemed as useless, because it failed to take into account how tired a user was, or how bad the weather was. In other cases, data was considered useful when it accounted for memories of their trips and their bodily sensations. In practice, tracking is open to multiple (re)interpretations, resulting from the different contexts under which data is tracked and interacted with.

Self-tracking, under Lupton's notion of live data, is meaningless in itself. It is enriched with meaning and value, as the result of complex interactions between the data generated by these devices and the context under which tracking is carried out [57]. This holistic view of self-tracking was insufficiently examined within our own study with *Habito* – as commonly within Personal Informatics literature[56]. Little knowledge was gained as to how activity tracking is integrated in people's everyday lives, and how it affects people through their practices [22], an idea captured by the *lived informatics* framing of Rooksby and colleagues [76].

In this chapter, I present an observational study of activity tracking that attempts to explore the embodied nature of activity tracking – namely, how people's use of activity trackers is shaped by, and entangled with the surrounding environment of use. I report on a study that combines behavioral data from trackers with video recordings from wearable cameras, in an attempt to understand how activity tracking unfolds in users' everyday lives. The use of wearable cameras may seem counter-intuitive for investigating the use of a device designed to be unobtrusive, but we found this method to be remarkably insightful.

On one side, studies of technology in use are notoriously difficult, having to rely on ethnographic approaches – such as observations of participants as they engage with digital technologies, or on self-reported data - either through interviews or online surveys (a common approach within self-tracking literature) [88]. These reports, while having considerable strengths, rarely cover the contextual and immediate experiences with trackers. As Brown and colleagues argument, subtle but significant interactions with technology may be lost as the data is based on self-reports rather than actual behaviors as they unroll within the real world [23]. This problem is emphasized when attempting to study the use of technologies designed to be inconspicuous and largely invisible when in use. There is no one fixed place where the use of activity trackers can be observed and focusing on one aspect of use, such as explicit checking of sleep or step counting data may lead to overlooking other interactions with these devices [30].

Through the use of video methods, our study took a look at a number of practices that surround the use of these tools while being used in user's daily lives. Second, it provided moment-by-moment details of how activity trackers and the surrounding environment of use, are connected in use. I discuss the methodological considerations of this approach, and conclude with suggestions for how this method may be employed to further our understanding of activity tracker use in practice.

4.2 – Activity Tracking *in vivo*

This article is organized in four main sections. The first section presents a review of three different methodological approaches for the inquiry of activity tracking: *in the wild, in everyday life* and *in vivo*. We argue that early research on activity tracking has mostly focused on assessing the impact of these tools for changing people's behaviors, while recent research has increasingly focused on investigating the real-life practices emerging from owning and using an activity tracker.

The second section describes a study in which we use video methods to investigate some of the motivations that drive the daily use of activity trackers, and how this is shaped by different contexts of use. 12 individuals were provided with wearable cameras and monitored two days in their lives, collecting in total, 244 incidents where trackers were used. These recordings were combined with behavioral data from trackers as well as interviews, providing a detailed view on a number of practices that surround tracker use in daily life.

The third section presents our findings and discussion of our results. Our findings are split into two main sections: First, we provide an overall analysis of the basic functions of the different trackers used by our participants, and how they were used. This is followed by a deeper analysis of the role of trackers in users' everyday activities and routines. Our paper concludes with a proposal of three directions for the design of activity trackers.

This article was published in CHI 2018 [33] with co-authors Dr. Evangelos Karapanos and Prof. Dr. Marc Hassenzahl. I led the field study, as well as the collection and analysis of data. The paper was written collaboratively with Dr. Karapanos. The manuscript was revised by Prof. Dr. Marc Hassenzahl.

Activity Tracking *in vivo*

Rúben Gouveia
Madeira-ITI
Funchal, Portugal
rubahfgouveia@gmail.com

Evangelos Karapanos
Cyprus University of Technology
Limassol, Cyprus
e.karapanos@gmail.com

Marc Hassenzahl
University of Siegen
Siegen, Germany
marc.hassenzahl@uni-siegen.de

ABSTRACT
While recent research has emphasized the importance of understanding the lived experience of personal tracking, very little is known about the everyday coordination between tracker use and the surrounding environment. We combine behavioral data from trackers with video recordings from wearable cameras, in an attempt to understand how usage unfolds in daily life and how it is shaped by the context of use. We recorded twelve participants' daily use of activity trackers, collecting and analyzing 244 incidents where activity trackers were used. Among our findings, tracker use was strongly driven by reflection and learning-in action, contrasting the traditional view that learning is one of deep exploration, following the collection of data on behaviors. We leverage on these insights and propose three directions for the design of activity trackers: *facilitating learning through glances*, *providing normative feedback* and *facilitating micro-plans*.

Author Keywords
Physical Activity Tracking; Personal Informatics;

ACM Classification Keywords
H.5.2. User Interfaces: Evaluation/Methodology.

INTRODUCTION
Over recent years, physical activity trackers have received an upsurge of interest both in research and practice. Consumer interest in commercially available devices has tripled over the last couple of years [45]. Ownership rates followed similar trends, with one in every two U.S adults claiming to own, or have owned, an activity tracker [44]. This trend is accompanied by a rhetoric about the potential health benefits gained from the use of these devices. Individuals have been found to walk more [9], lose weight [2] and feel more in control of their behaviors [8] when consistently monitoring their step count. These benefits seem to prevail over months or years of continued monitoring [2,50].

CHI 2018, April 21–26, 2018, Montreal, QC, Canada

ACM ISBN 978-1-4503-5620-6/18/04...$15.00
https://doi.org/10.1145/3173574.3173936

However, more critical research repeatedly pointed out the complexity of tracking as a social practice and called for a deeper understanding of how trackers become a part of people's everyday life [see 16,18,47]. For instance, our understanding of self-tracking practices has moved beyond the conventional vision of trackers as mediators of change [10,47]. Trackers are used to understand routines, learn something out of interest or simply have fun and satisfy curiosity, as opposed to a stricter regulation of behaviors. Activity tracking happens alongside other daily activities. Users consult their trackers to gain momentary insights in-situ, when involved in an activity, as opposed to the conventional idea of deep, retrospective explorations of own data [23,47].

Unfortunately, most of the research highlighting the complexity of activity tracking as social practice relied on self-report data, gathered either through interviews or online surveys. Typically, participants are asked to recount typical use patterns and resulting experiences. These reports, of course, suffer from potential recall and social biases. For example, the interplay between technology and the particular context of use may be lost when studies solely focus on self-report [35].

In this paper, we report on an *in vivo* study of activity tracking. We provided 12 individuals with wearable cameras and monitored two days in their lives with tracking. We combined these recordings with behavioral data from trackers as well as interviews, providing us with a detailed view on how trackers are used in everyday life.

In the remainder of the paper, we will first provide an overview of empirical studies of activity tracking. We identified three methodological paradigms: the study of activity tracking *in the wild*, *in everyday life* and *in vivo.* Second, we describe a number of practices that surround the use of activity trackers. Most notably, we find tracking to be driven by reflection and *learning-in-action* rather than by learning through retrospective, deep exploration. Third, we leverage these findings and propose three directions for the future design of activity trackers: *facilitating learning through glances*, *providing normative feedback on goal accomplishment*, and *facilitating micro-plans*.

BACKGROUND
In this section, we discuss three paradigms for the inquiry of activity tracking: in the wild, in everyday life, and in vivo.

Activity tracking in the wild

Early studies of activity tracking often entailed the design, development and field trial of a novel research prototype. Driven by a theoretical concern, their goal was to assess the efficacy and user acceptance of the prototypes and their underlying design strategies. For instance, Consolvo and colleagues explored how mobile technology can motivate physical activity through social influence [13], goal-setting [38] and glanceable displays [14]. In [14], for instance, they designed *UbiFit*, a mobile application to provide feedback on one's physical activity through a glanceable, stylized representation on the background of a mobile phone, and employed it in a study with three experimental conditions to assess its impact on activity. More recent work has helped to untangle the design space of glanceable displays through the iterative design and analysis of twenty-one concepts of glanceable physical activity feedback [22]. Four of those concepts had been further prototyped and deployed in a comparative month-long study, inquiring into how users engaged with each prototype, the immediate impact on physical activity, and users' perceptions.

In the wild studies have been, and still are, instrumental to the development of the field. They move evaluation out of the lab, enabling the evaluation of novel technological systems in natural settings. Through quasi-experimental setups, and mixed-method approaches, they provide early insights both on which strategies work best, as well as why they work. However, these studies are limited in a number of ways. First, due to the complexity of the trial, and the technological limitations imposed by the maturity of the prototype, these studies are small in sample size and duration, typically involving 5 to 30 participants over a duration of 3 weeks to 3 months [51]. Second, participants are typically given a prototype, rather than purchasing a product on their own, and are incentivized to use the prototype for the duration of the study, which reduces the ecological validity of the study. Third, studies often featured a specific set of users with an already appropriate level of 'readiness' for change [33]. As a result, while these studies are extremely useful as efficacy evaluation of different design strategies [29], they have limited predictive power over the adoption and use of a tracker in 'real-life'.

The study of activity tracking in everyday life

With the widespread adoption of commercial activity trackers, researchers have increasingly shifted attention to the study of peoples' real-life practices emerging from owning and using a tracker. The focus of respective studies is not on the effectiveness and the efficiency of the technologies, but on how users appropriate those technologies, and how adoption, or non-adoption, is shaped by the context of use. Researchers produced models of how people use self-tracking tools informed by qualitative inquiry [16, 32]; investigated how people lapse and abandon these devices [29]; explored the ability of trackers to align with users' motivations and desires [25, 28]; and investigated tracking practices in everyday life [10,18,47] and in specific contexts, such as the workplace [11].

Most of the studies have been qualitative in nature, relying on users' self-reports, either through interviews or online surveys. For instance, Fritz et al. [18] interviewed 30 participants who purchased a tracker of their own volition and used it from 3 to 54 months. Their study revealed how the practices that surround long-term tracking were different from those of early adoption, and how non-beneficial practices were afforded by the design of the technology, such as "number fishing" which turned exercise from a meaningful intrinsic endeavor into an extrinsically-rewarded activity. Rooksby et al. [47] interviewed 22 people two times, separated by a month, and found five different motives for tracker use, from directive, to documentary and diagnostic tracking. Karapanos et al. [28] used a psychological needs framework to inquire into memorable experiences of 133 users through an online survey. Their study highlighted that tracking has a nuanced social component, from the sense belonging and social support provided through the online communities, to the stronger more direct social exchange among family members, when they purchased a tracker for a relative and joined in their efforts towards a better, healthier self.

Some studies leveraged the value of *secondary data*. For instance, Clawson et al. [12] analyzed 1600 advertisements of personal health tracking technologies on craigslist and found that individuals often abandon these, not due to technologies' failure, but often because they achieved their goal (e.g., lost weight), they desired an upgrade to a newer model, or because of unanticipated changes in their life (e.g. surgery). Choe et al. [10] analyzed video-recordings of Quantified Self talks and found that individuals often had a specific, personal, health-related goal, such as finding triggers for an allergy or the right drug dosage.

Other quantitative studies logged users' activities and interactions with the tracker in everyday life. For instance, Gouveia et al. [21] monitored the engagement of 256 users with an activity tracker called *Habito,* and its impact on users' physical activity over a period of ten months. They found that users rarely look back at their performance data. Over 70% of the interaction had been *glances*: brief, 5-sec sessions where users called the app to check how much they had walked so far without any further interaction. In addition, they found that current physical activity trackers work only for people in intermediary stages of behavior change (i.e., contemplation, preparation). Those displayed an adoption rate of 56%, contrary to that of 20% for people being in the remaining stages (i.e., precontemplation, action or maintenance). Moreover, contrary to a common assumption in personal informatics literature and Goal-Setting Theory [34], only 30% of users set their own daily step goal, while 80% of users who did so, never updated the goal again. Similarly, Meyer et al. [36] analyzed the behavioral data of 104 activity tracker users, over 14,413 days of use and found periodic breaks, lasting from a couple of hours to a couple of days, to be the norm in the use of activity trackers. They identified different patterns for activity tracker use such as try-and-drop, slow-starter, experimenter, hop-on hop-off, intermittent and power user.

Studies of activity tracking in everyday life advanced our understanding of how trackers are embedded and appropriated in users' routines. Due to their inexpensive format, qualitative studies based on interviews, surveys and secondary data, have inquired into the long-term adoption of and experiences with activity trackers, while quantitative, interaction logging studies provided a more realistic picture of how people actually engage with trackers, through the behavioral observation of many users, who acquired the trackers of their own volition, and used them over prolonged periods of time.

However, these studies are limited as well. As most of the studies rely on self-reports, asking people to retrospect on typical use patterns may suffer from recall bias [24]. Insights on actual behaviors and experiences are likely to be forgotten, overlooked, or avoided. While quantitative behavioral studies provide a more accurate picture of users' interactions, they provide an only limited understanding of the context that surround its use and the motives that drive it. For instance, while Gouveia's et al. [21] study classified users' engagements with trackers into *glance*, *review* and *engage* sessions, based on the duration of the usage session and the actions taken (e.g., whether users checked historical data), it provided limited understanding of users' goals within each usage session and the reasons that led them to check the tracker. As argued by McMillan et al. [35], "the coordination between technology and the surrounding environment of use may be lost when studies as solely focused on self-reported [and remote observation] methods."

The study of activity tracking in vivo

Motivated by the limitation of self-reports and remote observational studies of activity tracking, researchers turned to the use of video methods and direct, in-situ observation of users' interactions with activity trackers.

Patel and O'Kane [40], for instance, combined participant observation with interviews, to better understand how individuals use technology while exercising at the gym. Participants were asked to verbalize their thoughts and feelings while engaging with technology during a workout regime (e.g. engaging with the display of a treadmill or an activity tracker), while observed by a researcher. Similarly, Gorm and Shklovski combined participant observation with semi-structured interviews to better understand how trackers were used within a step-count campaign in the workplace. The first author participated in participants' "*work meetings, sat at a desk allocated to her in the open office space alongside the employees, joined in the lunch breaks, department meetings and Friday breakfasts, and generally partook in the daily life of the office during 12 workdays*" [23, p.151]. By using an in-situ approach, both studies provided insights into how trackers were used in these specific contexts.

While these studies are prime examples of in-situ inquiry, shadowing an individual over the course of a full day seems infeasible and may constrain how interactions with technology naturally occur [40]. One solution is the approach taken by Gorm et al [24]. They used a participant-driven photo elicitation method. Twenty-five novice users of trackers took photos of their "private" self-tracking experiences over the course of five months. Participants were instructed to take photos of events or experiences they felt were related to their activity tracking and send them via e-mail to the first author once a week and to include one or two sentences describing the photo or any thoughts they might like to share. Participants further reflected on their pictures in a follow-up interview, explaining why they took a certain photo. One of the identified limitations was the long follow-up time of interviews (in some cases, pictures were analyzed nearly 5 months after having been taken), as some participants had difficulties to recall the reasons that led them to capture a certain picture. Moreover, this method is limited in the sense it only highlights practices that are chosen to be shared by users, and in that they provide only a limited snapshot.

To counter for those limitations, Brown et al. [7] proposed the use of video methods in the study of mobile technology in everyday life. Mondada [37], for instance, used video-recordings to understand how phone calls unroll in work settings. Individuals were found to engage in *multi-activities* during calls, which could be interrupted, suspended, accelerated or perturbed by incoming calls. McMillan et al. [35] used video recording to understand how the use of the mobile phone becomes integrated into ongoing activities (e.g. how maps are used for route-finding), while Pizza et al. [42] examined how different features offered by smartwatches (e.g. time, notifications, activity tracking) were used in users' daily lives. Video methods offer naturalistic, visual perspectives on the use of technology as data are collected in-situ [37]. Moreover, these video logs may be revisited [26], thus creating the conditions for more precise recall of activities of interest, and allowing researchers and participants to become aware of aspects that might have been overlooked.

Despite the premise of video methods, according to the authors' knowledge, such methods have not been employed in the study of activity tracking yet.

AN IN VIVO STUDY OF ACTIVITY TRACKING

Participants

Twelve individuals were recruited in the study (5 female, 7 male, median age=28, min=21, max=41). They had all been using an activity tracker already for a minimum of 4 months (max=14, median=7 months). Our goal was to understand the practices surrounding these devices even after months of use. All participants were Portuguese. Nine had a full-time job, three were students.

Table 1. Summary of participant information

	Device	Months of use	Age	Gender	Occupation	Stage of behavior change	Daily step goal
P1	HR	9	33	M	Professor	Action	12,500
P2	HR	6	30	F	Teacher	Main.	18,000
P3	HR	4	23	F	Usability	Prep.	9,000
P4	HR	5	26	M	Research	Cont.	8,800
P5	HR	7	31	M	Security	Main.	15.000
P6	HR	5	22	M	Student	Cont.	12,000
P7	HR	8	29	F	Retail	Main.	13,000
P8	Flex	14	41	M	Lawyer	Main.	11,500
P9	Flex	7	25	M	Student	Prep.	15,000
P10	Flex	10	38	M	Designer	Prep.	8,500
P11	Flex	4	24	F	Student	Prep.	11,000
P12	Flex	6	21	F	Designer	Main.	12,000

The recruitment took place through local mailing lists and social media. We limited our search to *Fitbit* users to facilitate the comparison among participants' practices. However, we included two different versions from the same brand, varying considerably in features. Seven participants owned a *Fitbi*t *Charge HR* and five a *Fitbit Flex*. *Fitbit Flex* displays users' step count through five LEDs, each lighting up for achieving 20% of a daily walking goal. *Fitbit Charge HR* displays a wider range of metrics on users' physical activity – i.e. numerical step count, heart rate, caloric count, distance and floors, as well as the time and notifications on incoming calls. Users navigate through these metrics by tapping the *Charge HR*, or clicking on a side button (as seen in Figure 1). All *HR* users had the time as their primary watch face – having to click, or tap on their device to see physical activity feedback. These distinctions are taken into account when interpreting the results of our study. Both devices offer more elaborate feedback on physical activity, such as visualizations of one's step and sleep data over time, through a mobile and web application. Our analysis is constrained to the use of the wristband. We focused on the wristband, as prior work has shown that tracker usage is confined to short interactions [21], where users check their current activity levels with no further exploration of data.

We did not sample for participants with specific levels of physical activity or fitness and health goals. However, in line with previous research [21], we measured participants' stage of behavior change towards physical activity [43] in an attempt to understand how individuals' commitment to exercise influenced their use practices. As expected, our recruitment was biased towards physically active people: six out of twelve were in advanced stages of behavior change (i.e. action and maintenance), with the remaining in the intermediate stages (i.e. contemplation and preparation). Most set a higher daily walking goal than the 10K steps that are typically suggested by medical practitioners [49] (median = 12,000). All participants were rewarded with a 40€ voucher for taking part in the study.

Figure 1. Fitbit Flex (left), Fitbit Charge HR (right)

Method

Our study consisted of two phases: *a recording* and a *reconstruction* phase.

Recording phase

Throughout the recording phase, participants carried a *Xiaomi Yi* wearable camera and a bag, containing an external battery bank, allowing for up to 8 hours of video recording. The camera was mounted vertically, slightly above the chest of the participant. This setting allowed us to capture participants' interactions with their tracker, as well as the environment (see Figure 2).

Figure 2. Participants wore a camera during two days, providing insights into how tracker use unfolds in daily life

The first author conducted a meeting with all participants, in which he introduced them to the study apparatus. Participants were given a camera and asked to record on two days for eight hours, between 9am and 5pm. Participants were told to turn the camera off when they felt they needed privacy. They were free to choose the day in which they wanted to start recording. At the end of each day, the researcher met them to collect the recordings and to discuss any technical problems. Due to technical issues, two participants were only able to record approximately six hours of footage on their second day of recording.

We edited the video footage and extracted incidents in which participants interacted with their activity trackers. We refer to those as *usage sessions*. A usage session was defined by the moment in which the user brought the tracker at eyesight, to the moment the participant lowered his arm to its original position (similarly to [42] and [22]). We also extracted the fifteen seconds preceding, and following a usage session, in an attempt to gain insights on the aspects that lead up to and followed each usage session. We logged the duration, time of occurrence and nature of use of the session – i.e. *where* a session took place, *who* was present and *what* feedback was checked during a session (as in [31]). Each daily footage was edited by at least two individuals, immediately after being collected. We found no incidents of participants removing their tracker, either temporarily or permanently, during the recording.

Finally, we collected, and analyzed participants' physical activity data during recording. Participants were asked to log into the *Fitbit API* and grant access to their step count data. We used the Fitbit Intraday API to collect minute-by-minute data of participants' step count, as well as the corresponding levels of intensity for each minute – sedentary, lightly active, moderately active and vigorously active (as described in [19]).

Reconstruction phase

On the third day, participants took part in an interview. Participants were presented with their usage sessions and asked to recall these moments, focusing on the reasons that led them to interact with their tracker and how they thought the surrounding environment (e.g. location, ongoing activity and people) shaped their action. Participants were also asked to describe what they were doing prior to interaction. Insights were attached to the corresponding usage session.

Inspired by the Day Reconstruction Method [27], we displayed usage sessions chronologically, featuring the time in which they unfolded. Prior work has found temporal cues, such as timestamps in pictures or videos, to assist in the recollection of events and prevent misjudgments, such as the temporal misplacement of such events [20]. Video footage supported a more specific recall of the moments in which participants engaged with their trackers [20] and allowed them to take the leading role in the interview and provide their own insights into how usage unfolded. Each interview lasted about of 40 minutes.

The analysis of interviews grows out of thematic analysis. Interviews were transcribed, coded and organized into emerging themes (closely following the phases of thematic analysis suggested in [6]). Iterative rounds of discussion and refinement were performed between authors, looking for salient themes from interviews and observations of usage sessions. The prevalence of each theme was measured by counting the number of occurrences in which a theme appeared in the reflections of a usage session.

FINDINGS AND DISCUSSION

Tracker use

We collected 184 hours and 15 minutes of footage from all participants. Each participant recorded about 15 hours of footage. Participants were found to interrupt recording during situations considered inappropriate and in which they did not feel comfortable in wearing the apparatus (e.g. in a bathroom or work meeting).

A total of 244 usage sessions were identified. 82% (201 of 244) referred to the *Fitbit HR*, while 18% referred to *Fitbit Flex* (43 of 244). Participants engaged with the *HR* (N=7) for a median of 1.8 times per hour (IQR=1.7-1.9), while they engaged with the *Flex* (N=5) for 0.6 times per hour (IQR=0.4-0.7). A Mann-Whitney test revealed a significantly higher frequency of engaging with the *HR* (Mann-Whitney U=0, p<.01) as compared to the *Flex*.

While *HR* owners were found to engage more frequently with their devices, this did not mean they checked their physical activity more often. We grouped each individual usage session with the *HR* into three distinct categories, regarding the feedback offered by this device: a) sessions in which participants checked the time and engaged no further, b) sessions in which participants checked physical activity feedback and the time, and c) sessions in which notifications on incoming phone calls were checked. We further logged the time spent within each of these screens. An overview of this data is displayed in Table 2.

Approximately half of the usage sessions with the *HR* were time checking, with no engagement with further feedback (102 of 201, 51%). Physical activity related feedback was checked in approximately one third of usage sessions (65 of 201, 32%), while 11% (22 of 201) were triggered by notifications from incoming phone calls. The screen was not visible in the remaining sessions (12 of 201, 6%).

Table 2. Hourly checking rates and duration for different types of feedback checking

	Fitbit HR		***Fitbit Flex***	
Tracker use	**Hourly checking rate (IQR)**	**Duration, in sec. (IQR)**	**Hourly checking rate (IQR)**	**Duration, in sec. (IQR)**
Time	0.9 (0.8-1.1)	2.1 (1.7-2.7)	-	-
Physical activity	0.6 (0.6-0.6)	4.9 (3.4-6.7)	0.6 (0.4-0.7)	2.8 (2.2-3.7)
Notifications	0.2 (0.1-0.3)	3.5 (2.4-4.1)	-	-

Physical activity checking, with the *HR*, was not an isolated practice from time checking. Participants engaged with their trackers for a median of 5 seconds when checking physical activity feedback. Time checking accounted for approximately 25% (IQR=0-50) of the duration of these sessions (median=1.7sec). Only in 39% of the sessions (n=25) users spend less than a second in time checking.

As one may note in the data, while users engaged more frequently with Fitbit HR, no significant differences were found in the frequency of checking physical activity feedback between the two devices (Mann-Whitney U=16.5, p<.05), despite the the wider range of metrics offered by the *HR* (i.e. numerical step count, heart rate, distance walked, caloric count and floors climbed). Participants however, had longer usage sessions when checking their physical activity feedback on the *HR*, as compared to the *Flex* (Mann-Whitney U=0, p<.01; see Table 2).

Interaction to accomplish goals

Participants were more likely to engage with physical activity feedback when being physically active, as compared to when engaging in sedentary behaviors (see Table 3). In fact, participants checked their trackers 1.1 (IQR=0.8-1.3) times per hour while physically active, and only 0.3 times per hour while sedentary (IQR=0.2-0.4, Mann-Whitney U=10, p<.01).

Table 3. Engagement with physical activity feedback over different levels of physical activity

Level of physical activity (% of sessions, IQR)	Hours spent, of 18h of recording (IQR)	Hourly checking rate (IQR)	Duration, in sec. (IQR)
Sedentary (33%, 18-43%)	9.6 (8.6-10.6)	0.3 (0.2-0.4)	4.3 (2.9-6.0)
Lightly active (37%,14-44%)	4.6 (4.1-5.4)	0.8 (0.2-0.9)	3.4 (2.2-5.9)
Moderately active (20% , 11-40%)	1.4 (1.3-1.6)	1.5 (0.7-2.7)	3.6 (2.8-4.7)
Vigorously active (11% , 0-28%)	0.4 (0.2-0.5)	2.7 (0-6.9)	3.7 (3.1-4.2)

We found engagement during moments of physical activity to be linked to participants' desire to maintain or achieve optimal levels of performance. Participants often set strict goals, and engaging with the tracker served to ensure the goal was achieved before an activity was finished, e.g.:

[P9] "it felt like I had been walking forever, but I wanted to be sure I was already there (3km goal) before heading back home".

Figure 3. Users were nine times more likely to check their tracker during vigorous physical activity than when sedentary. Such engagements increased in frequency near goal completion and fueled users' motivation to meet their goals

Such engagements increased near goal completion and fueled participants' motivation to go beyond their levels of comfort in order to meet their goal, e.g.:

[P9] "the smell of the finish line kept me going… it's all good after I hit those numbers".

Trackers were also used to mediate the impact of an upcoming course of actions, regarding target behaviors. P7, for instance, was found to engage in heart rate checking before adjusting the speed of a treadmill (see Figure 3):

[P7] "it gives a sense of security… seeing that you're doing OK… sticking to your target, before taking that next step".

Learning in a glance

We found that glances serve towards learning gains. All participants, except one, were found to check their activity levels during, as well as right before the start and after the end of a particular activity, such as household chores [P3], commuting to work [P2], or doing groceries [P11] to gain insights on these activities (see Figure 4). These strategic engagements occurred frequently, accounting for approximately one third of participants' usage sessions (median=37%, IQR=22-67%).

[P11] "I am still surprised to learn how many steps I get each day with little things like chores around the house or walking around with my dog. Seeing these little things keeps it simple and interesting"

Figure 4. Trackers were checked strategically towards gaining insights on behaviors – such as right before starting, and after finishing walking a dog, to know how many steps were gained.

This strategic engagement with the tracker hints at a second path to knowledge, next to the deep exploration of past activity data through rich visualizations, one that is more flexible and that can better account for the complexity of daily life. For instance, while a tracker may segment physical activities based on location or time, users' inquiries were often more specific (e.g., how much did I walk while doing household chores). Moreover, participants frequently commented on the variability of even daily routine activities.

[P2] "… It depends on so many things, I might get some cleaning done if I wake up earlier, or be in a hurry and just have time to get dressed and leave, or even leave earlier than usual to run some tasks"

Three participants (P1,P2,P5) were found to track such routine activities to uncover the impact of their variability. P1, for instance, engaged with his tracker when arriving at work – to check for variations in his step count, due to a detour in his daily commute to work (see Figure 5). Another two (P3,10) were found to engage with their tracker when anticipating variations to a routine activity. For instance, P3, a primary teacher, engaged with the tracker before and after her class, to measure the physical activity gained while students performed an exam. These sessions accounted for one fifth of participants' sessions (median=20%, SD=11-20%), indicating that routines are subject to frequent change (as indicated in [17]), with individuals searching for information towards *repairing*, *striving* or *expanding* routines affected by change:

[P3] "I was wondering how many steps I'd get in the class, with the exam going on… checking it before (the start of the class), helped me set a baseline… to come back and see how many steps I gained"

Non-routine activities also led to such strategic engagement with the tracker, geared towards learning. P12 checked his tracker before and after taking a hike, while P6 used the same strategy to track his steps while helping a friend with moving his furniture to his new house.

[P6] "... I was curious to see how much we walked and decided to make some bets (…) surprisingly, we had more than 10,000 steps! My guess wasn't even close"

Figure 5. Trackers were used to uncover variations in routine activities. Examples ranged from checking how many extra steps were gained after taking part in a longer-than-habitual workout (left), and during a detour to work (right).

Glancing at physical activity and time concurrently

We noticed that participants would often glance at the time and their physical activity levels concurrently, or very closely after each other. Approximately half of each participants' engagements with physical activity feedback on Fitbit HR followed time checking (median=44%, SD=25-56%), for at least 2 seconds (median=3.2, SD=2.7-3.7). Participants often commented on the impulsiveness of their engagement with physical activity feedback.

[P5] "… it happens again and again. I check the time, but get dragged to my step count or the distance…it's a click away"

Moreover, we found that 20% of the sessions in which HR users engaged with physical activity feedback (N=65) were either preceded (9 of 13) or followed (4 of 13), within the following 15 seconds, by a usage session in which they checked the time.

We noticed three, interweaving reasons which led users to combine time and physical activity feedback within single or shortly separated sessions: a) assessing the attainability of goals, b) planning future activity, and c) reflecting on activity levels.

Assessing the attainability of goals

Time-checking served to estimate how likely one was to meet a daily goal given the distance walked so far. As one participant noted:

[P1] "… it's not only about how many steps I've gained, but also how much time it took me to get there… and how much (time) I still have to complete my goal"

In this respect, users developed a strategy to counteract the inefficiency of the tracker. While the tracker merely provided descriptive data (i.e., how much they had walked by that point in time), users desired normative data (i.e., "is this good enough?"). During the interviews, we realized that users often had a strong awareness of how much they should have walked by a given point in time in order to meet their goal. For instance, P7 noted:

"On a normal day, I should have 7,000 steps by mid-day… 10,000 around 5 [pm], and 12,000 around 8 [p.m]"

Further, users would also often think about their upcoming plans to estimate the likelihood of meeting their goal.

[P5] "I was going to spend the whole day sitting on my butt, so I knew it was a long shot [reaching step count goal]. There was little space to make up for it, having to pick up the kids from school, cooking, cleaning"

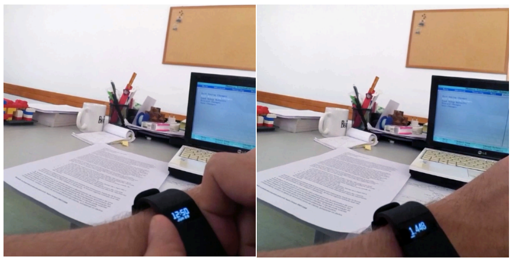

Figure 6. Users often combined PA feedback with time checking to estimate the likelihood of meeting goals. While trackers merely provided descriptive data (i.e., how much was walked by a point in time), users desired normative data (i.e., is this good enough?).

Planning future activity

Besides estimating the likelihood of meeting their goal, users would often use interaction to plan future activity, accounting for half of the sessions in which each participant combined time and physical activity checking (median=50%, IQR=25-67%). Such plans would vary in their temporal proximity and extend, from small detours in the near future such as grabbing some water while waiting for the printer, to ones more significant and distant in time, such as reaching 1000 steps in the next hour, or planning a visit to the gym for the evening, *e.g.*:

[P1] "I try getting out of the office on my afternoon breaks… to make up for heptic mornings. If my steps are low I'll say to myself 'lets try to get 1000 steps done in the next hour'… other days 200 will be enough. It depends on how I'm doing so far"

[P10] "I can't - and don't want to stop work or being with my family to walk every time my step is lower than I expect (...) it does, however, get me thinking about how I can get some steps in (...) a longer walk home, taking a dog for a walk after work."

While in some cases these plans were followed strictly (e.g., meeting one's 3000-step goal before getting off the treadmill), in most of the cases, the plans were flexible and responsive to the variability of users' routines.

[P10] "my rule of thumb is to have 1000 steps for every hour spent at work, but you never know, it might be a busier day (…) I'll tell myself 'ok, let's take a 10-minute break' and go for a walk if I see I'm falling behind"

In those circumstances, users would often adjust their expectations, or form new plans, when the opportunity arose in order to fulfill their plan. For instance, P2 was not near fulfilling her plan as she waited for her friends, she checked the time, and she realized she had 10 minutes for a short walk before her friends were due to arrive.

Figure 7. Users maintained an awareness of their activity levels and formed *micro-plans* such as grabbing water while waiting for the printer, or walking 1000 steps in the next hour.

Reflecting on activity levels

Participants also combined time and physical activity feedback to reflect about points in time, or activities, laying in the past – such as activity levels over a day, or within previously performed activities (e.g. looking back at the steps one has performed while at home [P4], or at the gym [P5]). P5, for instance, engaged with his tracker - while on vacations, to reflect on the amount of physical activity he had managed to include in his morning, and how this compared to his typical morning activity levels.

[P2] "… 5000 steps by 10am! with the vacations going on, I had no clue of how many steps to expect… it got me thinking of where I got them from. Probably from all the walking around the house or the late dancing"

Others engaged with time and physical activity feedback data to assess their performance in recently performed activities. P5, for instance, checked his step count and time before and after a jog, reflecting on the amount of steps gained within that timeframe: "*3km in 20 minutes. pretty slow today*".

Mitigating waiting time and alleviating boredom

Participants were found to engage with their physical activity feedback during "dead times" [40, pg. 9] - such as commuting via public transport, or waiting for an upcoming activity, accounting for approximately 10% of the sessions in which physical activity feedback was checked (median=11%, IQR=3-24%). "Dead times", as considered by Perry et al., refer to the time that occurs between tasks and activities, in which participants have little control over the resources they had to hand. P3, for instance, was found to check her step count while waiting for a bus. P1 checked his step count and time while waiting for a printing service.

[P1] "I try to get 12,500 steps before leaving work… there are moments, here and there where I try to get some steps if I see myself really behind… like grabbing some water while the printer is working, but I'll only bother if I'm more than 2000 or 3000 steps away"

P2 engaged with his tracker while waiting for a client meeting.

[P2] "I played around with it while I was waiting. (...) it helped kill the time, I didn't have much to do (…) I rather spend my time checking and thinking of my health then checking what other people are doing on Facebook."

Others were found to engage in tracker checking to alleviate boredom. P9 checked his tracker while in class; P3 while taking part in a meeting.

Figure 8. Participants were found to turn to their trackers to help mitigate waiting time . While engaging shortly with trackers, these led to longer-lasting refletions on one's data.

Contrary to the user-initiated engagement with physical activity data during "dead times", users often commented upon *engagement anxiety* during cognitively demanding tasks. Engagement with physical activity data was seen as an activity competing for resources – not necessarily due to the time spent within usage sessions, but due to the potential outcome of checking, such as feeling pressured or enforced to keep up to a certain walking goal, disregarding their ongoing availability (e.g., [P9] "*it's keeps me asking if I've done enough, and pushing me to do more. it gets tiring after a while*"). In fact, two participants (P8, P9) noted that they avoid checking their trackers during activities that demanded their focus, such as computer programming and during meetings. All in all, these results highlight the variability of daily life and the need for systems that sense and adapt to users' context. On the one hand, trackers need to be more accommodating, accepting that not all moments are equal and be less demanding in situations that offer little opportunity for physical activity. On the other hand, trackers may leverage "dead times" into opportunities for engaging with one's data, learning from past behaviors, and even motivating short bursts of physical activity.

DISCUSSION

Our findings suggest specific design considerations, such as facilitating learning through glances, providing normative feedback on goal accomplishment, and facilitating micro-plans.

Facilitating learning through glances

Personal informatics literature has assumed two primary modes of interaction with respective systems, guided by our *reflective* versus our *impulsive* thinking and decision making systems, respectively [48]. The reflective mode was long assumed to be the dominant mode of interaction: people would collect data, then explore and review them in retrospect (i.e., days, weeks), to identify patterns in their behaviors and plan alternative future courses of action. As such, most of today's personal informatics tools support this process through the high-level aggregation of data over time (e.g. step count over a week or a month) and considerable amount of efforts have focused on the design of personal visualizations to support reflection, interpretation and reminiscing based on data [15].

In contrast, the *impulsive* mode assumes people to process data and take action quickly and unconsciously. From a design perspective, the shifts emphasis from supporting learning to supporting the self-regulation of behavior [1, 22]. Recent research has highlighted the prevalence of the *impulsive* over the *reflective* mode. Gouveia et al. [21] found over 70% of usage sessions with a tracker to related to the so-called *"glances"* – brief, 5-second sessions where users checked their current activity levels with no further exploration. Glances were assumed to support the frequent regulation of behavior; people would estimate how likely they are to meet their activity goal by the end of the day, and introduce new actions when needed.

Our study, however, revealed that such glances may also serve towards learning. Participants were found to check their activity levels right before the start and after the end of a particular activity, such as household chores or commuting to work, with the goal of understanding their actual and potential contribution to their physical activity levels. Such interactions occurred frequently, accounting for approximately one third usage sessions. This points at the need to develop mechanisms to support learning through these frequent glances.

One could imagine intelligent systems – identifying significant correlations among users' data (as in [4]) or feedback that adapts to the particular context of use. But while current technology may help disentangling physical activity in relation to one's location or time of the day, we noticed that participants' inquiries were often simple, yet more precise. More than wanting to know how much they had walked at a certain venue (e.g. work), they wanted to gain specific insights into performed activities (e.g. how much did I walk while doing household chores?). Inferring the start and end time of such activities would be difficult given today's technology. We thus suggest a semi-automated approach, where the tracker enables users to quickly mark moments such as the start and end time of an activity, group these moments into an activity, and support users in labeling, annotating, inquiring into, and comparing different activities.

Providing normative feedback on goal accomplishment

We found that users often check time in conjunction with physical activity to estimate how likely they were to meet their daily goal given the distance walked so far. In this respect, users developed a strategy to counteract the inefficiency of the tracker. While the tracker merely provided descriptive data (i.e., how much they had walked by that point in time), users desired normative data (i.e., "is this good enough?"). This inefficiency of activity trackers had been previously noted by using the example of Fitbit Flex's wristband which features five LEDs that illuminate for each 20% of a daily walking goal achieved. As mentioned, "*even this seemingly simple display requires some quite difficult projections, if one wants to use it for immediate self-regulation*" [22, p.146]. As a solution, they proposed *Normly*, which employs a large database of other people's walking trajectories over the course of a day, and compares the distance one has walked at a given time in a day to that walked by others having the same goal, at the same time in the day. Thus, at each moment a user engages with *Normly*, she receives simple, normative feedback – that she is either doing better or worse than others, at this specific moment.

We contribute an additional concept aimed at providing normative data on goal accomplishment. *Predicto* leverages on users' aggregated accounts of data to forecast the likelihood of goal completion. *Predicto* draws inspiration from prediction markets [5], which leverage on aggregate information to produce predictions about future events (e.g. a political candidate's re-election, the victory of a sports team). The likelihood would be estimated on a number of conditions, such the number of consecutive days in which a user has reached his goals, or the upcoming plans for a day. *Predicto* takes into account the variability of routines [17] by constantly updated predictions (e.g. skipping a habitual trip to the gym lowers the chances of reaching a goal).

Facilitating micro-plans

Our study revealed that users often formed *micro-plans* for the immediate future, such as reaching 1000 steps in the next hour. While the positive effects of proximal goals on motivation and performance have been highlighted by empirical research in goal setting and acknowledged by

Goal Setting Theory [34], the majority of today's trackers, rather without much thought, adopt a daily step goal.

One might wonder about how technology could further support practices such as the one identified here – that of *micro-plans*, short courses of action planned within daily routines geared towards meeting one's levels of physical activity. Prior work has explored ways to make goal-setting more agile to accommodate the variability of daily life. Munson et al. [38] for instance, proposed the idea of *secondary goals* as fallbacks in days of reduced physical activity, while Konstanti and Karapanos proposed the idea of *micro-updates* – daily step goals that expire at the end of the day, inspired by users' practices (e.g., "*when I know I will be seated a lot e.g. long car trip, I adjust my goal downwards").* However, no work has explored how trackers may further support individuals in planning their days in more detail. Given our findings and Goal Setting Theory, such mechanisms are expected to positively impact users' motivation and performance.

However, a crucial interaction design challenge is how to support this practice while maintaining users' flexibility in adjusting their plans? We found that in most of the cases, users treated those as flexible and responsive to the variability of their routines. They often adjusted their expectations, formed new plans, or enforced nested actions when the opportunity arose in order to fulfill their plan.

Figure 9. *Mikro* compares one's progress to a micro-goal (outer ring) to the time remaining to complete a goal (inner ring) . *Mikro* also allows users to make short adjustments to goals.

We detail below two concepts that aim at facilitating the formation and execution of micro-plans. The first one, *Mikro* (see Figure 9), enables users to set a micro-plan manually. The user selects the number of steps and the duration of a micro-plan. *Mikro* then displays the remaining steps and time for the micro-plan along with the total number steps walked in the day. *Mikro* further supports flexible goals. In a similar way alarms allow for 5 extra minutes of snoozing, users are allowed to make small adjustments their goals at any point of their day. The second one, *Mikromoves*, builds upon the commercial application Moves (see [21] for a similar approach), that segments physical activity over the different locations one visits in the course of the day. *Mikromoves* automatically sets a micro-plan for each new location the user visits (e.g., Welcome to *Work*, let's walk 2300 steps over the next 8 hours). The distance and duration are estimated based on past visits at the location, while the user may adapt those values, which also contributes to training *Mikromoves* prediction algorithm.

LESSONS LEARNT AND LIMITATIONS

This paper presented an *in vivo* study of activity tracking. Through the use of video methods, the study took a close look at a number of practices that surround the use of activity trackers in daily life. Most notably, we found the use of these devices to be strongly driven by reflection and *learning-in-action*, contrasting the traditional view that learning is one of deep, retrospective exploration of data.

One should note a number of limitations in this study. First, this is a study of the practices of successful adopters of activity trackers. We chose to do so in order to shed light into the practices that we need to support in order to sustain prolonged use [see also 18]. Yet, one has to take into account that not everyone who tries a wearable activity monitor continues to use it in the long term [21,30]. Understanding failed practices is also a much needed endeavor. Secondly, our study involved only twelve participants and only two days of their engagement with the tracker. We believe these are direct outcomes of some of the challenges of the adopted method. In particular, participation may be hindered due to surveillance and privacy concerns of using a wearable camera [39] and is likely to produce increased discomfort if an extended study duration is required.

In closing, our study took a step forward towards understanding how tracker use is enmeshed within everyday life, and how these devices could be better designed to support long-term use. Activity tracking, as a practice, is diversified, dependent upon and threaded into what goes on around us. Following these developments is difficult. Yet, as these devices become ever more central to the ongoing discourse on behavior change and patient-driven healthcare, a richer understanding of the lived dynamics of activity tracking trackers is crucial.

AKNOWLEDGMENTS

This research was supported by the Portuguese Foundation of Science and Technology grant SFRH/BD/89796/2012.

REFERENCES

1. Alexander T. Adams, Jean Costa, Malte F. Jung, and Tanzeem Choudhury. (2015). Mindless computing: designing technologies to subtly influence behavior. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing* (UbiComp'15), pp. 719–730 https://doi.org/10.1145/2750858.2805843

2. Jeremy D. Akers, Rachel A. Cornett, Jyoti S. Savla, Kevin P. Davy, and Brenda M. Davy. (2012). Daily self- monitoring of body weight, step count, fruit/vegetable intake, and water consumption: a feasible and effective long-term weight loss maintenance approach. *Journal of the Academy of Nutrition and Dietetics*,*112*(5), pp. 685-692. https://doi.org/10.1016/j.jand.2012.01.022
3. Raymond C. Baker, and Daniel S. Kirschenbaum. (1993). Self-monitoring may be necessary for successful weight control. *Behavior Therapy*, *24*(3), pp. 377-394. https://doi.org/10.1016/S0005-7894(05)80212-6
4. Frank Bentley, Konrad Tollmar, Peter Stephenson, Laura Levy, Brian Jones, Scott Robertson, Ed Price, Richard Catrambone, and Jeff Wilson. (2013). Health Mashups: Presenting Statistical Patterns between Wellbeing Data and Context in Natural Language to Promote Behavior Change. *Transactions on Computer-Human Interactions* (ToCHI'13), pp. 1–13. http://dx.doi.org/10.1145/2503823
5. Joyce E. Berg, and Thomas A. Rietz. (2003). Prediction markets as decision support systems. *Information systems frontiers*, *5*(1), pp. 79-93. https://doi.org/10.1023/A:1022002107255
6. Virginia Braun, and Victoria Clarke. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, *3*(2), pp. 77-101. http://dx.doi.org/10.1191/1478088706qp063oa
7. Barry Brown, Moira McGregor, and Eric Laurier. (2013). iPhone in vivo: video analysis of mobile device use. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI'13), pp. 1031-1040. https://doi.org/10.1145/2470654.2466132
8. Lora E. Burke, Jing Wang, and Mary Ann Sevick. (2011). Self-monitoring in weight loss: a systematic review of the literature. *Journal of the American Dietetic Association*, *111*(1), pp. 92-102. https://doi.org/10.1016/j.jada.2010.10.008
9. Robert A. Carels, Lynn A. Darby, Sofia Rydin, Olivia M. Douglass, Holly M. Cacciapaglia, and William H. O'Brien. (2005). The relationship between self-monitoring, outcome expectancies, difficulties with eating and exercise, and physical activity and weight loss treatment outcomes. *Annals of Behavioral Medicine*, *30*(3), pp. 182-190. https://doi.org/10.1207/s15324796abm3003_2
10. Eun Kyoung Choe, Nicole B. Lee, Bongshin Lee, Wanda Pratt, and Julie A. Kientz. (2014). Understanding quantified-selfers' practices in collecting and exploring personal data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI'14), pp. 1143-1152. http://doi.acm.org/10.1145/2556288.2557372
11. Chia-Fang Chung, Nanna Gorm, Irina A. Shklovski, and Sean Munson. (2017). Finding the Right Fit: Understanding Health Tracking in Workplace Wellness Programs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI'17), pp. 4875-4886. https://doi.org/10.1145/3025453.3025510
12. James Clawson, Jessica A. Pater, Andrew D. Miller, Elizabeth D. Mynatt, and Lena Mamykina. (2015). No longer wearing: investigating the abandonment of personal health-tracking technologies on craigslist. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing* (UbiComp'15)*,* pp. 647-658. https://doi.org/10.1145/2750858.2807554
13. Sunny Consolvo, Katherine Everitt, Ian Smith, and James A. Landay. (2006). Design requirements for technologies that encourage physical activity. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI'06), pp. 457-466. https://doi.org/10.1145/1124772.1124840
14. Sunny Consolvo, David W. McDonald, Tammy Toscos, Mike Y. Chen, Jon Froehlich, Beverly Harrison, Predrag Klasnja, Anthony LaMarca, Louis LeGrand, Ryan Libby, Ian Smith, and James A. Landay. (2008). Activity Sensing in the Wild: a Field Trial of UbiFit Garden. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI'18), 1797–1806. http://doi.org/fj37wd
15. Daniel Epstein, Felicia Cordeiro, Elizabeth Bales, James Fogarty, and Sean Munson. (2014). Taming data complexity in lifelogs: exploring visual cuts of personal informatics data. In *Proceedings of the ACM Conference on Designing Interactive Systems* (DIS'14), pp. 667-676. https://doi.org/10.1145/2598510.2598558
16. Daniel A. Epstein, An Ping, James Fogarty, and Sean A. Munson. (2015). A lived informatics model of personal informatics. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp'15),* pp. 731-742. https://doi.org/10.1145/2750858.2804250
17. Martha S. Feldman. (2000). Organizational Routines as a Source of Continuous Change. *Organization Science*, 11(6), pp. 611 - 629. https://doi.org/10.1287/orsc.11.6.611.12529
18. Thomas Fritz, Elaine M. Huang, Gail C. Murphy, and Thomas Zimmermann. (2014). Persuasive technology in the real world: a study of long-term use of activity sensing devices for fitness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing*

Systems (CHI'14), pp. 487- 496. https://doi.org/10.1145/2556288.2557383

19. Get Activity Intraday Time Series. (n.d.). Retrieved 19 September, from https://dev.fitbit.com/reference/web-api/activity/ - get-activity-intraday-time-series
20. Rúben Gouveia and Evangelos Karapanos. (2013). Footprint tracker: supporting diary studies with lifelogging. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI'13), pp. 2921-2930. https://doi.org/10.1145/2470654.2481405
21. Rúben Gouveia, Evangelos Karapanos, and Marc Hassenzahl. (2015). How do we engage with activity trackers?: a longitudinal study of Habito. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing* (UbiComp'15), pp. 1305-1316. https://doi.org/10.1145/2750858.2804290
22. Rúben Gouveia, Fábio Pereira, Evangelos Karapanos, Sean A. Munson, and Marc Hassenzahl. (2016). Exploring the design space of glanceable feedback for physical activity trackers. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing* (UbiComp'16), pp. 144-155. https://doi.org/10.1145/2971648.2971754
23. Nanna Gorm and Irina Shklovski. (2016). Steps, Choices and Moral Accounting: Observations from a Step-Counting Campaign in the Workplace. In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work & Social Computing* (CSCW'16), pp. 148-159. https://doi.org/10.1145/2818048.2819944
24. Nanna Gorm and Irina Shklovski. (2017). Participant Driven Photo Elicitation for Understanding Activity Tracking: Benefits and Limitations. In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work & Social Computing* (CSCW'17), pp. 1350-1361. https://doi.org/10.1145/2998181.2998214
25. Daniel Harrison, Paul Marshall, Nadia Bianchi-Berthouze, and Jon Bird. (2015). Activity tracking: barriers, workarounds and customisation. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp'15),* pp. 617-621. https://doi.org/10.1145/2750858.2805832
26. Christian Heath, Paul Luff, P and Marcus S. Svensson. (2007). Video and qualitative research: analysing medical practice and interaction. *Medical education*, *41*(1), 109-116. http://dx.doi.org/10.1111/j.1365-2929.2006.02641.x
27. Daniel Kahneman, Alan B. Krueger, David A. Schkade, Norbert Schwarz and Arthur A. Stone. (2004). A survey method for characterizing daily life experience: The day reconstruction method. *Science*, 306(5702), pp. 1776-1780. https://dx.doi.org/10.1126/science.1103572
28. Evangelos Karapanos, Rúben Gouveia, Marc Hassenzahl, and Jodi Forlizzi. (2016). Wellbeing in the making: peoples' experiences with wearable activity trackers. *Psychology of well-being*, *6*(1), pp. 1-17. https://dx.doi.org/10.1186%2Fs13612-016-0042-6
29. Predrag Klasnja, Sunny Consolvo, and Wanda Pratt. (2011). How to evaluate technologies for health behavior change in HCI research. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI'11), pp. 3063-3072. https://doi.org/10.1145/1978942.1979396
30. Amanda Lazar, Christian Koehler, Joshua Tanenbaum, and David H. Nguyen. (2015). Why we use and abandon smart devices. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp'15),* pp. 635-646. https://doi.org/10.1145/2750858.2804288
31. Matthew L. Lee and Anind K. Dey. (2007). Providing good memory cues for people with episodic memory impairment. *In Proceedings of the ACM SIGACCESS conference on Computers and accessibility* (Assets'07), pp.131-138. http://dx.doi.org/10.1145/1296843.1296867
32. Ian Li, Anind Dey, and Jodi Forlizzi. (2010). A stage-based model of personal informatics systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI'10), pp. 557-566). https://doi.org/10.1145/1753326.1753409
33. James J. Lin, Lena Mamykina, Silvia Lindtner, Gregory Delajoux, and Henry B. Strub. (2006). Fish'n'Steps: Encouraging physical activity with an interactive computer game. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp'15), pp.* 261-278. http://dx.doi.org/10.1007/11853565_16
34. Edwin A. Locke and Gary P. Latham. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. American psychologist, 57(9), pp. 705-715. http://dx.doi.org/10.1037/0003-066X.57.9.705
35. Donald McMillan, Moira McGregor, and Barry Brown. (2015). From in the wild to in vivo: Video Analysis of Mobile Device Use. In *Proceedings of the ACM International Conference on Human-Computer Interaction with Mobile Devices and Services* (MobileHCI'15), pp. 494-503. ACM. https://doi.org/10.1145/2785830.2785883
36. Jochen Meyer, Merlin Wasmann, Wilko Heuten, Abdallah El Ali, and Susanne C.J. Boll. (2017). Identification and Classification of Usage Patterns in Long-Term Activity Tracking. In *Proceedings of the*

SIGCHI Conference on Human Factors in Computing Systems (CHI'17)*,* pp. 667-678. https://doi.org/10.1145/3025453.3025690

37. Lorenza Mondada. (2008). Using video for a sequential and multimodal analysis of social interaction: Videotaping institutional telephone calls. In *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research,* 9(3). http://dx.doi.org/10.17169/fqs-9.3.1161
38. Sean A. Munson and Sunny Consolvo (2012). Exploring goal-setting, rewards, self-monitoring, and sharing to motivate physical activity. In *Pervasive computing technologies for healthcare (PervasiveHealth'12),* pp. 25-32. https://doi.org/10.4108/icst.pervasivehealth.2012.248691
39. David H. Nguyen, Gabriela Marcu, Gillian R. Hayes, Khai N. Truong, James Scott, Marc Langheinrich, and Christof Roduner. (2009). Encountering SenseCam: personal recording technologies in everyday life. In *Proceedings of the International Conference on Ubiquitous Computing* (UbiComp '09), pp. 165-174. http://dx.doi.org/10.1145/1620545.1620571
40. Misha Patel and Aisling Ann O'Kane. (2015). Contextual influences on the use and non-use of digital technology while exercising at the gym. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI'15), pp. 2923-2932. https://doi.org/10.1145/2702123.2702384
41. Mark Perry, Kenton O'hara, Abigail Sellen, Barry Brown, and Richard Harper. (2001). Dealing with mobility: understanding access anytime, anywhere. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *8*(4), pp. 323-347. http://dx.doi.org/10.1145/504704.504707
42. Stefania Pizza, Barry Brown, Donald McMillan, and Airi Lampinen. (2016). Smartwatch in vivo. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI'16), pp. 5456-5469). https://doi.org/10.1145/2858036.2858522
43. James O. Prochaska, Wayne F. Velicer (1997). The transtheoretical model of health behavior change. *American journal of health promotion*, *12*(1), pp. 38-48. https://doi.org/10.4278/0890-1171-12.1.38
44. PWC Consumer Intelligence. (2016). The Wearable Life 2.0: Connected living in a wearable world. 55. https://www.pwc.com/us/en/industry/entertainment-media/assets/pwc-cis-wearables.pdf
45. David Riley. Owner of Activity Trackers, Smartwatches Expected to Jump at Least 4 Percent this Holiday Season. *The NPD Group*. https://www.npd.com/wps/portal/npd/us/news/press-releases/2015/ownership-of-activity-trackers-smartwatches-expected-to-jump-at-least-4-percent-this-holiday-season-according-to-npd/
46. Yvonne Rogers. (2011). Interaction design gone wild: striving for wild theory. *Interactions* 18(4), 58.
47. John Rooksby, Mattias Rost, Alistair Morrison, and Matthew Chalmers Chalmers. (2014). Personal Tracking as Lived Informatics. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'14), pp.1163–1172. http://doi.acm.org/10.1145/2556288.2557039
48. Fritz Strack and Roland Deutsch. (2004). Reflective and impulsive determinants of social behavior. *Personality and social psychology review*, *8*(3), pp. 220-247. https://doi.org/10.1207%2Fs15327957pspr0803_1
49. Catrine Tudor-Locke, C. and David R. Bassett Jr. (2004). How many steps/day are enough? *Sports medicine*, 34(1), pp.1-8. https://doi.org/10.2165/00007256-200434010-00001
50. Adam G. Tsai and Thomas A. Wadden. (2005). Systematic review: an evaluation of major commercial weight loss programs in the United States. *Annals of internal medicine*, *142*(1), 56-66. http://dx.doi.org/10.7326/0003-4819-142-1-200501040-00012
51. Muhammad U. Warraich. (2016). Wellness Routines with Wearable Activity Trackers: A Systematic Review. In *Proceedings of Mediterranean Conference on Information Systems,* (MCIS'16). http://aisel.aisnet.org/mcis2016/35

4.3 – Discussion of results and conclusion

Our study revealed some of the complexities of activity tracking, while extending some of the underlying ideas of the use of these devices. Within this section, I discuss some of our findings in line with Personal Informatics (and activity tracking) literature, and explore some of the implications of these findings for the design of activity tracking devices: namely, how activity tracking devices should be designed to supporting just-in-time, contextualized moments of learning.

Supporting contextualized moments of learning

In a second attempt to inquire into the daily use of activity trackers, we turned to video methods and direct, in-situ observations of users' interactions to understand what motivates people to engage with activity trackers in their everyday life, and how the surrounding environment shapes the use of these devices. Most notably, we found the use of activity trackers to be strongly driven by *learning-in-action*. Participants interacted with their trackers to learn about ongoing activities – such as how much they were walking while cleaning their house, or while commuting to work. More than mediators of change, trackers, in these cases, were used to "simply" mediate learning.

These insights contrast with traditional views of how activity tracking tools are used. For example, one of the most cited models of Personal Informatics use was proposed by Li and colleagues [51]. This model describes five stages that people go transition through when using Personal Informatics tools: (a) *preparation*, where people decide what personal information they want to track; (b) *collection*, where they start collecting data about themselves; (c) *integration*, where the data is prepared to be reflected upon; (d) *reflection*, where people retrospect on their behaviors*;* and (e) *action*, where decisions are made about how to change a certain behavior. By placing action as an end-step, this model emphasizes behavior change as a natural outcome of the use of these tools. Self-knowledge and self-reflection, under this vision, are simply a means to progress towards a specific behavioral goal (e.g. walking more, or losing weight).

Our study, however, hints at an alternative way of using activity trackers: one in which learning occupies a central role (and end-goal) in the use of these devices. This aligns with recent talks about the need for Personal Informatics systems to move beyond behavior change and self-improvement and to consider supporting discoveries of the "self" (e.g. [75], [22]). In their article "*Know Thyself: A Theory of the Self for Personal Informatics*" [75], Rapp and Tirassa propose shifting the traditional framing of personal informatics tools from a subclass of behavior change technologies, to a class of systems that provides individuals with better knowledge of themselves. The "*self*", according to Rapp and Tirassa, should be considered as more than a "static, frozen entity that can be revealed by looking at data" (p. 344). Instead, it should be considered as an entity with its own personal needs, that is constantly evolving an being re-created through its interactions with the surrounding world. Personal Informatics tools, under this view, should help people learn about their everyday lives and build better understandings of themselves.

As suggested in our paper, one way of supporting self-discovery and learning with trackers is to enhance the data collected by these devices with contextual details from the external environment. However, more than simply showing a user how active he has been within a certain location (which might not be the most effective way of engaging a user in learning, as seen in our study with *Habito* and in [52]), contextual details should be tied to the activity the user is (or was) performing when checking (or reviewing) his data. For example, a teacher who wants to explore how active she is while lecturing a class should be given the opportunity to find answers to questions like, How does the duration of the class and number of students influence how physically active I am? How does this compare to the other classes that I teach? How was I feeling at that time? What tasks was I assigning the students with? What was happening to me during that day? While the first questions might be addressed by answered easily through today's technology (e.g. by tracking the duration of the class and the number of steps taken within), the last lead to suggestions of semi-automated tracking approaches, where users can add label, annotate and inquire into their internal states, like emotions, thoughts, and intentions.

These activities could further be memorized by activity tracking systems, and used to present users with "*just-in-time*" feedback to motivate self-reflection, and even behavior change [46].

Supporting brief, just-in-time, uses of activity trackers

Personal Informatics literature has long assumed that people gain insights as the result of deep and careful reflections on their behaviors [51]. As such, most of today's personal informatics tools support this process through the high-level aggregation of data over time (e.g. step count over a week or a month) and considerable amount of efforts have focused on the design of personal visualizations to support reflection, interpretation and reminiscing (e.g. [25]).

Considerably less effort has been placed in supporting quick opportunities for learning, and action as they arise in people's everyday lives. Our studies, however, showed that the use of trackers might actually be dominated by this type of use. First, in our study with *Habito*, we found over 70% of the usage of *Habito* to be driven by glances – brief, 5-second sessions where individuals check ongoing activity levels with no further interaction. Glances were assumed to support the frequent regulation of behavior; people would estimate how likely they were to meet their goals, and introduce action when needed [32]. Our *Invivo* study, however, found these quick moments of interaction with trackers to be also used towards learning and self-discovery. Users checked their trackers to learn how active they had been while performing a certain activity – such as while teaching a class, or cleaning a house.

These insights raise questions towards how to develop interfaces that support brief moments of learning and action. While we are in no way trying to undermine the importance of deep explorations of data, our results point to the importance of engaging in conversations of how activity tracking devices can support brief moments of action and reflection as they arise within people's everyday lives. In the next chapter of this dissertation, we explore ways in which activity trackers can be designed to support these moments of everyday, quick learning and action.

Chapter 5.

Exploring the Design Space of Glanceable Feedback for Physical Activity Trackers

5.1 – Introduction

Our *Habito* and *Invivo* studies highlighted the brief, situated nature of activity tracking use. In both studies, activity tracker use was driven by *glances* – brief, no longer than 5-second sessions where individuals checked their ongoing activity levels with no further interactions. This form of interaction accounted for 70% and 75% of the usage sessions in our *Habito* and *Invivo* studies, respectively. *Glances* supported frequent regulations of behavior, as well as moments of learning and self-discovery. People would estimate how likely they were to meet their goals and introduced action when needed, and also engaged in quick moments of learning and self-discovery – such as learning how active one was while cleaning a house, or how many steps were gained during routine commutes to work.

In this chapter, I focus on a challenge highlighted by these results: if *glances* and quick interactions seem to be the predominant way in which users regulate and learn about their behaviors, then how can we design feedback interfaces that support this type of interactions?

In studies of Personal Informatics systems, a number of researchers have pointed towards the value of glanceable feedback and quick interactions (e.g. [20],[62]). For instance, Consolvo et al. [17] describe three locations where self-monitoring tools commonly provide feedback: (1) in-application feedback (i.e. portraying feedback within an application, such as how we did with our *Habito* study); (2) on application websites (i.e. displaying feedback in cloud-based services, such as websites) and; (3) glanceable displays. Glanceable displays, as defined by Consolvo and colleagues, refer to *quickly* and *easily* consumable feedback that is displayed outside of applications, in locations where users will frequently see, whether they go to applications or not. One well cited example is the *UbiFit Garden* [20], a mobile phone application intended to encourage regular physical activity. A key component of the *Ubifit Garden* application is the feedback about physical activity that is displayed on the background of users' mobile phone. In this way, users have frequent encounters with their physical activity behaviors – either when checking their phone to make a phone call, check the time, send a message, or use the phone for any other purpose. A 3-month field study of *Ubifit Garden* demonstrated the power of this approach. This always-available display was found to act as a persistent reminder to stay engaged and committed to physical activity. Since then, a number of additional researchers have investigated then effectiveness of glanceable displays.

Glanceable displays have the potential for encouraging healthy behaviors. Recent research has found people to increase their daily water intake [1], engaging in regular breaks from sedentary behaviors [40] and increasing awareness of physical activity levels [53] while being presented with glanceable feedback on their behaviors. However, research on glanceable displays is limited in many ways. On the one hand, glanceable displays have been the least explored form of feedback within personal informatics and health informatics literature. Instead, a considerable amount of effort has focused on the design and use of personal visualizations to support reflection, interpretation and reminiscing (e.g. [25], [80]). On the other, most research on glanceable displays has focused on uncovering the overall effects of these displays towards behavior change. Considerably less is known as to how these displays should be designed, and the effects of particular design choices. In particular, literature lacks detailed inquiries into "the design space of glanceable behavioral feedback, guidelines for what makes feedback glanceable, and an understanding of the effects of different glanceable feedback displays" [35].

In this chapter, I present a study that explores how glanceable displays can be designed to influence people's behaviors. This study contributes with an exploration of the underlying characteristics for the design of glanceable behavioral feedback interfaces. Further, through a field study, we deploy four different glanceable interfaces in the wild to better understand how different types of glanceable feedback affect users' engagement and physical activity. We present the results of this study, and insights on the effectiveness of these interfaces for encouraging physical activity.

5.2 – Exploring the Design Space of Glanceable Feedback for Physical Activity Trackers

This article is organized in three main sections. The first section presents an exploration of the design space of glanceable feedback for activity trackers, which resulted in the proposal of six underlying design qualities for these interfaces: (1) being *abstract*, (2) *integrating with existing activities*, (3) *supporting comparisons to targets and norms*, (4) being *actionable*, (5) having the capacity to lead to *checking habits* and (6) acting as a *proxies to further engagement.*

The second section describes four concepts which we prototyped and deployed in a field study with twelve participants over 28 days: (1) *TickTock*, which portrayed periods in which one was physically active over the past hour; (2) *Normly*, which compared one's goal completion to that of others having a similar goal; (3) *Gardy*, which abstracted physical activity levels through a garden, blossoming as users progressed towards their goal, and; (4) *Goal Completion*, which portrayed one's progress towards a daily goal through a bar that fills up as a user reaches his goal. The former interface was used as a baseline, against we compared the remaining glanceable interfaces.

The third section presents our findings and discussion of our results. Our findings are split into two main sections: First, we provide an overall analysis of the overall engagement with each interface over the 28 days of study, as well as an analysis of users' physical activity levels while using each interface. This if followed by a deeper analysis of participants' experiences with each of the interfaces. Our paper concludes with a discussion of the effects of each interface, as well as an overall analysis on the effects of glanceable representations of physical activity feedback.

This article was published in Ubicomp 2016 [35] with co-authors Fábio Pereira, Dr. Evangelos Karapanos, Dr. Sean Munson and Prof. Dr. Marc Hassenzahl. The design exploration was conducted by the first three authors. Pereira and I led the field study, as well as the collection of data. Dr. Karapanos and I analyzed the data. The paper was written collaboratively with Dr. Karapanos, Dr. Munson and Prof. Dr. Hassenzahl.

Exploring the Design Space of Glanceable Feedback for Physical Activity Trackers

Rúben Gouveia[1], Fábio Pereira[1], Evangelos Karapanos[1,2], Sean A. Munson[3], Marc Hassenzahl[4]
[1]Madeira-ITI, [2]Cyprus University of Technology, [3]University of Washington, [4]Folkwang University of Arts
{rubahfgouveia, fpereirauma, e.karapanos}@gmail.com, smunson@uw.edu, marc.hassenzahl@folkwang-uni.de

ABSTRACT
Recent research reveals over 70% of the usage of physical activity trackers to be driven by *glances* – brief, 5-second sessions where individuals check ongoing activity levels with no further interaction. This raises a question as to how to best design glanceable behavioral feedback. We first set out to explore the design space of glanceable feedback in physical activity trackers, which resulted in 21 unique concepts and 6 design qualities: being *abstract*, *integrating with existing activities*, *supporting comparisons to targets and norms*, being *actionable*, having the capacity to lead to *checking habits* and to act as a *proxy to further engagement*. Second, we prototyped four of the concepts and deployed them in the wild to better understand how different types of glanceable behavioral feedback affect user engagement and physical activity. We found significant differences among the prototypes, all in all, highlighting the surprisingly strong effect glanceable feedback has on individuals' behaviors.

Author Keywords
Physical activity tracking; glanceable displays; behavioral feedback interfaces; personal informatics.

ACM Classification Keywords
H.5.2. User Interfaces: Evaluation/Methodology.

INTRODUCTION

People increasingly adopt technologies to track their everyday behavior [39]. Personal informatics tools rest on the assumption that people develop a better understanding of their habits through self-monitoring, which in turn promotes self-knowledge, reflection and ultimately change upon undesirable habits [28]. Examples are counting steps to increase levels of physical activity [11] or measuring water spent in the shower to reduce waste [17].

Since knowledge of existing behavioral patterns seems at the heart of self-tracking, according tools focus on the rich visualization and the deep exploration of personal data [14,

UbiComp '16, September 12-16, 2016, Heidelberg, Germany
© 2016 ACM. ISBN 978-1-4503-4461-6/16/09...$15.00
DOI: http://dx.doi.org/10.1145/2971648.2971754

Figure 1 – *TickTock* (left) and *Normly* (right), two of the concepts developed as watchfaces. *TickTock* portrays periods in which one was physically active over the past hour. *Normly* compares one's goal completion to that of others having a similar walking goal.

7]. This implies a certain way of using such tools. First people collect data, then explore and review summaries of longer periods in retrospect (i.e., days, weeks) to identify patterns and plan alternative future courses of action [16]. For example, some people use tools provided by their phone companies to analyze their monthly costs to pick the "best" tariff or to optimize own future usage behavior.

In addition to this rather analytical approach, people use self-tracking to monitor and regulate immediate behavior [8]. For example, somebody may have told Ruben that paced walking (e.g., 6 km/h) is a valuable opportunity to get a little more exercise throughout the day. Unfortunately, Ruben is a slow walker. To get into the habit, he measures his walking pace *while* walking home from work to keep up the speed. This scenario requires frequent feedback while actually being engaged in the activity of walking [8].

In the case of physical activity trackers, brief and frequent monitoring may in fact be the dominant mode of interaction. In a prior study [19], 70% of all interactions with an activity tracker were *glances* – brief, 5-second sessions where users checked their current activity levels with no further exploration or interaction.

While researchers have noted the value of glanceable feedback as a complement to the deeper and reflective analysis [11], research focusing specifically on glanceable activity feedback displays has been scarcer than research on deep, reflective feedback displays. In particular, literature lacks detailed inquiries into the design space of glanceable behavioral feedback, guidelines for what makes feedback glanceable, and an understanding of the effects of different glanceable feedback displays.

To provide a better understanding of glanceable behavioral feedback, we first explored the design space of glanceable feedback in the context of physical activity trackers. We created a total of 21 concepts and a total of 6 design qualities through an iterative ideation and reflection process. We argue that glanceable feedback for behavior change should *be abstract*, *integrate with existing activities*, *support comparisons to targets and norms*, *be actionable*, and have the capacity to lead to *checking habits* and act as a *proxy to further engagement*. Second, we prototyped four of the concepts and deployed them in the wild to better understand how different types of glanceable feedback affect users' engagement and physical activity.

RELATED WORK

So far, the importance of glanceable feedback in behavior change tools has been noted by a number of researchers. Ham and Midden [20] emphasized the persuasiveness of glanceable feedback since it requires minimal attention to be perceived and processed. Consolvo et al. [11] found individuals to increase long-term commitment to physical exercise when presented with glanceable feedback. Mullet and Sano [32] further argue that the frequent monitoring of behavior can lead to early correction of slips and relapse.

But what makes feedback especially glanceable? Consolvo et al. [12] define "glanceability" in terms of how *quickly* and *easily* feedback is able to convey information after one pays attention. To accomplish high glanceability, feedback should be "reduced to the essence through a process of simplification and abstraction" [32]. Feedback should provide "just enough" to be perceived and processed [30]. A further quality of glanceable feedback is its ability to be perceived at the periphery of one's attention [5]. Feedback should be "*working in the background while we attend to foreground activities ... [enabling people] to get the essence of the information with a quick visual glance"* [29].

Empirical studies have provided support for the effectiveness of glanceable feedback. Jafarinaimi et al. [23] developed *Breakaway*, a small human sculpture aimed at encouraging regular breaks from work. *Breakaway* mimicked its user's posture throughout the day. It was placed on the office desk, offering persistent, yet unobtrusive and quickly consumable feedback. A case study with a single participant showed the likelihood of taking a break from work to increase when the sculpture slouched. In addition, the participant commented on how easily *Breakaway* could be ignored, when busy. In this case, healthy sitting is a secondary task to be monitored and regulated throughout the day while actually completing primary, work-related tasks.

Another example is Consolvo et al.'s [11] *UbiFit Garden*, a mobile application designed to support overall physical activity by tracking users' physical activity, and presenting feedback on the background screen of mobile phones. In a comparative study, participants using *UbiFit Garden* had higher activity levels than participants without persistent feedback on behaviors. The always-available information on activity levels acted as a reminder to stay engaged and committed to the goal of increasing physical activity. Fortmann et al. [15] created *WaterJewel*, a wearable wrist bracelet to motivate users to maintain adequate hydration levels throughout the day. *WaterJewel* has eight LEDs, which light up when users progress towards their daily goal of water intake. Participants using *WaterJewel* were more likely to accomplish their goals for water intake than participants who received the information on their phones.

All in all, research suggests that presenting abstract, easily consumable information, at locations where the individual is likely to gaze frequently positively affects self-regulation of particular behaviors.

Yet, while the strengths of glanceable feedback have been recognized, previous literature has highlighted the need to explore the efficacy of different forms of glanceable feedback. In Consolvo et al.'s study [11], for example, men were more skeptical of the garden display than women, raising questions about the effectiveness of different stories told through feedback. Are some forms of glanceable feedback more effective compared to others? [12].

In the remainder of the paper, we present our design space exploration, which led to 21 concepts and 6 design qualities important for glanceable feedback, followed by an empirical exploration of the four prototyped concepts.

DESIGN SPACE EXPLORATION: CONCEPTS AND QUALITIES

Our first goal was to explore the design space of glanceable feedback for activity trackers. Since wrist-worn devices (e.g., smartwatches, wristbands) are the most glanced mobile feedback displays available [31, 35], we focused our exploration on smartwatch interfaces. As a technology, smartwatches allow for the widest variety of ways to present feedback in glanceable ways.

The design space exploration was performed by the first three authors. Starting with a design brief of 'glanceable watchfaces reflecting physical activity', we followed an iterative process of synthesis and analysis, whereby new ideas were compared to each other to reveal the underlying differences and qualities of glanceability, followed by new rounds of ideation aimed at further deepening the understanding of each emerging quality. Existing research prototypes (e.g., *UbiFit*) or commercial products (e.g., *Fitbit*) were often used as reference points during the analysis, while theoretical frameworks and constructs (e.g., Cialdini's [9] scarcity principle) often helped us elaborate on the design qualities. This process led to a total of twenty-

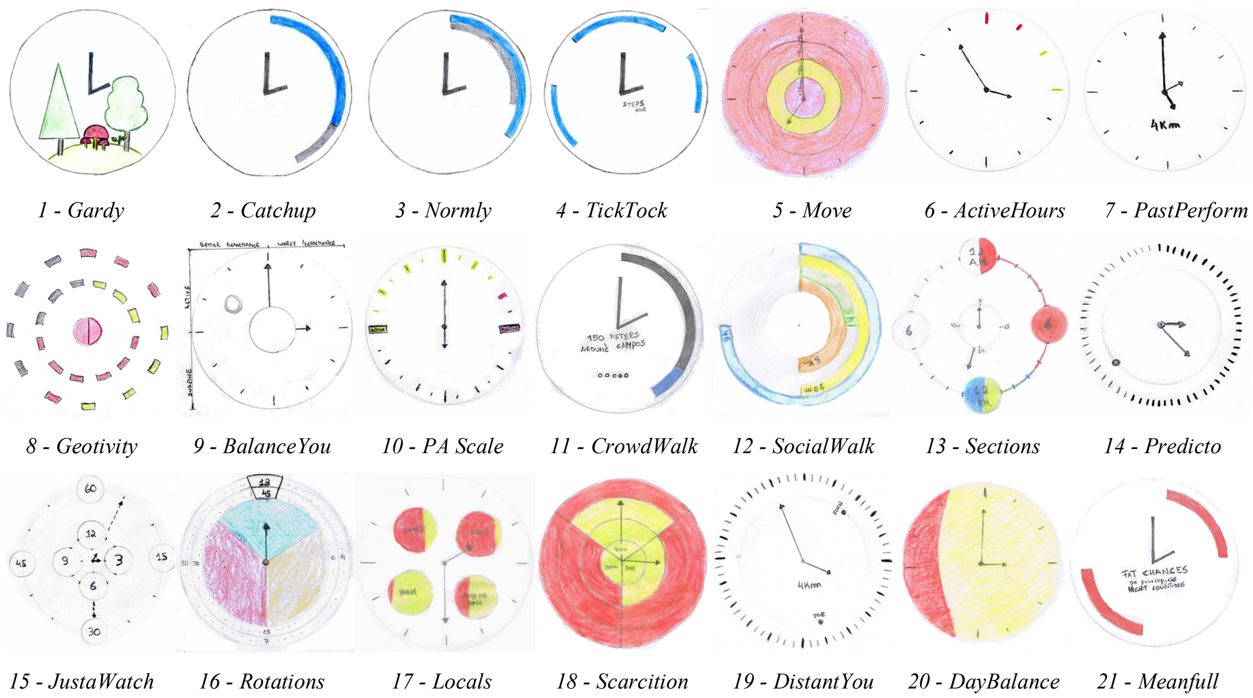

Figure 2. The 21 concepts of glanceable physical activity feedback

one concepts (see Figure 2) and six design qualities. We briefly summarize each of the six design qualities and illustrate them either with one of our 21 concepts or prototypes already existing in the literature (see Table 1).

Abstract
Abstraction of data is perhaps the most prevalent quality of glanceable displays [30,38]. A number of existing prototypes and products apply this principle. Abstracting data, as opposed to displaying raw data, allows users to process and perceive information with minimal consciousness [20], enabling quick awareness and reflection on one's behaviors [11].

To support abstraction, all 21 concepts convey step count through abstract forms, such as circles (e.g., Fig.2.12), or stylized representations (e.g., Fig.2.1). *Gardy* (Fig.2.1), for instance, uses the metaphor of a blossoming garden to highlight one's progress towards goal completion – a simplified variant of UbiFit Garden's abstraction of user's activity levels [11]. Similarly, *Geotivity* and *SocialWalk* (Fig. 2.8 and Fig. 2.12), use shapes to represent different facets of one's physical activity – *Geotivity* displays the moments in which one was active and sedentary (green and red rectangles) over the course of a day, while *SocialWalk* displays different aspects of one's physical activity, such as the total distance walked or time sedentary, through circles.

Integrates with existing activities
Another principle that often came out in our analysis of the emerging concepts was that of *integration with existing activities*. Embedding feedback into frequently occurring activities makes the feedback more likely to be glanced. In fact, glanceable displays have been commonly placed in frequently accessed locations - such the background of one's mobile phone [11] or the periphery of one's vision [5]. Prior work has found that users check their smartwatch 60-80 [35] and 95 [31] times in a day, with more than half of the usage being fueled by checking the time, or triggered by an incoming notification. Following upon this, we decided to integrate all 21 concepts with the practice of checking the time; feedback was placed on the periphery or the background of the primary screen of the smartwatch, whose main function was to tell the time.

Support Comparisons to Targets and Norms
Activity trackers commonly provide descriptive feedback – they tell us how much we have walked but not whether this is enough [33]. Feedback that presents *progress in comparison to a target* can be easier for the user to process, helping the user evaluate their behavior relative to a certain goal rather than presenting raw data requiring further inferences. Consider, for instance, *Fitbit Flex's* glanceable feedback. The wristband features five LEDs that illuminate for each 20% of a daily walking goal achieved. However, even this seemingly simple display requires some quite difficult projections, if one wants to use it for immediate self-regulation. Since for an office worker physical activity is not a constant background task, users need to estimate how likely it is to meet the daily goal based on the distance walked so far and opportunities to walk in the future.

Table 1. We identified 6 underlying design qualities in our 21 concepts

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Being abstract		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Integrating with activities		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Comparison to targets and norms			■	■		■	■	■	■	■	■		■	■				■	■	■	■	
Being Actionable						■						■										
Leading to checking habits	Novelty	■										■	■					■				
	Scarcity				■	■	■												■			
Proxy to further engagement															■							■

Normative comparisons can reduce this burden of projection. For instance, *PastPerform* (see Fig 2.7) and C*atchup* (see Fig 2.2) compare the distance walked so far to the distance walked at the same time yesterday, or at a day when one's goal was barely met, respectively. Following the same logic, *Normly* (see Fig 2.3) employs a large database of other people's walking on different days and compares at every glance, the distance one has walked so far to that of other users, who usually are equally active. *DistantYou* (see Fig 2.19) follows the same approach as *Normly*, but highlights the specific time in which other people met their goal. *ActiveHours* (Fig 2.6) and *Sections* (Fig 2.13) further attempt to project norms by highlighting how balanced a user has been (i.e. active vs inactive) over the course of an hour, while *PA Scale* (Fig 2.10), *BalanceYou* (Fig 2.9) and *DayBalance* (Fig 2.20) highlight how balanced a user has been over the course of a full day.

All the previously described interfaces provide normative, directly interpretable feedback that helps users maintain an awareness of their performance at a glance.

Actionable

Another quality that often surfaced in our exploration was that of *actionable* feedback. Effective glanceable feedback interfaces should not only inform but also instigate short, goal related actions [25]. An example is *CrowdWalk* (Fig 2.11), which presents in a brief text walking challenges one may perform from the current location, and visualizes the contribution these would make towards meeting the daily goal. For instance, as users enter a building, *CrowdWalk* may suggest taking the stairs; when entering a supermarket, users may be challenged to leave their shopping cart behind while walking back and forth to gather items. As another example, *Move* (Fig 2.5) suggest moments, every 15 minutes, where a user should try to fit in exercise over the course of a day. *Move* takes into account users' calendar, and levels of past activity to make such recommendations.

Leads to checking habits

While glancing is the dominant form of interaction with smartwatches [35] and physical activity trackers [19], prior work has shown the frequency of glances as well as the overall engagement with feedback to decrease over time [19]. This drop in engagement may have detrimental effects on behavior change as individuals quickly relapse once self-monitoring stops [36], while the frequent monitoring of one's behaviors can help prevent relapse. We thus argue that glanceable feedback should be able to sustain the frequency of glancing over the long run, or in other words to *instigate checking habits* [41]. Prior work has suggested this to be feasible. For instance, Oulasvirta et al. [34] linked the information gratification users derive from social media updates and incoming emails on their smartphones to the creation of "checking habits: brief, repetitive inspections of dynamic content quickly accessible on the device".

Our ideation process resulted to two approaches for the creation of checking habits: *novelty*, and *scarcity*.

Novelty asks: what if the feedback provided by an activity tracker constantly presents new information? This is a well-employed strategy in the computer gaming and airline industries, which regularly update content to sustain interest in games or safety instructions. According to Oulasvirta et al. [34], the gratification people derive from encountering novel content as they check their smartwatch would reinforce the habit of checking for new information. Gouveia et al. [19] employed this strategy in the design of the *Habito* mobile app, which, among other features, presented users with textual messages providing feedback about their physical activity. They found that when users read a novel message, they would take less time to come back to the app than when encountering a message they had read before. In the case of glanceable displays, feedback should be short and quickly apprehensible. For instance, *Locals* (Fig. 2.17) portrays random places where a user has walked over the course of the day, indicating his activity (and inactivity) levels within. *CrowdWalk* (Fig 2.11) further leverages on novelty by constantly updating the walking activities suggested to the user. *SocialWalk* (Fig 2.12) compares a user's progress towards goal completion to the progress of random friends. *Locals, CrowdWalk and SocialWalk* leverage on the idea of novelty by updating the places, activities and friends, respectively, multiple times per day. *Gardy* (Fig 2.1) further supports novelty by introducing new elements into users' garden as they progress towards their walking goal.

Scarcity suggests that checking habits may be created if feedback is turned into a scarce resource [13]. Scarcity is a powerful persuasion strategy – individuals are, for instance, more likely to subscribe to a workshop if they know seats are limited [9]. Existing media already apply this principle. For instance, individuals often endure TV commercials to assure they do not miss parts of an interrupted show. Likewise, social media users, such as those on Facebook, frequently reengage to ensure that they do not miss major content among many updates. Overall, people often build their revisit patterns around the update patterns of content to be viewed [4]. Building upon this principle, behavioral feedback could be displayed for a limited amount of time, thus reinforcing re-engagement habits and the frequent monitoring of behaviors. As an example, *TickTock* and *Scarition* (Fig 2.4 and 2.18) portray moments in which a user was active over the past hour and, respectively, the same information but in comparison to his friends.

Acts as a proxy to further engagement

Prior work has found that individuals quickly lose interest in deep data exploration [24]. We argue that glanceable feedback can be designed with the goal of creating "aha" moments, thus acting as cues for further engagement with the feedback. One strategy could be to present information that raises questions rather than provides answers. For instance, *Meanfull* (Fig 2.21) highlights patterns in user data through textual messages (e.g., "Lazy Tuesdays..."), while offering users the opportunity to further explore the underlying data. Another strategy could be to present insights that surprise the user. For instance, *Predicto* (Fig 2.14) analyzes parameters such as past night's sleep quality, the weather over the upcoming day and existing patterns in physical activity to predict the activity levels of the upcoming day. When predictions challenge a user's expectations, the user may become interested to explore the grounds for this surprising prediction.

FIELD DEPLOYMENT OF 4 GLANCEABLE INTERFACES

Next, we wanted to evaluate some of the assumptions that were generated during the ideation phase, in the real world. We selected and prototyped four concepts and deployed them over 28 days with twelve participants. The goal of the study was to compare concepts in terms of their adoption, how participants engaged with them, and what impact they had on their physical activity. We did not design this study to evaluate each concept's efficacy towards behavior change, given the limited sample and short, seven-day exposure participants had to each of the interfaces. Rather, we wanted to inquire into participants' experiences with the four interfaces that go beyond their initial reactions.

Interfaces

We selected four of the twenty-one concepts based on two criteria: *diversity* and *feasibility*. First, we excluded certain concepts, as they were infeasible to prototype to a mature stage within our available resources. Next, we selected two concepts (goal completion and stylized representation) due to their similarity of existing work (*Fitbit Flex's* wristband LED feedback and *UbiFit Garden*, respectively). Finally, we selected two additional, diverse concepts that we deemed represented interesting design claims. We do not argue that these concepts represent the entire design space of glanceable interfaces. We also do not assume the interfaces as a direct representation the theories that motivated them. Their performance during the field study depended on their implementation as much as the design claims they encapsulate.

Figure 3. *Gardy* (left) and *Goal Completion* (right), two of the concepts developed as watchfaces

All interfaces were developed as watch faces for Android Wear. Each comprised the primary screen of the smartwatch. Their only interactive feature was to allow users to set a daily goal for physical activity. We developed, debugged, and field-tested all interfaces on the LG G Watch R to control for variations in interfaces across hardware or other confounders related to hardware variation [1].

TickTock

TickTock (see Fig. 1) portrays, in the periphery of the smartwatch, the periods in which one was physically active over the past hour. We expected *TickTock* to present two main advantages over the other interfaces. First, through turning the feedback into a scarce resource [9] – by constraining it to only the past hour – we expected to build "checking habits", i.e., frequent monitoring of the smartwatch to make sure that no feedback goes unnoticed. This increased frequency of self-monitoring may, in turn, lead to increases in individuals' physical activity. Secondly, we expected that presenting physical activity of only the past hour would inherently lead participants to strive for keeping a balance of physical activity throughout their days. For instance, if they notice that they have been inactive for the past hour, they may try to have a short walk. As a result, contrary to the remaining three interfaces which aim at assisting individuals in achieving a daily goal, *TickTock* may be pushing individuals to avoid prolong periods of sedentarism, which has been found to be a health risk factor independently of the amount of physical activity one performs over the course of a day [37].

Normly

Normly (see Fig. 1) compares at each glance one's daily progress to that of others having the same goal. To establish normative data, we leveraged a database of the daily walking progress of 25 individuals, on a total of approximately 20000 days. We split the database in 10 groups, reflecting the distance walked at the end of the day (i.e., 7km, 8km etc). We then split the data in 1-min intervals, averaging the values within each group. As a result, if a user defines a goal of 8 km/day, *Normly* will compare, at a resolution of 1-min, his daily progress to the average progress of people who walked 8 km by the end of the day. We expected this normative feedback would lead to more frequent action and increases in overall physical activity at the end of the day, for instance in comparison to *Goal Completion*, which simply presents but does not evaluate one's daily progress.

Gardy

Gardy (see Fig. 3) abstracts physical activity levels through a garden, blossoming as individuals' progress towards their daily walking goal. At the start of each day, the garden is bare, with elements such as leaves, mushrooms and trees appearing as they reach their goal. Such abstract, stylized representations have been previously found to sustain users' engagement, through fostering curiosity on users as they anticipate the unfolding of the story, while individuals tend to appreciate the attractiveness and variety of metrics conveyed in such displays [11]. Yet, little is known as to how individuals engage with such representations and the impact they have on users' behaviors.

Goal Completion

Goal Completion (see Fig. 3) presents one's progress towards their daily goal. Participants were presented with a preset goal of 10K steps [44] and were allowed to modify it. Ample evidence exists on the efficacy of goal setting [33] - individuals that set specific goals (e.g., walk 10K steps per day) to be more likely to enhance self-regulation and activate self-evaluations than those which set abstract goals as "do my best" or "try hard" [27]. We decided to set a challenging default goal that reflects medical practitioners' recommendations (i.e., 10K steps) as previous studies on activity tracking have found individuals to have limited understanding of their daily physical activity and to go with the preset goal, even when this is unrealistically low [19], while setting a challenging goal is strongly linked to greater performance [27]. *Goal setting* is no different to commercial prototypes (e.g., Fitbit's feedback on band). We included *Goal Setting* as a baseline, against which we could compare the remaining glanceable interfaces.

Participants

We recruited participants through the reddit community, via the *lggwatchr* subreddit. To qualify, participants had to own an *LG G Watch R* and be willing to commit to use the four interfaces for a total of 28 days. A total of 12 participants successfully completed the study (median age = 25, all male). Seven participants were located in the U.S., two in Canada (25%), and one in Italy and Sweden respectively. They all had prior experience with physical activity tracking. Participants were rewarded with a 40€ voucher upon successful completion.

Readiness to change

We did not limit our sample to participants of certain 'readiness' to change as we wanted to have a diverse sample. However, we measured the stage of behavior change individuals were in using a five-item questionnaire [26]. Our population was biased towards physically active people: no participant was in the precontemplation stage, three in the contemplation stage, two in preparation, two in action and five in maintenance. Prior work has shown activity trackers to work best for people in the intermediary stages of behavior change (contemplation and preparation); in other means, individuals that have the will but not yet the means to change their behaviors [19]. This has to be taken into consideration when interpreting our results. We chose not to use participants' readiness as a variable in our analysis process due to our limited sample size.

Procedure

We debriefed participants and assisted them in installing our application. They used each interface for seven days, followed by a Skype interview, which introduced the upcoming interface and inquired into their usage and experience with the past one. The order of interfaces was counterbalanced across participants. Each interview lasted up to 15 minutes. All interviews were audio-recorded and transcribed by two independent researchers.

Participants were asked to keep the interfaces for the full duration of the study. They were informed that their physical activity and smartwatch usage would be tracked. During our final interview, we asked participants to rank the interfaces in terms of general preference and motivation to exercise, from most to least preferred, and we allowed them to continue using all interfaces after the study elapsed.

We logged participants' physical activity and smartwatch use in order to compare our concepts in terms of adoption, engagement and impact on physical activity. To track participants' physical activity, we made use of Android's step counter, tracking the start and end time of walking activities as well as the number of steps taken while walking. Regarding smartwatch usage, we tracked the time and duration of individual usage sessions, as well as interactions within a session, such as swiping to settings or launching additional applications. A usage session was defined by the time the smartwatch screen was turned on (i.e. interactive mode), until the screen was turned off or timed out (i.e. ambient mode). We also tracked incoming notifications, in an attempt to distinguish smartwatch use motivated by checking notifications versus our interfaces.

Findings

We first summarize overall participant engagement with all interfaces and their physical activity over the course of the 28 days. Next we delve into engagement, experience and impact on physical activity of each of the four interfaces.

Overall engagement and physical activity

All in all, participants checked their smartwatch on average 107 times per day (SD=80), which is slightly higher than in previous studies [31]. Over 80% of all usage sessions were *glances*: sessions in which a participant briefly checks his smartwatch and lets the screen timeout, with no further interaction [6]. Such sessions were short, with a median duration of 7 seconds (SD= 10).

Participants primarily used their smartwatch to check the time or incoming notifications: interactions following a notification (up to one minute), accounted for 41% of all usage. Participants often commented that while they did not engage with the watch in order to check their physical activity, they often paid attention to physical activity feedback, which became a constant reminder to move:

> *I would actually look at the time, but I would also happen to look at the steps*. [P3]

> *I've always expected to see this information privately, such as on a website or my mobile. But, I feel it's a little bit more motivating to have it always 'in my face'*. [P8]

Overall, participants engaged fewer times per day and walked less per day while using Gardy than with any of the other interfaces (see Table 2). Pairwise comparisons with Bonferroni correction revealed significant differences in participant engagement between *Gardy* and *Normly* ($p<0.05$), *Gardy* and *Goal Completion* ($p<0.05$), *Gardy* and *TickTock* ($p<0.05$) and marginal differences in terms of physical activity ($p<0.10$) between *Gardy* and all of the remaining interfaces.

These findings were consistent with participant preferences. Normly was the most preferred prototype (for 9 of 12 participants), followed by TickTock (2 of 12) and Goal Completion (1 of 12). *Ticktock* was also the most controversial as 3 participants considered it their least preferred. The least preferred prototype was Gardy, with 8 participants considering it their least preferred.

Table 2. Mean daily usage sessions and step count per interface

	Normly	*TickTock*	*Goal Comp.*	*Gardy*
Usage sessions	122 (SD: 99)	110 (SD: 81)	108 (SD: 69)	86 (SD: 60)
Step count	5460 (SD: 4528)	5150 (SD: 4543)	5340 (SD: 4528)	3760 (SD: 3511)

Participant experiences with Normly

We expected that providing participants with normative feedback on their performance would lead to more frequent action and higher overall levels of physical activity, as compared to *Goal Completion*. This was not confirmed at an overall analysis, as an independent samples t-test showed no significant differences among the daily number of steps walked across both interfaces ($mean_{Normly}$ = 5460 steps, $mean_{Goal\ Compl.}$ = 5340 steps, $t(165) = -0.18$, $p=0.86$).

However, we noticed differences among participant behaviors based on how far ahead or behind others they were at each given moment. More specifically, we looked at participant physical activity upon interacting with the watch. Participants interacted with *Normly* a total of 9472 times. In 1855 of those (20%), they were up to 500 steps behind or ahead of others. In 5764 of the times (61%) they trailed behind others by over 500 steps, while in 1799 of the times (19%) were more than 500 steps ahead of others.

We found that, when close to others, participants would take a mean of 5 min after the interaction to start a new walk, and they would walk on average 394 steps. In contrast, when lagging behind by over 500 steps, participants would take significantly more time to start a new walk (mean=19 minutes, ($t(7614) = -10$, $p<0.01$) and walk significantly less steps (*mean*=156 steps, $t(7614) = 19.3$, $p<0.01$), as confirmed by independent samples t-tests. The same happened when participants were far ahead of others, where they would take 10min on average to start a new walk, $t(3649) = -13.1$, $p<0.01$, and walk for an average of 248 steps, $t(3649) = 9.94$, $p<0.01$.

Participants felt motivated to walk when sensing they could easily catch-up or stay ahead of others. This effect would disappear, though, once differences grew bigger in either direction:

> *If I was way far ahead, I wouldn't do much. If I was just a little ahead, I would try to walk and keep ahead.* [P3]

In certain ways, these findings are not surprising. More than providing normative feedback, *Normly* engaged participants in a competition with others, even though they had no relationship to or understanding of who these others were. Participants accepted these others as *similar to themselves* – knowing they shared they same walking goals, and competed with them on a daily basis.

> *... I mean, we have the same goal so we should be walking about the same* [P6]

> *I liked being able to see how good or bad I did against others at a glance (...) even though I didn't know them, It made me want to keep up with other people.* [P12]

From a social comparison perspective, individual motivation and performance is expected to be heightened when outperforming others is attainable but not certain [43]. However, in over 60% of the times individuals checked their watch, they trailed behind others considerably. In fact, participants achieved their daily goal on average only once over the seven days, and as a result were compared to others who consistently performed better, which had a toll on their motivation:

> *It was tough seeing others always ahead of me and knowing I couldn't catch up to them (because I was having a busy week). I just ignored how much others had walked and tried to focus only on mine* [P9]

We must note that participant's underachievement was emphasized as they were being compared to people which met that goal by the end of the day. This was not the case of participants, as they were trying to achieve it - either successfully or not.

A question raised is: if participants witnessed that they consistently underperformed compared to others and that this has a toll on their motivation, why didn't they decrease their daily walking goal? Our analysis suggests that participants wanted to retain their initial target, as they felt the reward of reaching a more demanding goal was more enticing than outperforming a less-competitive group, *e.g.*:

> *I was mostly behind [others], but I didn't really think about changing [my goal] (...) I know I can achieve 8000 steps, so why change it to 5000?(...) It's pretty sweet when I hit my goal before them* [P12]

These insights have implications for the design of normative glanceable feedback, suggesting a need for more dynamic systems that maintain comparative levels of performance for a higher percentage of the time. This might be achieved through deception (e.g., artificially lowering the performance of others to provide an opportunity for the participant to catch up).

Further, many felt frustrated with the flexibility of the interface, as they had to keep putting in steps throughout the day to keep up with others:

> *it's not easy to keep ahead (...) an hour ago I was 90 steps behind so I walked a bit until I was 100 steps ahead. But now I am already 80 steps behind! It is frustrating, but if I don't keep up they will get a lot of steps ahead* [P16]

Participant Experiences with TickTock

By displaying behavioral feedback for a limited amount of time, we expected *TickTock* to reinforce re-engagement habits. This was true as participants re-engaged with *TickTock* more frequently – on average every 9 min - as compared to *Goal Completion* (every 15 min, $t(16675)$ = 6.59, p<0.01). As one participant noted:

> *It only shows me how active I've been over the last hour, so I need to come back to it ever now and then to see how I'd been.* [P7]

Not only did *TickTock* lead to more frequent interactions, it also triggered more frequent walking activities. When using *TickTock*, participants would make on average 61 walking activities per day. An independent samples t-test revealed a significantly higher number of daily walking activities when participants used *TickTock* as compared to *Goal Completion* (mean=50, $t(162)$=-2.5, p<0.05), *Normly* (mean=51, $t(166)$=-2.3, p<0.05), and *Gardy* (mean=50, $t(166)$=-2.77, p<0.01). They would, however, walk for an average of 77 steps in each walking activity, which was significantly shorter than in *Goal Completion* (mean=106, $t(6910)$=4.8, p<0.01), *Normly* (mean=107, $t(8313)$=5.8, p<0.01), and *Gardy* (mean=99, $t(8678)$=4.6, p<0.01).

Our qualitative data suggest two main reasons for the effectiveness of *TickTock* on triggering short, frequent action from individuals. First, it strengthened individual *accountability* for maintaining minimum levels of physical exercise every hour by making this information explicit and easy to glance upon. Second, it rewarded short breaks from sedentary behaviors by making their impact *visible in an instant*:

> *It rewards my sporadic movements since I can see the colors change when I start moving.* [P3]

We further found that the feedback provided by *TickTock* was a significant predictor of later behavior. We performed a linear regression analysis to predict the time participants took until the next walk after interacting with *TickTock*, based on the feedback they received, namely their active time (min) over the past hour. The number of minutes from a participant looking at *TickTock's* feedback until their next walk can be predicted as 1.06 + 0.95 * active-time; $F(1,8465)$ = 26734, p<0.001, with a R^2 of 0.76. In other words, for every additional 10 min of physical activity that the participants saw they performed over the past hour, they would take an extra 9.5 min till their next walk (Fig. 4). Participants who saw that they walked 10 or less min over the past hour had a 77% probability of starting a new walk in the next 5 min. As one participant noted:

> *... every time I was at work and saw 0 steps in the last hour, it was a signal to get up and walk around.* [P17]

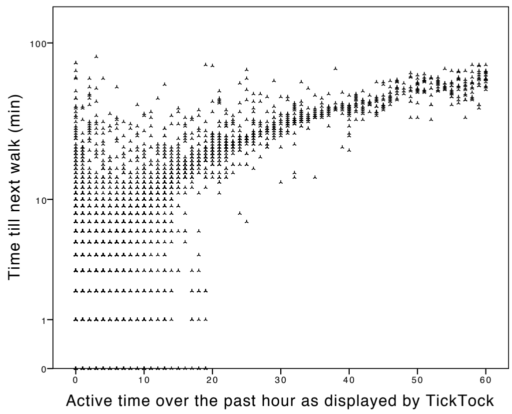

Figure 4. Witnessing that one was sedentary over the past hour would trigger physical activity in shorter period of time.

This push for frequent engagement, however, took a toll on participants' motivation, with some experiencing *reactance* and most reporting that they often felt a lack of credit for physical activity that took place earlier in the day:

> *When I looked and it said I had 0 steps over the last hour, I felt that I hadn't walked for the whole day, which was not the case, so I would think to myself: it's simply not showing the total steps from the whole day (...) I also had no clue how much I had walked over the day.* [P3]

In fact, *TickTock* was the most controversial interface, being the most preferred by two participants and the least preferred by four participants. In addition, three participants ranked it as the most motivating to exercise, while five ranked it as the least motivating. We found that these two groups of participants differed primarily on their fitness goal: participants who rated *TickTock* as motivating had already adopted the goal of breaking sedentary activity throughout the day as their primary motivation.

> *I'm not trying to hit a target so I don't really care about the total [steps] (...) I care more about seeing the steps in the last hour and keeping balanced during my day.* [P11]

In contrast, participants that rated *TickTock* as the least motivating were driven towards larger, daily goals. They found TickTock inflexible and unforgiving on days where their schedule did not allow for frequent physical activity:

> *It's less flexible depending on the day I'll have ahead of me. If I had a goal, I could adjust it on busier days, but in this one I can't really do that. If I stop for 60 min I've gone sedentary.* [P17]

As expected, these two groups differed in terms of their behaviors. An independent samples t-test of found the 'anti-sedentary' group to engage with *TickTock* more frequently (N=165) than the 'daily goal' group (N=75, $t(54) = 5.36$, $p<0.01$), perform more physical activities in the course of their days (mean=82) than they 'daily goal' group (N=51, $t(53) = 5.02$, $p<0.01$), and have marginally higher step count (N=6529) than the 'daily goal' group (N=4176, $t(54) = 1.83$, $p<0.10$) of participants which considered it the least motivating.

Participant Experiences with Gardy

Contrary to our expectations, *Gardy* was the least preferred interface and least motivating to exercise (for 8 and 7 participants, respectively), with participants engaging and performing significantly less physical activity with this interface as compared to the remaining (see section 'Overall engagement and physical activity').

Moreover, single linear regressions revealed that participant engagement decreased over the course of the seven days, by an average of 11 sessions per day (N_{Engage}=86–11*day, $F(1,82) = 11.93$, $p<0.01$, $R^2 = 0.13$). The number of steps would also decrease by an average of 442 steps per day (N_{steps}=3760–442*day, $F(1,82) = 5.62$, $p<0.05$, $R^2 = 0.06$).

Participants displayed an initial interest in the interface to see how the garden fills up. Some participants would even lower their goal to explore all the stages of the story, e.g., *"[P9]: To be honest, I lowered my goal to get to the last screen faster"*. However, after encountering all stages of the story, their engagement with *Gardy* would be halved.

> *I feel my interest wore off after time (...) probably after I figured out the cycle (...) it's fun to figure out what is going to show up next, but after you get the hang of it, it kind of loses a bit of interest* [P9]

Participants further reported difficulties in estimating exactly their progress over the course of a day, as *Gardy* did not provide numerical feedback on one's step count. In fact, many participants complemented *Gardy* with an external numerical step count (e.g., Google Fit).

> *I knew it changed at every 20% of my goal but I couldn't know how much I walked, precisely. I'm sure I could figure that out, but not by just glancing at it* [P12]

Finally, the public nature of the watch, combined with *Gardy's* simple graphical representation, had a significant impact on participants' attitudes towards the interface. For some, being public was a benefit as it spurred discussion, especially in the presence of children:

> *The garden is definitely the one that attracts more attention (...) I work at a dining and some kids came up with their parents and asked me what it is. I feel good having it full when I explain, it's double rewarding… having them see I've reached my goal.* [P15]

For others, however, it was demotivating as they felt the design of *Gardy* was inconsistent with their self-identity. This would have an impact on its adoption, as participants often reported avoiding checking their watch in public:

> *I would avoid looking at the garden with other people around (...) I would hide it beneath my jacket (...) my own watchface is much simpler and not childish (...).* [P7]

Participant Experiences with Goal Completion

Contrary to *TickTock*, *Goal Completion* seemed to work best for people who preferred defined daily walking goals.

> *I like having a hard goal to hit. It motivates me more than just seeing numbers*. [P4]

Participants appreciated it's minimalistic graphical representation, at which they glanced frequently to maintain an awareness of physical activity and to reassure themselves that they had adequate progress:

> *I feel I was glancing quite often to see where I was (...) by quickly looking at the circle I could tell if I was around 15%, 30% or 50% of my goal.* [P4]

They often developed shortcuts in their decision-making, such as the following one, who developed the strategy of having a short walk if goal completion was less than 50%:

> *I would try to walk when the circle was only half full* [P4]

Interestingly, when interacting with Goal Completion, participants performed the fewest updates of their daily walk goal (N=13) among all interfaces (N_{Normly}=35, N_{Gardy}=66, $N_{TickTock}$=20). A plausible explanation for this was the lack of novelty of Goal Completion, as all participants had prior exposure to similar feedback through their own activity trackers.

> *I can't say it took me by surprise, I already track my progress (...) It just makes it a bit more glanceable (...) I don't feel it gives me the extra push like the rest do.* [P10]

DISCUSSION AND CONCLUSION

Our design space exploration of glanceable feedback for physical activity trackers resulted in 21 concepts and six overall design qualities. Based on this we prototyped and deployed four concepts "in the wild", representing different elements of the design space. We found that, as expected, integrating feedback with frequently performed activities, such as checking the time, provides a promising path for self-monitoring tools. Participants engaged with their watches about 100 times per day, which is substantially higher than the number of times people engage with an activity tracker app on a smartphone [19]. While checking activity levels was most of the time not their primary intention, they would still glance at it, which impacted their subsequent behavior.

In our analysis of how people responded to the different prototypes, their use to support self-regulation was striking. When using *TickTock*, people who saw that they had a sufficient number of active minutes in the preceding hour were less likely to initiate a new walk, while an individual who had not been active was more likely to initiate a new walk soon.

In people's reactions to *Normly*, however, we see how some presentations of a lack of activity can rather be demotivating. Users took less time to start a new walk, and walked for longer distances, when they were closely behind or even ahead of others. If the difference was large in either direction, however, the feedback did not inspire new walks: the user was either comfortably ahead or too far behind to catch up. These findings corroborate social comparison research – motivation increases when outperforming others is attainable but not certain [43]. These demotivating examples are common in social comparison. In over 60% of the glances at *Normly*, participants saw themselves substantially underperforming. Rather than presenting demotivating feedback in these instances, feedback should maximize its effect on behaviors. One approach, as we discussed earlier, could be the use of *benevolent deception* – for instance, artificially lowering the performance of others, or changing how it is portrayed, to communicate an opportunity for the user to catch up [3,10].

Our study further showed how the different interfaces support self-regulation of different targets, and thus lead to different behavioral patterns. For instance, displaying behavioral feedback for a limited amount of time, as in the case of *TickTock*, led participants to *re-engage* and *walk more frequently*. In contrast, feedback about completion of traditional step goals best supported reaching one's target step count. These are quite subtle effects designers have to consider. Aligning measures and feedback with the desired behavior is key.

Previous research led us to expect *Gardy* to be a popular interface. Participant's responses, however, did not support this. First, this serves as an important reminder that interfaces for smartwatches are more public than smartphones. They must be evaluated not just for their efficacy, but also for their fit with user self-identify [15] and even fashion [42]. Second, while *Gardy*'s stylized representation created some interest in the beginning, it could not sustain interest. After observing one full cycle, participants got bored. More variation, as offered by *UbiFit Garden*, would be important here. How to fit this on a watch interface, however, remains a challenge. Third, many participants experienced difficulties in evaluating their exact progress and planning actions over the course of a day *Gardy's* vague representation. Participants seem to mainly associate exact measurements with a tracker and expect according feedback. This may be a consequence of the all-male, already physically active sample of participants, who in fact already owned smart watches. However, this does not imply that vague feedback is wrong. It can be a way to motivate other people (e.g., novices), who do not respond favorably to a framing of activity in terms of numbers and performance. Fourth, the semantic link between a garden and physical activity is rather weak. Because of this, the garden does not offer the most meaningful story. Letting a garden grow through activity appears slightly arbitrary. Other "stories", such as tending to a Tamagotchi-like dog, which wants and needs to be walked, might be more acceptable and interesting over a longer period of time.

All in all, our study shows that glanecable feedback has a positive effect in general, through its increased availability. More importantly, we showed how subtle differences in interaction emerge, depending on the exact concept (i.e., form) chosen. While some may argue that the how doesn't matter as long as people become more active, we believe that especially for a more sustained use the exact mechanism invoked matter. While the garden may not have been the wisest choice, a vague, varying, more story-like concept could be more motivating than, for example, social comparison in the long run. The story unfolds, while social comparison simply becomes demotivating the moment one realizes that there is no chance of getting ahead of others. This hints a noteworthy limitation of our study. While it was in the wild, it still featured only seven days of using each interface. This prevents drawing any conclusions about long term behaviors from the present results [21].

While future research is needed to assess long-term use and effects, as well as differences in more diverse populations, our study outlines a rich design space for these further explorations. The results of the field study show the importance of aligning feedback with the desired behavior, and highlights opportunities to present more motivating feedback and in ways that are have greater potential to sustain user interest.

ACKNOWLEDGMENTS

This research was supported by the Portuguese Foundation of Science and Technology grants CMUP-EPB/TIC/0063/2013 and SFRH/BD/89796/2012.

REFERENCES

1. Adam Noah, J., Spierer, D. K., Gu, J., & Bronner, S. (2013). Comparison of steps and energy expenditure assessment in adults of Fitbit Tracker and Ultra to the Actical and indirect calorimetry. Journal of medical engineering & technology, 37(7), pp. 456-462.
2. Adams, A. T., Costa, J., Jung, M. F., & Choudhury, T. (2015). Mindless computing: designing technologies to subtly influence behavior. In Proceedings of Ubicomp'15, pp. 719-730.
3. Adar, E., Tan, D. S., & Teevan, J. (2013). Benevolent deception in human computer interaction. In Proceedings of CHI'13, pp. 1863-1872.
4. Adar, E., Teevan, J., & Dumais, S. T. (2008). Large scale analysis of web revisitation patterns. In Proceedings of CHI'08, pp. 1197-1206.
5. Bakker, S., van den Hoven, E., & Eggen, B. (2015). Peripheral interaction: characteristics and considerations. Personal and Ubiquitous Computing, 19(1), pp. 239-254.
6. Banovic, N., Brant, C., Mankoff, J., & Dey, A. (2014). ProactiveTasks: the short of mobile device use sessions. In Proceedings of MobileHCI'14, pp. 243-252.
7. Bentley, F., Tollmar, K., Stephenson, P., Levy, L., Jones, B., Robertson, S., Price, E, Catrambone, R., & Wilson, J. (2013). Health Mashups: Presenting statistical patterns between wellbeing data and context in natural language to promote behavior change. Transactions on Computer-Human Interaction, 20(5), 30.
8. Choe, E. K., Lee, B., Kay, M., Pratt, W., & Kientz, J. A. (2015). SleepTight: low-burden, self-monitoring technology for capturing and reflecting on sleep behaviors. In Proceedings of Ubicomp'15, pp. 121-132.
9. Cialdini, R. B., & James, L. (2009). Influence: Science and practice (Vol. 4). Boston, MA: Pearson education.
10. Colusso, L., Hsieh, G., & Munson, S. A. (2016). Designing Closeness to Increase Gamers' Performance. In Proceedings CHI'16, pp. 3020-3024.
11. Consolvo, S., Klasnja, P., McDonald, D. W., Avrahami, D., Froehlich, J., LeGrand, L., Libby, R., Mosher, K., & Landay, J. A. (2008). Flowers or a robot army?:encouraging awareness & activity with personal, mobile displays. In Proceedings UbiComp'08, pp. 54-63.
12. Consolvo, S., Klasnja, P., McDonald, D. W., & Landay, J. A. (2014). Designing for Healthy Lifestyles: Design Considerations for Mobile Technologies to Encourage Consumer Health and Wellness. Human–Computer Interaction, 6(3-4), pp. 167-315.
13. Döring, T., Sylvester, A., & Schmidt, A. (2013). A design space for ephemeral user interfaces. In *Proceedings of* TEI'13*,* pp. 75-82.
14. Epstein, D., Cordeiro, F., Bales, E., Fogarty, J., & Munson, S. (2014). Taming data complexity in lifelogs: exploring visual cuts of personal informatics data. In Proceedings of Ubicomp'14, pp. 667-676.
15. Fortmann, J., Cobus, V., Heuten, W., & Boll, S. (2014, November). Waterjewel: Design and evaluation of a bracelet to promote a better drinking behaviour. In Proceedings of Mobile and Ubiquitous Multimedia, pp. 58-67.
16. Fritz, T., Huang, E. M., Murphy, G. C., & Zimmermann, T. (2014). Persuasive technology in the real world: a study of long-term use of activity sensing devices for fitness. In Proceedings of CHI'14, pp. 487-496.
17. Froehlich, J. E., Larson, E., Campbell, T., Haggerty, C., Fogarty, J., & Patel, S. N. (2009). HydroSense: infrastructure-mediated single-point sensing of whole-home water activity. In Proceedings of Ubicomp'14, pp. 235-244.
18. Goffman, E. (1963). Stigma: Notes on the management of spoiled identity. Englewood Clifls, NJ: Prentice-Hall
19. Gouveia, R., Karapanos, E., & Hassenzahl, M. (2015) How Do We Engage with Activity Trackers? A Longitudinal Study of Habito. In Proceedings of Ubicomp'15, pp. 1305-1316.
20. Ham, J., & Midden, C. (2010). Ambient persuasive technology needs little cognitive effort: the differential effects of cognitive load on lighting feedback versus factual feedback. In Persuasive Technology, pp. 132-142.
21. Hazlewood, W. R., Stolterman, E., & Connelly, K. (2011). Issues in evaluating ambient displays in the wild: two case studies In Proceedings of CHI'11, pp. 877-886.
22. Houben, S., Golsteijn, C., Gallacher, S., Johnson, R., Bakker, S., Marquardt, N., Capra, L., & Rogers, Y. (2016). Physikit: Data Engagement Through Physical Ambient Visualizations in the Home. In Proceedings of CHI'16, pp. 1608-1619.
23. Jafarinaimi, N., Forlizzi, J., Hurst, A., & Zimmerman, J. (2005). Breakaway: an ambient display designed to change human behavior. In Proceedings of CHI'05, pp. 1945-1948.
24. Karapanos, E., Gouveia, R., Hassenzahl, M., & Forlizzi, J. (2016). Wellbeing in the Making: Peoples' Experiences with Wearable Activity Trackers. *Psychology of Well-Being*, *6*(1), pp. 1-17.
25. Katz, D., Dalton, N., Holland, S., O'Kane, A., & Price, B. A. (2016). Questioning the Reflection Paradigm for

Diabetes Mobile Apps. In EAI International Conference on Wearables in Healthcare.

26. Kearney, J. M., de Graaf, C., Damkjaer, S., & Engstrom, L. M. (1999). Stages of change towards physical activity in a nationally representative sample in the European Union. Public health nutrition, 2(1a), pp. 115-124.
27. Latham, G. P., & Locke, E. A. (1991). Self-regulation through goal setting. Organizational behavior and human decision processes, 50(2), pp. 212-247.
28. Li, I., Dey, A., & Forlizzi, J. (2010). A stage-based model of personal informatics systems. In Proceedings of CHI'10, pp. 557-566.
29. Mankoff, J., Dey, A. K., Hsieh, G., Kientz, J., Lederer, S., & Ames, M. (2003, April). Heuristic evaluation of ambient displays. In Proceedings of CHI'2003, pp. 169-176.
30. Matthews, T., Blais, D., Shick, A., Mankoff, J., Forlizzi, J., Rohrbach, S., & Klatzky, R. (2006). Evaluating glanceable visuals for multitasking. Technical Report EECS-2006-173. UC Berkeley.
31. Min, C., Kang, S., Yoo, C., Cha, J., Choi, S., Oh, Y., & Song, J. (2015). Exploring current practices for battery use and management of smartwatches. In Proceedings of ISWC'15, pp. 11-18.
32. Mullet, K. & Sano, D. (1995) Designing visual interfaces: Communication oriented techniques. Sunsoft Press
33. Munson, S. A., & Consolvo, S. (2012). Exploring goalsetting, rewards, self-monitoring, and sharing to motivate physical activity. In Proceedings of PervasiveHealth'12, pp. 25-32.
34. Oulasvirta, A., Rattenbury, T., Ma, L., & Raita, E. (2012). Habits make smartphone use more pervasive. Personal and Ubiquitous Computing, 16(1), pp. 105-114.
35. Pizza, S., Brown, B., McMillan, D., & Lampinen, A. (2016). Smartwatch in vivo. In Proceedings of CHI'16, pp. 5456-5469.
36. Prochaska, J. O., DiClemente, C. C., & Norcross, J. C. (1992). In search of how people change: applications to addictive behaviors. American psychologist, 47(9), pp. 1102-1114.
37. Proper, K. I., Singh, A. S., Van Mechelen, W., & Chinapaw, M. J. (2011). Sedentary behaviors and health outcomes among adults: a systematic review of prospective studies. American journal of preventive medicine, 40(2), pp. 174-182.
38. Rogers, Y., Hazlewood, W. R., Marshall, P., Dalton, N., & Hertrich, S. (2010, September). Ambient influence: Can twinkly lights lure and abstract representations trigger behavioral change?. In Proceedings of Ubicomp'10, pp. 261-270.
39. Rooksby, J., Rost, M., Morrison, A., & Chalmers, M. C.(2014). Personal tracking as lived informatics. In Proceedings of CHI'14, pp. 1163-1172.
40. Schneider, H., Moser, K., Butz, A., & Alt, F. (2016). Understanding the Mechanics of Persuasive System Design: A Mixed-Method Theory-driven Analysis of Freeletics. In Proceedings of CHI'16, pp. 309-320.
41. Stawarz, K., Cox, A. L., & Blandford, A. (2014). Don't forget your pill!: designing effective medication reminder apps that support users' daily routines. In Proceedings of CHI'14, pp. 2269-2278.
42. Stusak, S., Tabard, A., Sauka, F., Khot, R. A., & Butz, A. (2014). Activity sculptures: Exploring the impact of physical visualizations on running activity. Transactions on visualization and computer graphics, 20(12), pp. 2201-2210.
43. Suls, J., Martin, R., & Wheeler, L. (2002). Social comparison: Why, with whom, and with what effect?. Current directions in psychological science, 11(5), pp. 159-163.
44. Tudor-Locke, C., & Bassett Jr, D. R. (2004). How many steps/day are enough?. Sports medicine, 34(1), pp.1-8.

5.3 – Discussion of results and conclusion

One of the main goals of our study was to understand if activity tracking systems can be designed to influence people's behaviors through quickly-consumable and always-available representations of their physical activity (i.e. glanceable displays). Our study showed that they can, by highlighting *opportunities for goal-oriented actions*. Glanceable displays were frequently seen by participants, which in turn drove goal-pursuit behaviors, including noticing and taking advantage of opportunities for self-regulation and goal-directed actions (as similarly described in [17]). In this section I describe this effect, and describe how glanceable displays can serve as a way to complement the deeper and reflective way of using personal informatics, and activity tracking devices.

Highlighting opportunities for goal-oriented actions

One of the main strengths of glanceable displays is their high frequency of being seen. We found that having physical activity feedback integrated with a frequently performed activity – namely, time checking, increased the frequency of interacting with physical activity feedback. Participants interacted with their watches about 100 times per day, which is substantially higher than the number of times we found users to interact with a activity tracker application on a smartphone or on a wrist-worn activity tracker – respectively, 5 and 10 times per day, as seen in our *Habito* (see chapter 3) *and Invivo* (see chapter 4) studies.

As hypothesized by Consolvo and collegues [18], these frequent interactions with physical activity feedback seemed to "trigger goal-pursuit behaviors, including noticing and taking advantage of opportunities for goal-directed actions" (p. 244). An example can be seen in our *Normly* interface. *Normly* compared users' goal completion with others with a similar daily step count goal; whenever checking *Normly,* users could see if they were ahead, inline or behind other users with a similar goal. We found that participants were more likely to engage in short bursts of physical activity when seeing themselves closely behind or ahead of others. If the differences were large in either direction, users rather felt demotivated: they either felt comfortably ahead or too far behind to catch up. Participants' motivation to perform physical activity, in this case, was at its highest when identifying opportunities for goal-oriented action – namely, keeping closely ahead or behind other people with similar walking goals.

These instances, however, were an exception when interacting with *Normly*. In over 60% of the interactions with this interface, participants saw themselves substantially underperforming. Rather than presenting demotivating feedback in these instances, feedback should maximize its effect on behaviors. One approach could be to adapt the very nature of these displays. For instance, instead of being persistently displayed on the background of one's watch (or phone), this feedback could arise only during opportunities for *goal-oriented action* – such as when a user is closely ahead or behind his friends. This feedback could be hidden, or even adapted when user start falling further away. This strategy could minimize the potential aversion and reactance of having always-visible, "non-motivating" feedback of one's physical activity (as noted in [47]).

Transitioning glances to deeper reflections on one's health

As mentioned in Chapter 4, personal informatics literature has assumed two modes of interacting with respective systems: the *reflective*, in which insights are gained through deep reviews and explorations of data; and the *impulsive*, in which people process data and take action quickly and unconsciously. As such, most of today's systems are built to support either one or another mode of use. For instance, a number of researchers have investigated how to best support the *reflective* mode of use through rich visualizations of historical information (e.g. [25]), or the *impulsive* mode of use through glanceable visualizations of physical activity feedback (e.g. [20]).

I argue that future work should focus on connecting these two modes of use. While glanceable displays support the *impulsive* mode of use by providing frequently visible and abstract representations of self-monitoring feedback, these displays are often limited in their support for reflection. Glanceable displays typically are low on information and are limited in terms of the interactions they support (as noted in [18]).

We understand glanceable displays as "portals" to deeper engagements with activity tracking data. In addition to supporting in-the-moment motivation, we also desire to use frequent, short glances with trackers to promote moments of exploration and learning, in which individuals engage with their data towards developing newfound self-knowledge – such as identifying and making sense of trends and patterns in one’s behaviors [51]. One strategy could be to present feedback that acts as cues for further reflections and engagement with feedback. For instance, a activity tracker may notify a user about his or her high sedentary levels and only through further interaction this story becomes more telling, by for example, providing physical activity levels over the past 30 minutes, creating more opportunities to reflect about reasons for the current lack of physical activity (e.g., place, time, habits). Glanceable displays would, under this perspective, unveil the beginning of a story, which will further unravel when people chose to engage further.

Chapter 6.

Discussion and Conclusion

My dissertation contributes to an examination of how people use activity trackers within their daily lives and how the design of these devices can be improved towards supporting their uses and needs. Beginning with two inquiries of how people interact with activity trackers (see Chapter 3 and 4), my work has contributed to a renewed understanding of how people derive insights and change their behaviors as the result of using an activity tracker. Based on these insights, I have implemented and evaluated a number of novel designs which address some of the ways in which people gain value from the use of their trackers (Chapter 5).

Within this chapter, I return to the overarching research questions of my dissertation and discuss the ways in which they were addressed, and how they contribute to an existing body of work in activity tracking, and personal informatics.

6.1 – How do people interact with physical activity trackers in their daily lives? (RQ1)

People's use of activity trackers often does not follow the traditional *reflective* process, in which people collect, then carefully explore and review their data in order to plan future courses of action [51]. Instead, we found the use of activity trackers to be predominantly *impulsive*, where users simultaneously reflect, learn and change their behaviors as they collect data. As discussed in Chapter 3, over 70% of the interactions with a mobile activity tracker were *glances* - brief, up to 5 second long interactions with trackers, in which people's interactions did not involve the exploration of historical data. Glances were assumed to support the frequent regulation of behaviors; people would estimate how likely they were to meet their goals, and introduce action when needed. In Chapter 4, we showed that these brief moments of interaction with trackers did not only support action, but also learning and moments of self-discovery.

These results highlight the fact that people do not necessarily have to engage in deep reviews and explorations of data to act upon and learn about their behaviors. Learning and action, also happens in-situ and impromptu, as people collect data. This suggests a need for designs of tracking tools that support brief moments of learning and action as they arise in people's lives. This may involve the development of novel feedback displays (such as glanceable displays), or even the automatic surfacing of insights (as discussed in Chapter 5).

As further discussed in Chapter 3, we found a significant amount of people to abandon the use of an activity tracker within early days, or weeks of use. We showed that these abandonments were strongly connected to people's motivations towards behavior change. Users in intermediate stages of behavior change – characterized by having the intention but not yet the means to become physically active, had an adoption rate of 56%, whereas those in initial and advanced phases of behavior change – characterized by being unwilling to become physically active or already having physical activity incorporated as an intrinsically motivated practice, had adoption rates of approximately 20%.

Further, people were found to engage less with their data as they become more self-reliant and see themselves succeeding towards their goals. As discussed in chapter 3, trackers were used as "deficit technologies", to which people mostly turn when needing support – such as during low levels of goal completion. On one hand, the gradual disengagement with trackers could be seen as a case of success (in the cases where physical activity becomes an intrinsically motivated practice). However, research has repeatedly found that a minimum of engagement with trackers is important to prevent relapses to previous behaviors [71]. In Chapter 3, we have suggested three ways for designing trackers that sustain users' engagement:

Support people in their different needs: In Chapter 3, we highlight the need to design trackers that adapt to people's different motivations towards behavior change, and different goals when using a tracker. Besides being designed as tools to support the process of behavior change, activity trackers should also be designed to instill a desire for change and help prevent relapses in behaviors. Further, trackers should also be designed to support uses beyond behavior change. As highlighted in Chapter 4, people use trackers to address several, overlapping goals besides behavior change – such as learning and self-discovery;

Creating feedback that creates strong checking habits. In Chapter 3, we show that showing people novel content on their activity trackers (such as messages which are constantly updating their content) has the potential to sustain their engagement with a tracker, both on a single session level (e.g., duration of an interaction with a tracker), as well as in terms of overall patterns of interaction (e.g., time to next usage). Building on these findings, and on recent work that highlighted smartphones' capacity to create checking habits through dynamic content, such as social media updates and incoming emails [67], we suggested using novel, always-updating content to sustain engagement with trackers.

Transitioning glances to moments for reflective engagement. While explorations of historical data is important for reflection, and thus, crucial for behavior change, we found only a minority of participants to engage in explorations of their data. We argue that future work should focus on connecting people's brief interactions with trackers to moments of deeper explorations of their data. For instance, motivated by the way movie trailers portray small and enticing segments of upcoming movie scenes to instill curiosity among viewers, we ask: what if activity trackers provide snippets of information to foster users' interest in exploring additional data? As an example, trackers could highlight trends in users' data (e.g. Home has been your most active location of the week. On average, 400meters more than other locations) while offering users the opportunity to explore the underlying historical data (e.g., graphs comparing walking distances across different locations) [34:5].

6.2 – How is the use of activity tracking devices integrated into the fabric of people's daily lives? (RQ2)

People's use of activity trackers was deeply entangled with their daily routines and practices. In Chapter 4, we showed that people make sense of their data, and decide for future plans of action, as the result of complex interactions between their data and the surrounding environment. For instance, many people interacted with their data to learn about ongoing activities – such as how active they were while teaching a class, or during house chores. Through this work, we saw an opportunity to shift activity tracking devices from records of personal data and goal completion to tools that examine people's routines, habits and present opportunities for improvement when adequate. For example, a teacher who wants to explore how active she is while lecturing a class should be given the opportunity to find answers to questions like, How does the duration of the class and number of students influence how physically active I am? How does this compare to the other classes that I teach? How was I feeling at that time? What tasks was I assigning the students with? What was happening to me during that day? This may involve enhancing data collection with contextual details – such as details from the activities under which data is tracked; or through the use of semi-automated tracking approaches, where users can add label, annotate and inquire into their internal states, like emotions, thoughts, and intentions.

We also found the surrounding environment to impact the ways in which people planned, and took action upon their behaviors. For instance, we found many people to use their trackers to form *micro-plans*, short courses of action planned within daily routines geared towards meeting one's levels of physical activity. Such plans would vary in their temporal proximity and extend, from small detours in the near future such as grabbing some water while waiting for the printer, to ones more significant and distant in time, such as reaching 1000 steps goal during an upcoming trip to the gym. These plans were found to be flexible, accounting for people's surrounding contexts and to the variability of their daily routines.

While the positive effects of proximal goals on performance have been highlighted by research in goal setting (e.g. [54]), the majority of today's trackers continue to adopt a daily step goal. Our results, however, point towards the need to transition to strategies that support short, flexible courses of action planned within daily routines. As an example, we have proposed *CrowdWalk* [66], a mobile application that infers users' location and presents a list of walking activities that can be initiated from one's current location. For instance, a user may be challenged to take their dog for a walk after a long period of TV watching. *CrowdWalk* aim at fostering an alternative approach to the dominant narrative of self-improvement and behavior change. Although Crowdwalk records user's walking distance, the goal is not to evaluate performance, but rather to raise awareness towards short courses of action that can be taken within people's daily lives.

6.3 - How can we design physical activity feedback that help people learn and take action upon their behaviors? (RQ3)

Building on the finding that over 70% of the interactions with an activity tracker were *glances* – brief moments of interaction with trackers, in which people attempted to take action and learn about their behaviors, we asked: how can we design user interfaces that support brief moments of interaction with data for learning and action? In Chapter 5 we did exactly this; we explored how glanceable displays can be designed to best impact people's behaviors. Through a iterative process of synthesis and analysis, we devised six design qualities for glanceable feedback:

- **Integration with frequently performed activities:** Embedding feedback into frequently accessed locations makes feedback more likely to be seen. In Chapter 5, we found that integrated feedback as the background of a watch, increased the frequency of interacting with physical activity feedback. People interacted with their watches about 100 times per day, which is substantially higher than the number of times we found users to interact with a activity tracker application on a smartphone or on a wrist-worn activity tracker – respectively, 5 and 10 times per day, as seen in our Chapter 3 and Chapter 4;
- **Abstraction of data:** Abstracting physical activity data through, for instance, forms, images or animations, can help people gain quick awareness on their behaviors (as similarly described in [19]). In Chapter 5, we showed ways of abstracting physical activity feedback through forms, such as circles or stylized representations of a blossoming garden;
- **Supporting comparisons to targets and norms:** Feedback that presents progress in comparison to a normative target can help users maintain an awareness of their performance at a glance. In Chapter 5, we found such comparisons to have a strong effect on people's behaviors. For instance, we found that users were motivated to perform physical activity when finding opportunities to keep their performances closely in line with the those of other people which they shared similar walking goals with;
- **Actionable feedback:** As suggested in Chapter 3, glanceable feedback should not only inform people about their behaviors, but also help them identify

opportunities for action. As discussed in chapter 5, one idea could be to present people with walking challenges at points of decision making - such as suggesting them to take the stairs, while waiting for the elevator, or to leave their shopping cart behind while walking back and forth to gather items at a supermarket;

- **Driving checking habits:** In Chapter 3, we show that showing people novel content on their activity trackers – such as novel feedback messages, has the potential to sustain their engagement with a tracker. Building on these findings, we suggested using novel, always-updating content to sustain engagement with trackers over the long-term, and;
- **Acting as a proxy to future engagement:** We suggest that glanceable displays can be designed with the goal of connecting people's brief interactions with moments of deeper explorations of their data. As discussed in Chapter 5, one idea could be to present brief snippets of information to users (e.g. Today you walked less than usual), while offering users the opportunity to further explore the underlying data (e.g. by connecting these snippets to richer visualizations of their physical activity). When presented with these snippets, the user may become interested to explore the grounds for these insights.

Next, we evaluated a number of different designs in a field study. Overall, our study showed that glanceable displays, due to their increased availability, have the potential of highlighting frequent opportunities for target-oriented action – such as keeping in line with the performance of people with similar goals; or keeping a sufficient number of active minutes in the preceding hour. In our analysis of how people responded to our different designs, their use to support the self-regulation was striking. For instance, people who saw that they had a sufficient number of active minutes in the preceding hour were less likely to initiate a new walk, while those who had not been active were more likely to initiate a new walk soon. These results show the importance of aligning feedback with desired behaviors, and highlighting opportunities for pursuing goal-oriented action.

6.4 – Conclusion

As the rates of chronical diseases, such as obesity, cardiovascular disease and diabetes continue to increase, the development of tools that support people in achieving healthier habits is becoming ever more important. Activity tracking devices hold the potential of helping pursue healthier lifestyles. However, for this promise to be fulfilled, these systems need to be well designed, not only in terms of how they implement specific behavior change techniques, but also in how they integrate into people's daily lives and address their daily needs.

To this end, our understanding of how and why people use these devices must continue to evolve as activity trackers become increasingly prevalent in people's daily lives. Designers and researchers should acknowledge the *lived* perspective of tracking. Trackers, nowadays, are deeply entangled in people's daily activities and in their overall lives. People make sense of their data, engage in learning and act upon their behaviors as the result of complex interactions between their data and the surrounding environment. These advances are likely to continue changing how people interact, reflect and act upon their data.

What is clear from our work is that following these developments is difficult. Yet, as these devices become ever more central, and present in people's everyday lives, understanding their everyday use is crucial. I have argued towards a need for complementing (or even moving away from) traditional methods for evaluating and investigate the daily use and efficacy of these tools (such randomized controlled studies), with methods that manage to capture the lived perspective of tracking. I have shown how methods – such as log-based and video-based methods, can help us gain a better understanding of the values that people derive from their everyday tracking practices, and how physical activity is brought about as the result of interactions with different components of activity tracking systems.

All in all, I believe that understanding the use of these devices in the everyday life, is a precondition towards the development of tools that best help people derive value from their tracking. While plenty of great work has already been conducted within the HCI community, further work is still needed to better understand how to design activity tracking devices that are truly effective for delivering patient-driven healthcare.

References

[1] Jeremy D. Akers, Rachel A. Cornett, Jyoti S. Savla, Kevin P. Davy, and Brenda M. Davy. 2012. Daily Self-Monitoring of Body Weight, Step Count, Fruit/Vegetable Intake, and Water Consumption: A Feasible and Effective Long-Term Weight Loss Maintenance Approach. *Journal of the Academy of Nutrition and Dietetics* 112, 5 (May 2012), 685-692.e2. DOI:https://doi.org/10.1016/j.jand.2012.01.022

[2] Jeremy D. Akers, Rachel A. Cornett, Jyoti S. Savla, Kevin P. Davy, and Brenda M. Davy. 2012. Daily Self-Monitoring of Body Weight, Step Count, Fruit/Vegetable Intake, and Water Consumption: A Feasible and Effective Long-Term Weight Loss Maintenance Approach. *Journal of the Academy of Nutrition and Dietetics* 112, 5 (May 2012), 685-692.e2. DOI:https://doi.org/10.1016/j.jand.2012.01.022

[3] Simon Attfield, Gabriella Kazai, and Benjamin Piwowarski. 2011. Towards a science of user engagement (Position Paper).

[4] Raymond C. Baker and Daniel S. Kirschenbaum. 1993. Self-monitoring may be necessary for successful weight control. *Behavior Therapy* 24, 3 (June 1993), 377–394. DOI:https://doi.org/10.1016/S0005-7894(05)80212-6

[5] Vítor Belim, Olga Lyra, Pedro Teixeira, Ana Caraban, Maria José Ferreira, Rúben Gouveia, Andrés Lucero, and Evangelos Karapanos. 2014. Beyond gamification: sociometric technologies that encourage reflection before behavior change. In *Proceedings of the 11th Conference on Advances in Computer Entertainment Technology - ACE '14*, 1–6. DOI:https://doi.org/10.1145/2663806.2663828

[6] Frank Bentley, Konrad Tollmar, Peter Stephenson, Laura Levy, Brian Jones, Scott Robertson, Ed Price, Richard Catrambone, and Jeff Wilson. 2013. Health Mashups: Presenting Statistical Patterns Between Wellbeing Data and Context in Natural Language to Promote Behavior Change. *ACM Trans. Comput.-Hum. Interact.* 20, 5 (November 2013), 30:1–30:27. DOI:https://doi.org/10.1145/2503823

[7] Marleen H van den Berg, Johannes W Schoones, and Theodora PM Vliet Vlieland. 2007. Internet-Based Physical Activity Interventions: A Systematic Review of the Literature. *Journal of Medical Internet Research* 9, 3 (September 2007), e26. DOI:https://doi.org/10.2196/jmir.9.3.e26

[8] Dena M. Bravata, Crystal Smith-Spangler, Vandana Sundaram, Allison L. Gienger, Nancy Lin, Robyn Lewis, Christopher D. Stave, Ingram Olkin, and John R. Sirard. 2007. Using Pedometers to Increase Physical Activity and Improve Health: A Systematic Review. *JAMA* 298, 19 (November 2007), 2296. DOI:https://doi.org/10.1001/jama.298.19.2296

[9] Lora E. Burke, Jing Wang, and Mary Ann Sevick. 2011. Self-Monitoring in Weight Loss: A Systematic Review of the Literature. *Journal of the American Dietetic Association* 111, 1 (January 2011), 92–102. DOI:https://doi.org/10.1016/j.jada.2010.10.008

[10] Ana Caraban, Maria José Ferreira, Rúben Gouveia, and Evangelos Karapanos. 2015. Social toothbrush: fostering family nudging around tooth brushing habits. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and*

Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers - UbiComp '15, 649–653. DOI:https://doi.org/10.1145/2800835.2809438

[11] Robert A. Carels, Lynn A. Darby, Sofia Rydin, Olivia M. Douglass, Holly M. Cacciapaglia, and William H. O'Brien. 2005. The relationship between self-monitoring, outcome expectancies, difficulties with eating and exercise, and physical activity and weight loss treatment outcomes. *Ann Behav Med* 30, 3 (December 2005), 182–190. DOI:https://doi.org/10.1207/s15324796abm3003_2

[12] Eun Kyoung Choe, Nicole B. Lee, Bongshin Lee, Wanda Pratt, and Julie A. Kientz. 2014. Understanding Quantified-selfers' Practices in Collecting and Exploring Personal Data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '14), 1143–1152. DOI:https://doi.org/10.1145/2556288.2557372

[13] Chia-Fang Chung, Nanna Gorm, Irina A. Shklovski, and Sean Munson. 2017. Finding the Right Fit: Understanding Health Tracking in Workplace Wellness Programs. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (CHI '17), 4875–4886. DOI:https://doi.org/10.1145/3025453.3025510

[14] James Clawson, Jessica A. Pater, Andrew D. Miller, Elizabeth D. Mynatt, and Lena Mamykina. 2015. No Longer Wearing: Investigating the Abandonment of Personal Health-tracking Technologies on Craigslist. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (UbiComp '15), 647–658. DOI:https://doi.org/10.1145/2750858.2807554

[15] R Collins, R Peto, S MacMahon, J Godwin, N Qizilbash, R Collins, S MacMahon, P Hebert, K.A Eberlein, J.O Taylor, C.H Hennekens, N.H Fiebach, N Qizilbash, and C.H Hennekens. 1990. Blood pressure, stroke, and coronary heart disease. *The Lancet* 335, 8693 (April 1990), 827–838. DOI:https://doi.org/10.1016/0140-6736(90)90944-Z

[16] John Concato, Nirav Shah, and Ralph I. Horwitz. 2000. Randomized, Controlled Trials, Observational Studies, and the Hierarchy of Research Designs. *New England Journal of Medicine* 342, 25 (June 2000), 1887–1892. DOI:https://doi.org/10.1056/NEJM200006223422507

[17] Sunny Consolvo. 2012. Designing for Healthy Lifestyles: Design Considerations for Mobile Technologies to Encourage Consumer Health and Wellness. *Foundations and Trends® in Human–Computer Interaction* 6, 3–4 (2012), 167–315. DOI:https://doi.org/10.1561/1100000040

[18] Sunny Consolvo. 2012. Designing for Healthy Lifestyles: Design Considerations for Mobile Technologies to Encourage Consumer Health and Wellness. *Foundations and Trends® in Human–Computer Interaction* 6, 3–4 (2012), 167–315. DOI:https://doi.org/10.1561/1100000040

[19] Sunny Consolvo, Predrag Klasnja, David W. McDonald, Daniel Avrahami, Jon Froehlich, Louis LeGrand, Ryan Libby, Keith Mosher, and James A. Landay. 2008.

Flowers or a robot army?: encouraging awareness & activity with personal, mobile displays. 54. DOI:https://doi.org/10.1145/1409635.1409644

[20] Sunny Consolvo, David W. McDonald, Tammy Toscos, Mike Y. Chen, Jon Froehlich, Beverly Harrison, Predrag Klasnja, Anthony LaMarca, Louis LeGrand, Ryan Libby, Ian Smith, and James A. Landay. 2008. Activity Sensing in the Wild: A Field Trial of Ubifit Garden. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '08), 1797–1806. DOI:https://doi.org/10.1145/1357054.1357335

[21] John O. Cooper, Timothy E. Heron, and William L. Heward. 1987. *Applied behavior analysis*. Merrill Publishing Co, Columbus, OH, England.

[22] Dan Cosley, Elizabeth Churchill, Jodi Forlizzi, and Sean A. Munson. 2017. Introduction to This Special Issue on the Lived Experience of Personal Informatics. *Human–Computer Interaction* 32, 5–6 (November 2017), 197–207. DOI:https://doi.org/10.1080/07370024.2017.1324787

[23] Andy Crabtree, Steve Benford, Chris Greenhalgh, Paul Tennent, Matthew Chalmers, and Barry Brown. 2006. Supporting ethnographic studies of ubiquitous computing in the wild. In *Proceedings of the 6th ACM conference on Designing Interactive systems - DIS '06*, 60. DOI:https://doi.org/10.1145/1142405.1142417

[24] Daniel A. Epstein, An Ping, James Fogarty, and Sean A. Munson. 2015. A Lived Informatics Model of Personal Informatics. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (UbiComp '15), 731–742. DOI:https://doi.org/10.1145/2750858.2804250

[25] Daniel Epstein, Felicia Cordeiro, Elizabeth Bales, James Fogarty, and Sean Munson. 2014. Taming Data Complexity in Lifelogs: Exploring Visual Cuts of Personal Informatics Data. In *Proceedings of the 2014 Conference on Designing Interactive Systems* (DIS '14), 667–676. DOI:https://doi.org/10.1145/2598510.2598558

[26] Rhian E. Evans, Henrietta O. Fawole, Stephanie A. Sheriff, Philippa M. Dall, P. Margaret Grant, and Cormac G. Ryan. 2012. Point-of-Choice Prompts to Reduce Sitting Time at Work. *American Journal of Preventive Medicine* 43, 3 (September 2012), 293–297. DOI:https://doi.org/10.1016/j.amepre.2012.05.010

[27] Louis Faust, Rachael Purta, David Hachen, Aaron Striegel, Christian Poellabauer, Omar Lizardo, and Nitesh V. Chawla. 2017. Exploring Compliance: Observations from a Large Scale Fitbit Study. In *Proceedings of the 2nd International Workshop on Social Sensing - SocialSens'17*, 55–60. DOI:https://doi.org/10.1145/3055601.3055608

[28] Bj Fogg. 2009. A behavior model for persuasive design. In *Proceedings of the 4th International Conference on Persuasive Technology - Persuasive '09*, 1. DOI:https://doi.org/10.1145/1541948.1541999

[29] Thomas Fritz, Elaine M. Huang, Gail C. Murphy, and Thomas Zimmermann. 2014. Persuasive Technology in the Real World: A Study of Long-term Use of Activity Sensing Devices for Fitness. In *Proceedings of the SIGCHI Conference on*

Human Factors in Computing Systems (CHI ’14), 487–496. DOI:https://doi.org/10.1145/2556288.2557383

[30] Nanna Gorm and Irina Shklovski. 2017. Participant Driven Photo Elicitation for Understanding Activity Tracking: Benefits and Limitations. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (CSCW ’17), 1350–1361. DOI:https://doi.org/10.1145/2998181.2998214

[31] Rúben Gouveia and Evangelos Karapanos. 2013. Footprint Tracker: Supporting Diary Studies with Lifelogging. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI ’13), 2921–2930. DOI:https://doi.org/10.1145/2470654.2481405

[32] Rúben Gouveia, Evangelos Karapanos, and Marc Hassenzahl. 2015. How Do We Engage with Activity Trackers?: A Longitudinal Study of Habito. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (UbiComp ’15), 1305–1316. DOI:https://doi.org/10.1145/2750858.2804290

[33] Rúben Gouveia, Evangelos Karapanos, and Marc Hassenzahl. 2018. Activity Tracking *in vivo*. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI ’18*, 1–13. DOI:https://doi.org/10.1145/3173574.3173936

[34] Rúben Gouveia, Fábio Pereira, Ana Caraban, Sean A. Munson, and Evangelos Karapanos. 2015. You have 5 seconds: designing glanceable feedback for physical activity trackers. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers - UbiComp ’15*, 643–647. DOI:https://doi.org/10.1145/2800835.2809437

[35] Rúben Gouveia, Fábio Pereira, Evangelos Karapanos, Sean A. Munson, and Marc Hassenzahl. 2016. Exploring the Design Space of Glanceable Feedback for Physical Activity Trackers. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (UbiComp ’16), 144–155. DOI:https://doi.org/10.1145/2971648.2971754

[36] Regina Guthold, Gretchen A Stevens, Leanne M Riley, and Fiona C Bull. 2018. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1·9 million participants. *The Lancet Global Health* 6, 10 (October 2018), e1077–e1086. DOI:https://doi.org/10.1016/S2214-109X(18)30357-7

[37] Pedro C Hallal, Lars Bo Andersen, Fiona C Bull, Regina Guthold, William Haskell, and Ulf Ekelund. 2012. Global physical activity levels: surveillance progress, pitfalls, and prospects. *The Lancet* 380, 9838 (July 2012), 247–257. DOI:https://doi.org/10.1016/S0140-6736(12)60646-1

[38] Daniel Harrison, Paul Marshall, Nadia Bianchi-Berthouze, and Jon Bird. 2015. Activity Tracking: Barriers, Workarounds and Customisation. In *Proceedings of the*

2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp ’15), 617–621. DOI:https://doi.org/10.1145/2750858.2805832

[39] Catherine Hoffman. 1996. Persons With Chronic Conditions: Their Prevalence and Costs. *JAMA* 276, 18 (November 1996), 1473. DOI:https://doi.org/10.1001/jama.1996.03540180029029

[40] Nassim Jafarinaimi, Jodi Forlizzi, Amy Hurst, and John Zimmerman. 2005. Breakaway: an ambient display designed to change human behavior. In *CHI ’05 extended abstracts on Human factors in computing systems - CHI ’05*, 1945. DOI:https://doi.org/10.1145/1056808.1057063

[41] Evangelos Karapanos. 2016. Designing for Different Stages in Behavior Change. *arXiv:1603.01369 [cs]* (March 2016). Retrieved February 15, 2019 from http://arxiv.org/abs/1603.01369

[42] Evangelos Karapanos, Rúben Gouveia, Marc Hassenzahl, and Jodi Forlizzi. 2016. Wellbeing in the Making: Peoples’ Experiences with Wearable Activity Trackers. *Psychology of Well-Being* 6, 1 (June 2016), 4. DOI:https://doi.org/10.1186/s13612-016-0042-6

[43] Morwenna Kirwan, Mitch J Duncan, Corneel Vandelanotte, and W Kerry Mummery. 2012. Using Smartphone Technology to Monitor Physical Activity in the 10,000 Steps Program: A Matched Case–Control Trial. *Journal of Medical Internet Research* 14, 2 (April 2012), e55. DOI:https://doi.org/10.2196/jmir.1950

[44] Predrag Klasnja, Sunny Consolvo, and Wanda Pratt. 2011. How to Evaluate Technologies for Health Behavior Change in HCI Research. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI ’11), 3063–3072. DOI:https://doi.org/10.1145/1978942.1979396

[45] Predrag Klasnja, Eric B. Hekler, Elizabeth V. Korinek, John Harlow, and Sonali R. Mishra. 2017. Toward Usable Evidence: Optimizing Knowledge Accumulation in HCI Research on Health Behavior Change. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI ’17*, 3071–3082. DOI:https://doi.org/10.1145/3025453.3026013

[46] Predrag Klasnja, Eric B. Hekler, Saul Shiffman, Audrey Boruvka, Daniel Almirall, Ambuj Tewari, and Susan A. Murphy. 2015. Microrandomized trials: An experimental design for developing just-in-time adaptive interventions. *Health Psychology* 34, Suppl (December 2015), 1220–1228. DOI:https://doi.org/10.1037/hea0000305

[47] Predrag Klasnja and Wanda Pratt. 2012. Healthcare in the pocket: Mapping the space of mobile-phone health interventions. *Journal of Biomedical Informatics* 45, 1 (February 2012), 184–198. DOI:https://doi.org/10.1016/j.jbi.2011.08.017

[48] Margaret J. Knapp. 1994. A 30-Week Randomized Controlled Trial of High-Dose Tacrine in Patients With Alzheimer’s Disease. *JAMA: The Journal of the American Medical Association* 271, 13 (April 1994), 985. DOI:https://doi.org/10.1001/jama.1994.03510370037029

[49] Nicholas Lane, Mashfiqui Mohammod, Mu Lin, Xiaochao Yang, Hong Lu, Shahid Ali, Afsaneh Doryab, Ethan Berke, Tanzeem Choudhury, and Andrew Campbell.

2011. BeWell: A Smartphone Application to Monitor, Model and Promote Wellbeing. In *Proceedings of the 5th International ICST Conference on Pervasive Computing Technologies for Healthcare*. DOI:https://doi.org/10.4108/icst.pervasivehealth.2011.246161

[50] Amanda Lazar, Christian Koehler, Joshua Tanenbaum, and David H. Nguyen. 2015. Why We Use and Abandon Smart Devices. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (UbiComp '15), 635–646. DOI:https://doi.org/10.1145/2750858.2804288

[51] Ian Li, Anind Dey, and Jodi Forlizzi. 2010. A Stage-based Model of Personal Informatics Systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '10), 557–566. DOI:https://doi.org/10.1145/1753326.1753409

[52] Ian Li, Anind K. Dey, and Jodi Forlizzi. 2012. Using context to reveal factors that affect physical activity. *ACM Transactions on Computer-Human Interaction* 19, 1 (March 2012), 1–21. DOI:https://doi.org/10.1145/2147783.2147790

[53] James J. Lin, Lena Mamykina, Silvia Lindtner, Gregory Delajoux, and Henry B. Strub. 2006. Fish'n'Steps: Encouraging Physical Activity with an Interactive Computer Game. In *UbiComp 2006: Ubiquitous Computing*, Paul Dourish and Adrian Friday (eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 261–278. DOI:https://doi.org/10.1007/11853565_16

[54] Edwin A. Locke and Gary P. Latham. 2002. Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist* 57, 9 (2002), 705–717. DOI:https://doi.org/10.1037/0003-066X.57.9.705

[55] David R Lubans, Philip J Morgan, Deborah Dewar, Clare E Collins, Ronald C Plotnikoff, Anthony D Okely, Marijka J Batterham, Tara Finn, and Robin Callister. 2010. The Nutrition and Enjoyable Activity for Teen Girls (NEAT girls) randomized controlled trial for adolescent girls from disadvantaged secondary schools: rationale, study protocol, and baseline results. *BMC Public Health* 10, 1 (December 2010). DOI:https://doi.org/10.1186/1471-2458-10-652

[56] Deborah Lupton. 2016. You are Your Data: Self-Tracking Practices and Concepts of Data. In *Lifelogging*, Stefan Selke (ed.). Springer Fachmedien Wiesbaden, Wiesbaden, 61–79. DOI:https://doi.org/10.1007/978-3-658-13137-1_4

[57] Deborah Lupton. 2018. How do data come to matter? Living and becoming with personal data. *Big Data & Society* 5, 2 (July 2018), 205395171878631. DOI:https://doi.org/10.1177/2053951718786314

[58] Deborah Lupton, Sarah Pink, Christine Heyes LaBond, and Shanti Sumartojo. 2018. Digital Traces in Context: Personal Data Contexts, Data Sense, and Self-Tracking Cycling. *International Journal of Communications Special Issue on Digital Traces in Context, Vol 12, 647-665* (2018). Retrieved April 20, 2019 from https://openresearch-repository.anu.edu.au/handle/1885/154629

[59] Alessandro Marcengo and Amon Rapp. Visualization of Human Behavior data: the Quantified Self. 27.

[60] Jochen Meyer, Merlin Wasmann, Wilko Heuten, Abdallah El Ali, and Susanne C.J. Boll. 2017. Identification and Classification of Usage Patterns in Long-Term Activity Tracking. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (CHI '17), 667–678. DOI:https://doi.org/10.1145/3025453.3025690

[61] Sean Munson. *Mindfulness, Reflection, and Persuasion in Personal Informatics*.

[62] Sean Munson and Sunny Consolvo. 2012. Exploring Goal-setting, Rewards, Self-monitoring, and Sharing to Motivate Physical Activity. In *Proceedings of the 6th International Conference on Pervasive Computing Technologies for Healthcare*. DOI:https://doi.org/10.4108/icst.pervasivehealth.2012.248691

[63] Inbal Nahum-Shani, Eric B. Hekler, and Donna Spruijt-Metz. 2015. Building health behavior models to guide the development of just-in-time adaptive interventions: A pragmatic framework. *Health Psychology* 34, Suppl (December 2015), 1209–1219. DOI:https://doi.org/10.1037/hea0000306

[64] Davide Neves, Donovan Costa, Marcio Oliveira, Ruben Jardim, Ruben Gouveia, and Evangelos Karapanos. 2016. Motivating Healthy Water Intake through Prompting, Historical Information, and Implicit Feedback. *arXiv:1603.01367 [cs]* (March 2016). Retrieved May 3, 2019 from http://arxiv.org/abs/1603.01367

[65] Harri Oinas-Kukkonen and Marja Harjumaa. 2008. A Systematic Framework for Designing and Evaluating Persuasive Systems. In *Persuasive Technology*, Harri Oinas-Kukkonen, Per Hasle, Marja Harjumaa, Katarina Segerståhl and Peter Øhrstrøm (eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 164–176. DOI:https://doi.org/10.1007/978-3-540-68504-3_15

[66] Tiago Ornelas, Ana Caraban, Rúben Gouveia, and Evangelos Karapanos. 2015. CrowdWalk: leveraging the wisdom of the crowd to inspire walking activities. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers - UbiComp '15*, 213–216. DOI:https://doi.org/10.1145/2800835.2800923

[67] Antti Oulasvirta, Tye Rattenbury, Lingyi Ma, and Eeva Raita. 2012. Habits make smartphone use more pervasive. *Pers Ubiquit Comput* 16, 1 (January 2012), 105–114. DOI:https://doi.org/10.1007/s00779-011-0412-2

[68] N Owen, A Bauman, and W Brown. 2008. Too much sitting: a novel and important predictor of chronic disease risk? *British Journal of Sports Medicine* 43, 2 (December 2008), 81–83. DOI:https://doi.org/10.1136/bjsm.2008.055269

[69] Endeavour Partners. 2017. Inside Wearables: How the science of human behavior change offers the secret to long-term…. *Medium*. Retrieved May 1, 2019 from https://medium.com/@endeavourprtnrs/inside-wearable-how-the-science-of-human-behavior-change-offers-the-secret-to-long-term-engagement-a15b3c7d4cf3

[70] Josée Poirier, Wendy L Bennett, Gerald J Jerome, Nina G Shah, Mariana Lazo, Hsin-Chieh Yeh, Jeanne M Clark, and Nathan K Cobb. 2016. Effectiveness of an Activity Tracker- and Internet-Based Adaptive Walking Program for Adults: A

Randomized Controlled Trial. *Journal of Medical Internet Research* 18, 2 (February 2016), e34. DOI:https://doi.org/10.2196/jmir.5295

[71] James O. Prochaska, Carlo C. DiClemente, and John C. Norcross. 1992. In search of how people change: Applications to addictive behaviors. *American Psychologist* 47, 9 (1992), 1102–1114. DOI:https://doi.org/10.1037/0003-066X.47.9.1102

[72] James O. Prochaska and Wayne F. Velicer. 1997. The Transtheoretical Model of Health Behavior Change. *American Journal of Health Promotion* 12, 1 (September 1997), 38–48. DOI:https://doi.org/10.4278/0890-1171-12.1.38

[73] Amon Rapp, Federica Cena, Judy Kay, Bob Kummerfeld, Frank Hopfgartner, Till Plumbaum, Jakob Eg Larsen, Daniel A. Epstein, and Rúben Gouveia. 2016. New frontiers of quantified self 2: going beyond numbers. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct - UbiComp '16*, 506–509. DOI:https://doi.org/10.1145/2968219.2968331

[74] Amon Rapp, Federica Cena, Judy Kay, Bob Kummerfeld, Frank Hopfgartner, Till Plumbaum, Jakob Eg Larsen, Daniel A. Epstein, and Rúben Gouveia. 2017. New frontiers of quantified self 3: exploring understudied categories of users. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers on - UbiComp '17*, 861–864. DOI:https://doi.org/10.1145/3123024.3124456

[75] Amon Rapp and Maurizio Tirassa. 2017. Know Thyself: A Theory of the Self for Personal Informatics. *Human–Computer Interaction* 32, 5–6 (November 2017), 335–380. DOI:https://doi.org/10.1080/07370024.2017.1285704

[76] John Rooksby, Mattias Rost, Alistair Morrison, and Matthew Chalmers. 2014. Personal Tracking As Lived Informatics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '14), 1163–1172. DOI:https://doi.org/10.1145/2556288.2557039

[77] Patrick C. Shih, Kyungsik Han, Erika Shehan Poole, Mary Beth Rosson, and John M. Carroll. 2015. Use and Adoption Challenges of Wearable Activity Trackers. (March 2015). Retrieved January 25, 2019 from https://www.ideals.illinois.edu/handle/2142/73649

[78] Katherine A. Stamatakis, Timothy D. McBride, and Ross C. Brownson. 2010. Communicating Prevention Messages to Policy Makers: The Role of Stories in Promoting Physical Activity. *Journal of Physical Activity and Health* 7, s1 (March 2010), S99–S107. DOI:https://doi.org/10.1123/jpah.7.s1.s99

[79] Melanie Swan. 2013. The Quantified Self: Fundamental Disruption in Big Data Science and Biological Discovery. *Big Data* 1, 2 (June 2013), 85–99. DOI:https://doi.org/10.1089/big.2012.0002

[80] Alice Thudt, Dominikus Baur, Samuel Huron, and Sheelagh Carpendale. 2016. Visual Mementos: Reflecting Memories with Personal Data. *IEEE Trans. Visual. Comput. Graphics* 22, 1 (January 2016), 369–378. DOI:https://doi.org/10.1109/TVCG.2015.2467831

[81] Frank A. Treiber, Tom Baranowski, David S. Braden, William B. Strong, Maurice Levy, and Willie Knox. 1991. Social support for exercise: Relationship to physical activity in young adults. *Preventive Medicine* 20, 6 (November 1991), 737–750. DOI:https://doi.org/10.1016/0091-7435(91)90068-F

[82] Adam Gilden Tsai and Thomas A. Wadden. 2005. Systematic Review: An Evaluation of Major Commercial Weight Loss Programs in the United States. *Annals of Internal Medicine* 142, 1 (January 2005), 56. DOI:https://doi.org/10.7326/0003-4819-142-1-200501040-00012

[83] Catrine Tudor-Locke, David R. Bassett, Ann M. Swartz, Scott J. Strath, Brian B. Parr, Jared P. Reis, Katrina D. DuBose, and Barbara E. Ainsworth. 2004. A Preliminary study of one year of pedometer self-monitoring. *Annals of Behavioral Medicine* 28, 3 (October 2004), 158–162. DOI:https://doi.org/10.1207/s15324796abm2803_3

[84] Catrine Tudor-Locke and Lesley Lutes. 2009. Why Do Pedometers Work?: A Reflection upon the Factors Related to Successfully Increasing Physical Activity. *Sports Medicine* 39, 12 (December 2009), 981–993. DOI:https://doi.org/10.2165/11319600-000000000-00000

[85] Kathleen D. Vohs and Roy F. Baumeister. 2016. *Handbook of Self-Regulation, Third Edition: Research, Theory, and Applications*. Guilford Publications.

[86] Julie B. Wang, Lisa A. Cadmus-Bertram, Loki Natarajan, Martha M. White, Hala Madanat, Jeanne F. Nichols, Guadalupe X. Ayala, and John P. Pierce. 2015. Wearable Sensor/Device (Fitbit One) and SMS Text-Messaging Prompts to Increase Physical Activity in Overweight and Obese Adults: A Randomized Controlled Trial. *Telemedicine and e-Health* 21, 10 (October 2015), 782–792. DOI:https://doi.org/10.1089/tmj.2014.0176

[87] Darren ER Warburton, Sarah Charlesworth, Adam Ivey, Lindsay Nettlefold, and Shannon SD Bredin. 2010. A systematic review of the evidence for Canada's Physical Activity Guidelines for Adults. *Int J Behav Nutr Phys Act* 7, 1 (2010), 39. DOI:https://doi.org/10.1186/1479-5868-7-39

[88] Muhammad Usman Warraich. 2016. Wellness Routines with Wearable Activity Trackers: A Systematic Review. *MCIS 2016 Proceedings* (January 2016). Retrieved from https://aisel.aisnet.org/mcis2016/35

[89] 1948. Streptomycin Treatment of Pulmonary Tuberculosis: A Medical Research Council Investigation. *BMJ* 2, 4582 (October 1948), 769–782. DOI:https://doi.org/10.1136/bmj.2.4582.769

[90] Gartner Survey Shows Wearable Devices Need to Be More Useful. Retrieved January 25, 2019 from https://www.gartner.com/en/newsroom/press-releases/2016-12-07-gartner-survey-shows-wearable-devices-need-to-be-more-useful

[91] Quantified Self Guide. Retrieved February 6, 2019 from http://quantifiedself.com/guide/tag/gadget

[92] Connected Digital Fitness Device Growth Continues. Retrieved February 6, 2019 from https://www.npd.com/wps/portal/npd/us/news/press-

releases/2016/Year-Over-Year-connected-Digital-Fitness-Device-Growth-Continues-with-a-40-Percent-Dollar-Increase-in-the-First-Quarter-of-2016-According-to-NPD/

[93] FitLinxx Pebble Frequently Asked Questions Mission Health Store. Retrieved March 15, 2019 from https://store.mission-health.org/wellness/fitlinxx-pebble-frequently-asked-questions.html#_Toc384302725

[94] Walkadoo. Retrieved March 15, 2019 from https://walkadoo.meyouhealth.com/