TECHNISCHE
UNIVERSITÄT
DARMSTADT

PERVASIVE BEHAVIOR INTERVENTIONS

Using Mobile Devices for Overcoming Barriers for Physical Activity

Vom Fachbereich Elektrotechnik und Informationstechnik
der Technischen Universität Darmstadt
zur Erlangung des akademischen Grades eines
Doktor-Ingenieurs (Dr.-Ing.)
genehmigte Dissertation
von

DIPL.-INF. UNIV. TIM ALEXANDER DUTZ
Geboren am 20. Juli 1978 in Darmstadt

Vorsitz: Prof. Dr. techn. Heinz Koeppl
Referent: Prof. Dr.-Ing. habil. Ralf Steinmetz
Korreferent: Prof. Dr. rer. nat. Rainer Malaka

Tag der Einreichung: 14. September 2016
Tag der Disputation: 28. November 2016

Hochschulkennziffer D17
Darmstadt 2017

Für meine Eltern

Abstract

Extensive cohort studies show that physical inactivity is likely to have negative consequences for one's health. The World Health Organization thus recommends a minimum of thirty minutes of medium-intensity physical activity per day, an amount that can easily be reached by doing some brisk walking or leisure cycling. Recently, a Taiwanese-American team of scientists was able to prove that even less effort is required for positive health effects and that as little as fifteen minutes of physical activity per day will increase one's life expectancy by up to three years on the average. However, simply spreading this knowledge is not sufficient. Roughly one in three Europeans and US-Americans does not even meet the minimum recommendations for physical activity, although the majority of these people is aware of the damage that their behavior may do to their health. And this 'willful wrongdoing' does not only concern individuals: Due to the large number of inactive people, the problem of sedentary behavior affects societies as a whole, not the least by increasing public health costs.

But if it is not a lack of knowledge that causes this problem, what is? And what can be done to stimulate leisure-time physical activity? The Fogg Behavior Model (FBM), developed by psychologist and Stanford-lecturer B.J. Fogg, explains the factors that determine whether or not a given person will show a desired behavior. The core components of the FBM include a trigger that can be perceived by the target person and that she associates with the desired behavior, as well as her ability and motivation for this behavior at the time when the trigger reaches her. If the combined amount of ability and motivation exceeds a lower limit, the so-called activation threshold, then the triggered person will behave in the desired way; otherwise, she will not. Based on the understanding of human behavior that the FBM conveys, this thesis focuses on the question of how mobile devices can assist people in reaching the minimum amount of daily physical activity that is required for health benefits.

An in-depth analysis of the problem reveals that of the three possible strategies – trying to increase a user's ability for leisure-time physical activity, trying to increase her motivation for the same, and trying to increase her short-term awareness for its necessity and feasibility through triggers – the creation of adaptive triggers is the most promising approach. This task in turn consists of several sub-problems, such as the problem of how to recognize the user's current contextual situation, the problem of how to decide, whether or not the recognized situation is suited for an activation attempt, and the problem of interacting with the user in those cases in which an activation attempt seems worthwhile. Learning from the user's behavior and understanding her preferences and constraints is the key factor in the creation of accurate and reliable intervention mechanisms. To this end, smartphone sensors, wearables, and Web services are utilized for collecting information about the state of the user and her environment. This data is then analyzed by a supervised learning machine which, based on prior experience, estimates the probability for a successful activation attempt in the current situation. Ideally, the learner will identify a *kairotic* moment: A situation, in which a trigger is bound to initiate the desired behavior. If it does, it reaches out to the user.

Multiple types of such triggering mechanisms were embedded into the mobile exergame 'Twostone', an application that requires brisk walking or easy running from its users. During a field study with thirty participants, the performances of these different approaches were compared against one another. The study revealed a surprising result: Not the most-knowledgeable intervention mechanism emerged as a winner, but it was rather the triggering variant that relied on a reduced number of contextual information to achieve both the highest triggering success rates and the best user acceptance. The study also showed that intervention mechanisms can indeed increase the prevalence of a desired behavior, but only if the user has a positive attitude towards the respective activity. As such, both the conceptual model for technology-based interventive measures and the evaluation results that are presented in this thesis offer valuable insights for developers of devices and applications that aim to foster desired behaviors in general and increased levels of daily physical activity in particular.

Kurzfassung

Die gesundheitlichen Folgen körperlicher Inaktivität sind durch umfangreiche epidemiologische Kohortenstudien hinreichend belegt. Entsprechend empfiehlt die Weltgesundheitsorganisation jedem Erwachsenen ein Mindestmaß von täglich dreißig Minuten körperlicher Aktivität mittlerer Intensität, also beispielsweise zügiges Gehen oder gemütliches Radfahren. Einem taiwanesisch-amerikanischen Forscherteam gelang der Nachweis, dass bereits die Hälfte dieser Menge, also nur fünfzehn Minuten, genügen, um die eigene Lebenserwartung um bis zu drei Jahre zu erhöhen. Allerdings führt das Wissen über den Zusammenhang zwischen Bewegung und Gesundheit nicht zwingend zu einer Erhöhung des Aktivitätslevels. Etwa ein Drittel der europäischen und US-amerikanischen Bevölkerung leistet nicht einmal das erforderliche Mindestmaß, obwohl das Wissen um die möglichen Folgen dieses Verhaltens meist vorhanden ist. In der Häufung hat dieses bewusste Fehlverhalten Einzelner auch Auswirkungen auf die Gesamtgesellschaft, insbesondere durch einen Anstieg staatlicher Gesundheitsausgaben.

Wenn Bewegungsmangel aber nicht auf mangelnde Informiertheit zurückzuführen ist, worauf dann? Und welche alternativen Maßnahmen können helfen? Das Foggsche Verhaltensmodell (FBM) des an der Stanford-Universität lehrenden Psychologen B.J. Fogg beschreibt die Faktoren, die darüber entscheiden, ob eine Person ein gewünschtes Zielverhalten zeigt. Die wesentlichen Komponenten des FBM sind ein wahrnehmbarer und gedanklich mit dem Zielverhalten verbundener Auslöser, der Trigger, sowie Befähigung und Motivation für das gewünschte Verhalten zum Zeitpunkt des Auftretens eines solchen Triggers. Sind Befähigung und Motivation in ausreichendem Maße vorhanden und überschreiten in Kombination die sogenannte Aktivierungsgrenze, so zeigt die ‚getriggerte' Person das Zielverhalten – andernfalls nicht. Basierend auf diesem Verständnis menschlichen Verhaltens befasst sich die vorliegende Arbeit mit der Frage, wie mobile Endgeräte dazu genutzt werden können, um Personen zuverlässig zum erforderlichen Mindestmaß an körperlicher Aktivität anzuregen.

Die Analyse der Problemstellung macht deutlich, dass von den drei möglichen Ansätzen – Steigerung der Befähigung für körperliche Aktivität, Steigerung der Motivation und Steigerung des Bewusstseins für Notwendigkeit und Machbarkeit durch den Einsatz von Triggern – die Entwicklung adaptiver Triggering-Mechanismen am vielversprechendsten ist. Diese Herausforderung lässt sich ihrerseits in mehrere Teilprobleme unterteilen, etwa das Problem des Erfassens der gegenwärtigen Situation, das Problem der Entscheidung, ob ein Interventionsversuch unternommen werden sollte, sowie das Problem der eigentlichen Nutzerinteraktion. Aus dem Nutzerverhalten zu lernen, um Vorlieben und Möglichkeiten des Nutzers richtig einzuschätzen, ist dabei die Schlüsselfähigkeit erfolgreicher Maßnahmen. Vor diesem Hintergrund wird von Smartphonesensorik, Wearables und Web Services Gebrauch gemacht, um Informationen über den Nutzer und die aktuelle Situation zu erhalten. Ein überwachter Lerner analysiert diese Daten und entschiedet auf Basis zurückliegender Erfahrungen, ob ein Aktivierungsversuch erfolgen sollte. Im Idealfall stellt der Lerner dabei fest, dass ein *kairotischer Moment* vorliegt: eine Situation, in der ein Trigger zuverlässig das gewünschte Verhalten auslöst.

Mehrere unterschiedliche Triggering-Mechanismen wurden in das mobile Fitnessspiel ‚Twostone' eingebettet, eine Anwendung, die vom Nutzer zügiges Gehen oder lockeres Laufen erfordert. In einer umfangreichen Feldstudie mit dreißig Teilnehmern wurden alle Varianten miteinander verglichen und ein unerwartetes Ergebnis erzielt: nicht derjenige Mechanismus mit der umfangreichsten Wissensbasis erreichte die höchsten Erfolgsraten und die beste Nutzerakzeptanz, sondern eine Variante mit einer reduzierten Menge an Kontextwissen. Zudem zeigt die Evaluation, dass intelligente Trigger tatsächlich die Häufigkeit eines gewünschten Verhaltens erhöhen können, aber nur, sofern beim Nutzer bereits eine positive Grundeinstellung gegenüber diesem Zielverhalten vorhanden ist. Mit diesen Erkenntnissen leistet die vorliegende Arbeit einen wichtigen Beitrag zur zukünftigen Entwicklung interventiver Maßnahmen und insbesondere diejenigen technischen Anwendungen, die Personen dabei unterstützen sollen, das für positive Gesundheitseffekte erforderliche Mindestmaß an körperlicher Aktivität zu erfüllen, können darauf aufbauen.

Contents

1. Introduction

Maintaining physical health is a key factor in maintaining overall wellbeing and quality-of-life. However, having knowledge of behavior that promotes one's health and acting in a way to actually benefit from the positive effects of such behavior are two pairs of shoes. More often than not, short-term wants such as the desire for comfort, the longing for entertainment, or the craving for unhealthy food dominate our long-term health goals. For many people, health only becomes a real concern when it is lost.

Practicing frequent physical activity has been shown to significantly increase the chance of staying healthy. From the results of their extensive cohort study involving more than 400,000 individuals, a team of Taiwanese and U.S. American scientists deduced that as little as 15 minutes of moderate-intensity physical activity per day will extend an individual's lifespan for three years on the average [WWT+11]. On the contrary, sedentary behavior heightens the risk of suffering from various chronic diseases, such as coronary artery disease, hypertension, type 2 diabetes, colon cancer, and breast cancer [KGS00]. Despite a general awareness for the importance of leading an active life, barriers exist that keep many people from being physically active [OB92]. This work focuses on utilizing the means of contemporary technology – mainly mobile devices such as smartphones and wearables – to help overcome these barriers.

1.1 Motivation

The 2008-published, evidence-based report of the U.S. Physical Activity Guidelines Advisory Committee explains that a number of health benefits can be achieved by performing 30 minutes of physical activity a day for five days a week [US08]. According to this report, the health benefits that come with such amounts of physical activity include "*lower risk for all-cause mortality, coronary heart disease, stroke, hypertension, and type 2 diabetes*". Doubling the minimal recommendations and performing up to 300 minutes (five hours) of physical activity per week results in "*significantly lower rates of colon and breast cancer and the prevention of unhealthy weight gain or significant weight loss*" [US08, p.31/A-5]. In their global guidelines published in 2010, the World Health Organization (WHO) largely adopts the recommendations of the U.S. Advisory Committee and suggests at least 150 minutes of moderate-intensity physical activity per week, or 75 minutes of vigorous-intensity physical activity per week, or any combination of moderate-intensity and vigorous-intensity physical activities that lies in between these bounds [WHO10].

The terms 'moderate-intensity physical activity' and 'vigorous-intensity physical activity' are implicit references to the Metabolic Standard of Tasks (MET). This coding scheme was conceptualized by Stanford university emeritus William

Table 1: MET Value Table.

Physical Activity	**MET**
Sleeping	0.9
Resting Metabolic Rate (watching TV)	1.0
Working on PC	1.5
Tai Chi	3.0
Cleaning windows	3.2
Bowling	3.9
Walking (brisk, level ground)	4.3
Skiing (general)	7.0
Bicycling (general)	7.5
Running at 14 mph	23.0

MET values of different physical activities, selected from a total of 821 entries as specified in [AHH+11]

Haskell [AHL+93] and establishes a classification hierarchy of the intensities of all kinds of physical activity, whereby the intensity of an activity is derived from its rate of energy expenditure relative to the so-called Resting Metabolic Rate (RMR). The RMR is defined as the energy expenditure while sitting quietly, for instance while watching a (not too exciting) TV show, and has an associated MET value of 1.0. A physical activity with a MET value of 3.0, such as practicing Tai Chi, thus requires three times the energy expenditure of sitting quietly. The second update of the MET classification, published in 2011 [AHH+11], updated several of the original MET values and now lists values for a total of 821 different physical activities. Table 1 shows an excerpt of this list.

Table 2: Intensity Categories.

Category	**METs**
Low-intensity	0.9 to 2.9
Moderate-intensity	3.0 to 5.9
Vigorous-intensity	6.0 to 23.0

The three general categories of physical activity, according to [WHO10]

Based on the MET value hierarchy, the WHO groups physical activities into three categories: The (not explicitly defined) low-intensity group includes all those activities that have less than three times the intensity of rest, that is, activities that have a MET value of less than 3.0 assigned to them. Moderate-intensity activities are physical activities with a MET value in between 3.0 and 5.9. And finally, any activity with a MET value of 6.0 and above qualifies as a vigorous-intensity activity [WHO10]. A frequent example for a moderate-intensity activity is brisk walking on level ground [HLP+07]. According to [AHH+11], the physical activity with the highest intensity is running at 14 miles per hour (about 22.5 kilometers per hour), which is reflected by an associated MET value of 23.0. It is worth noting that even amateur athletes cannot necessarily perform all types of vigorous-intensity activities. Rather, the maximum capacity of a normal, healthy person will usually lie at about 10 to 13 METs. Only professional athletes can reach a capacity of 20 and above [Cam13].

In the aforementioned cohort study by Wen et al. [WWT+11], the authors compared the all-cause mortality of active and inactive individuals and found that even less than the recommended 30 minutes of medium-intensity physical activity per day can have significant positive health effects. According to their findings, exercising for only 15 minutes a day results in a 14% reduced risk of all-cause mortality, and every additional 15 minutes of exercise reduce the all-cause mortality risk by an additional 4% (up to a maximum of 100 minutes of exercise a day, after which no additional health benefits were found). Figure 1 shows the relationship between the amount of daily physical activity and the reduction of the all-cause mortality risk. The authors of the study go on to point out that "*compared with individuals in the inactive group, at age 30 years, life expectancy for individuals in the low-volume activity group* [author's note: individuals exercising for only 15 minutes/day] *was 2.55 years longer for men and 3.10 years longer for women, and life expectancy in those who met the recommended amount of daily exercise* [author's note: WHO recommendation of 30 minutes/day] *was 4.21 years longer for men and 3.67 years longer for women*" [WWT+11, p.1249].

Martinez-Gonzalez et al. conducted an extensive study on the prevalence of physical inactivity in 15 European Union member states, among them Germany, France, and Italy [MVS+01]. In each country, they selected a representative sample of roughly 1,000 persons, amounting to a total of 15,239 study participants, and used questionnaires and face-to-face interviews to determine each individual's average amount of daily physical activity. The results of the survey show that the prevalence of physical inactivity in Europe is comparable to that of the U.S., as described in [PMB99], in that about a third of the population must be considered inactive. As an interesting side note, an extreme range exists between the portion of physically inactive Finish and Portuguese (8.1% and 59.3%, respectively). In their paper, the authors use the convention MET-hours per week (MET·h·wk^{-1}) to estimate the amount and intensity of physical activity that an individual is performing per week on the average. This value is calculated by multiplying the hours dedicated to a specific physical activity per week by the MET value assigned to this specific activity. In this regard, the minimal amount of MET-hours per week required for health benefits according to [WWT+11] would be 9, and the WHO recommendation of 30 minutes of moderate-intensity physical activity per day translate to 18 MET-h·wk^{-1} (a total of 120/240 minutes per week of brisk walking with an associated MET value of roughly 4.5, numbers rounded for easier handling). Martinez-Gonzalez et al. conclude their study with the consideration that "*nevertheless, the amount of activity is low (...)*" and they go on to state that their work can be a "(...) *first step to define strategies to persuade populations to increase their physical activity*" [MVS+01, p.1142].

Successfully persuading someone to show an intended behavior means to change the person's attitude towards this specific behavior. If the person in question is aware of the meaningfulness of the desired behavior but nevertheless does not behave in such a way, then it must be assumed that barriers exist that keep her from doing so. Otherwise, if the person is not aware of the meaningfulness of the intended behavior, it may be sufficient to communicate the required knowledge. Initiating the intended behavior is then actually less a case of *persuading* and more a matter of *informing*. In this context, two questions arise: First, how widespread is the knowledge of the effects of physical inactivity on health and wellbeing and second, what barriers keep those people from being physically active that are fully aware of the negative effects of their behavior?

A worldwide study carried out with almost 20,000 university students from a total of 23 countries showed that knowledge of the effects of physical inactivity is not sufficiently widespread, even among the members of the privileged group of university attendees [HSS+04]. The authors found that in 15 of the considered 23 countries, more than half of the population was not aware of the connection between physical inactivity and the risk of suffering from heart diseases. Even more concerning, however, is another finding. The study states that "*nonetheless, knowledge was not associated with behavior*" and continues with pointing out that "*(...) knowledge of the relationship between physical activity and heart disease was not associated with leisure-time physical activity. These null findings are consistent with previous research [TOB+02]. Thus, although national economic development may relate to health knowledge, this does not translate at an individual level into participation in leisure-time physical activity.*

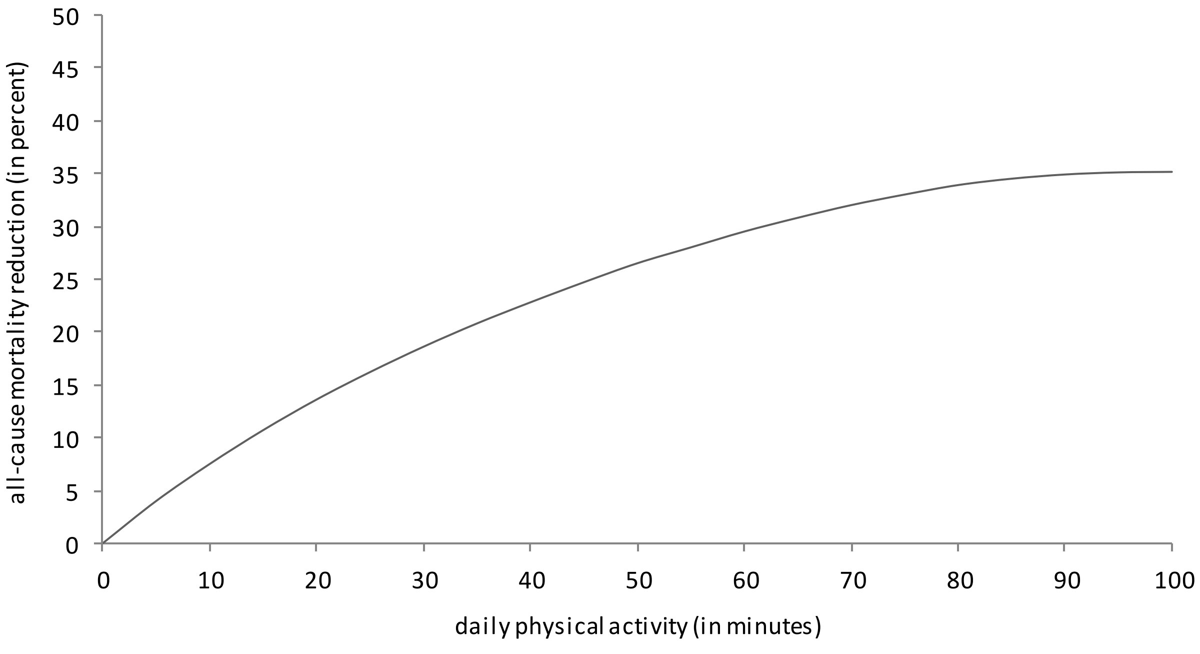


Figure 1: Daily Physical Activity and All-Cause Mortality Risk.

FIGURE NOTES – The graph shows an aggregated function of the health effects that come from performing medium-intensity and vigorous-intensity physical activity. Figure adapted from [WWT+11].

Improving knowledge about health effects should not be expected to be an effective physical activity promotion strategy, even in less developed countries" [HSS+04, p.188]. Apparently, simply creating awareness for the importance of physical activity will not necessarily result in significantly increased levels of activity.

Surveys on the sedentary behavior of Australian adults show that the top three reasons that keep people from being more active are a (self-perceived) lack of time, a lack of physical ability, and a lack of motivation, in this order [OB92]. Table 3 lists the most frequent arguments against regular exercise. A closer look reveals that indeed, *all* stated reasons for physical inactivity seem to be variations of either a perceived lack of opportunity (reasons 1, 5), a perceived lack of different kinds of ability (reasons 2, 6, 7, 8), or a lack of motivation (reasons 3, 4). By adding up the numbers [1], we find that a lack of opportunity is the main culprit (41.9% stated this as their problem), while interestingly, the lack of motivation is the least pressing issue (only 22.2% mentions). Although a response bias cannot be ruled out, it still seems reasonable to assume that in order for them to be successful, measures that aim at increasing people's amounts of physical activity should not just focus on solving the 'motivational problem', but also look into ways on how to increase their addressee's perceived amounts of opportunities.

Table 3: Reasons for Physical Inactivity.

Reported Reason		Absolute Amount	Relative Amount
1	No time	1,756	34.6
2	Physically unable	1,234	24.3
3	Don't want to exercise	679	13.4
4	Need encouragement	447	8.8
5	No chance to exercise	350	6.9
6	Exercising is too difficult	267	5.2
7	No facilities	128	2.5
8	No transport	118	2.3
9	Other reasons	365	7.1

Self-reported reasons for physical inactivity among Australian adults, n = 5,078, multiple answers possible, adapted from [OB92]

All interventive measures will be the most effective, however, if they can influence the attitude and behavior of persons at any time, and not just during specific hours of the day or week. From a technological point of view, this means that mobile devices, most notably smartphones and wearables, are the ideal tools for this task, as they have the highest 'pervasiveness quote' of all modern-day computational devices. Since the introduction of the first *iPhone* in 2007, smartphones have become incredibly prevalent. The fact that many people always carry their smartphones with them, the ongoing increase of their computational capabilities, and their contextual awareness through integrated sensors and a mobile access to Web services makes smartphones well suited for delivering health interventions [ICS+15]. Wearables such as smartwatches, wristbands, or chest straps can either be an alternative to smartphones for very specific purposes, i.e., for activity tracking, or used in conjunction with them and then function as an enhancement to their sensing and user interaction capabilities. How this interplay between smartphone and wearable can look like is demonstrated by Apple's *Activity application* for iOS-based devices. If available, it utilizes both the user's *iPhone* and her *Apple Watch* to assess physical activity levels, the user's heart rate, and other data, and it visualizes all this information to the user on request at any time. The *Activity app* even frequently urges the user to be more active with notifications, sounds, and soft vibrations of the watch. However, the way of how this feature is realized is unsatisfactory, as, for example, the *Apple Watch* also tries to activate users who are currently attending a theater performance or driving on the highway. This is just one example for how today's state-of-the-art is still far from tapping into the full technological potential of contemporary mobile devices when it comes to the development of reliable interventive applications [DGH+16].

[1] While multiple answers were allowed, adding up the numbers in order to gain a rough overview is made acceptable by the fact that less than two percent of the study participants gave two reasons and none of them gave more than two [OB92, p.307].

1.2 Thesis Contributions

This thesis presents a model for the creation of pervasive interventive measures, targeted primarily at counteracting the problem of physical inactivity with the help of mobile devices. As pointed out before, the barriers that keep people from being active can be grouped into the categories ability-related barriers, motivational barriers, and opportunity-related barriers. Analogously, the problem of how to reliably increase the physical activity levels of users with the help of contemporary technology can also be broken down into three sub-problems.

(1) CHALLENGE OF OVERCOMING ABILITY-RELATED BARRIERS – How to utilize contemporary technology to assist its users in overcoming any existing ability-related barriers that may prevent them from being more active?

(2) CHALLENGE OF OVERCOMING MOTIVATIONAL BARRIERS – How to utilize contemporary technology to sufficiently motivate users so that they are willing to perform at least the minimum amount of physical activity required for health benefits (9 MET-h·wk^{-1})?

(3) CHALLENGE OF OVERCOMING OPPORTUNITY-RELATED BARRIERS – How to utilize contemporary technology to create awareness for the importance and the feasibility of physical activity, i.e., how to use it to create reliable short-term awareness interventions ('triggers')?

As a matter of fact, these problems are not of equal importance. Rather, as will be explained in chapter three, overcoming opportunity-related barriers with the help of intelligent adaptive triggers should be the core task of any interventive measure. Consequently, the main contribution of this work is the conceptualization of an accurate mobile user triggering mechanism that encourages smartphone users to be physically active. In contrast, all considerations made on how to overcome motivational and/or ability-related barriers are merely aimed at increasing the probability of successful triggering. The mobile application *Twostone Interventive Measure* (Twostone-IM) is based on the concepts described in this thesis and was evaluated in an extensive two-week field study with more than 30 test users to demonstrate their validity. Figure 2 depicts the main building blocks of my work and how they relate to the structure of this document.

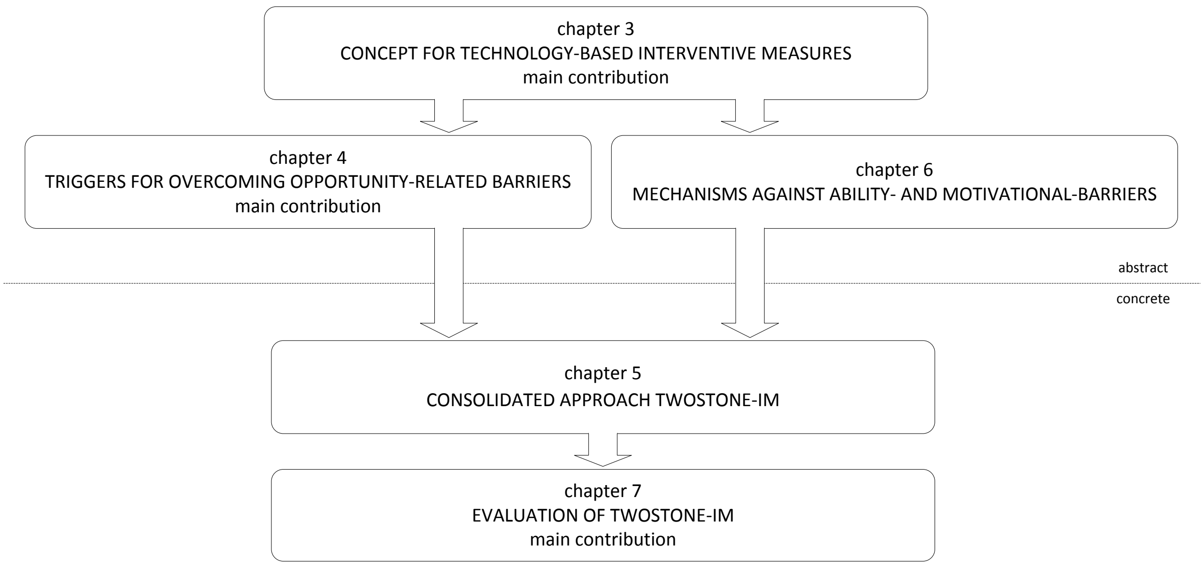


Figure 2: Thesis Contributions.

1.3 Organization and Conventions

This document follows the well-established formula for scientific technical writing: An introductory motivation chapter is followed by a description of the related work, followed by a theoretical concept for solving the problem at hand, followed by a description of this concept's actual (prototypical) implementation, followed by an evaluation of this prototype, and is finally concluded by a critical reflection on the achieved results.

More specifically, this first chapter of the thesis explains the motivation for this work. The subsequent second chapter focuses on an analysis of the related work, divided into the three categories user motivation, ability adaptation, and opportunistic interactions. The three fields are discussed from a technical standpoint, meaning that the respective sections mainly focus on the relevant technical advances in industry and science, e.g., on context-aware behavior changing smartphone applications. The third chapter discusses the problem of technology-based behavior interventions on an abstract level. The fourth and the sixth chapter focus on the three types of barriers that keep people from being physically active, and they describe specific approaches that are supposed to help overcome such barriers. It is worth noting that the fourth chapter, which focuses on the topic of creating well-timed short-term awareness interventions called 'triggers', is the more important one of the two; chapter six, discussing means to adapt to the user's ability and to increase her motivation, is more or less of supportive character. Chapter three explains the reasoning for this. The fifth chapter presents a consolidated approach that serves to unify the findings described in chapters three, four, and six within a prototypical application: The Android-based mobile exergame *Twostone* and an associated *Interventive Measure Application*; together they make up the *Twostone Interventive Measure* (Twostone-IM). Chapter seven presents the results of an end-user based study that compared the effects of several variants of this application. Finally, chapter eight concludes the thesis with a discussion and an outlook, reflecting on what was achieved and what new questions have emerged from my work.

In regard to format conventions, the layout of this document is largely dictated by the corporate design guidelines of my alma mater, the Technische Universität Darmstadt. This concerns the thesis' cover page, the font (Charter for standard text and Frontpage for headlines), the page spacing, and some other aspects. I am using *italic face* for direct quotations, also indicated by quotation marks, such as in this famous statement allegedly made by Abraham Lincoln: "*The problem with quotes found on the Internet is that they are often not true*". I am also using *italic* to highlight pivotal lines, words, or characters, such as names, formal definitions, and variable symbols. I am using neither **bold face** nor <u>underlines</u>. In-text citations are done using our institute's citation style (the KOM Citation Style, if you will): The last name initials of the first three authors followed by the last two digits of the publication year, such as in [WWT+11]. A plus sign indicates that there are actually more than three contributing authors. Sometimes I add also a page reference, i.e., in case of direct quotations, such as in [WWT+11, p.1249]. If you cannot find a specific reference index in the Bibliography, then this is probably because it is either a self-reference, or because I referenced one of the student theses that I supervised. In such cases, you will find the corresponding work in appendix D or E. Some figures and tables are adapted from other sources, which means that I redid (and possibly altered) a figure that I found somewhere else. This will be indicated by an "*adapted from*", followed by a reference index. If no reference is specified in the caption, then the respective figure is entirely my own creation. No figures in this document are directly copied from other sources. This also applies to all screenshots and images. However, for some of my figures I have used the 'Google Material Icons' (2) and the 'Android Device Art Generator' (3) with hopes to enhance my own limited design skills. Footnotes are meant to support the main text with additional information that may not be essential, but that I still find noteworthy.

I am using the generic she as a pronoun for no other reason than just because.

[2] https://design.google.com/icons/index.html
[3] https://developer.android.com/distribute/tools/promote/device-art.html

2. Related Work

The core question of this thesis is how the daily amount of physical activity can reliably be increased with the help of technology in general and with mobile devices in particular. As pointed out in the first chapter, a lack of knowledge about the potentially harmful effects of physical inactivity is not the problem. While roughly one in three Europeans and US-Americans does not even meet the minimum recommendations for physical activity [MVS+01], the majority of these people are well aware of the fact that their sedentary lifestyle is likely to have undesired consequences [HSS+04]. Apparently, barriers exist that prevent parts of the populace from being more active and technological measures that aim at increasing the physical activity levels of people must be equipped with the means to help their users overcome those barriers. In other words, they must be able to help them to achieve a desired behavior. As such, a precondition for the systematic development of such measures is an understanding of what determines human behavior.

The psychologist and Stanford University lecturer B.J. Fogg coined the term 'persuasive technology'. In his like-named monography, Fogg describes how computers can be used to change a person's attitude or behavior [Fog03]. The gist of his work is the Fogg Behavior Model, FBM for short. This conceptual model conveys the three factors that decide whether or not someone will show a desired behavior b [4]: First, the person's motivation m for behaving in the intended way; second, her ability a for doing so; and finally the occurrence of a trigger t that basically functions as a well-timed reminder. Fogg summarizes this interdependence with the equation $b = mat$ [5]. According to Fogg, the absence of a desired behavior can thus either be explained by a lack of motivation, and/or a lack of ability, and/or a lack of a trigger that calls the respective person to action. Figure 3 shows a graphical representation of the FBM.

The figure visualizes a person's motivation and ability for a specific behavior b at a given time δ [6] as the two axes on a coordinate plane. Moving away from the origin along the x- or y-axis symbolizes a person's increased ability or motivation for showing the intended behavior, respectively; moving closer to the origin translates to a lower ability or motivation. Marking a point in a significant distance from the origin thus implies that the respective person finds the desired behavior very easy to do at this point in time, and/or that she is highly motivated for performing it. For a given person, the ability and the motivation are dependent on two factors, namely the desired behavior $\bar{b}$, and the point in time δ when this behavior is supposed to occur. This becomes obvious when we compare a person's ability and motivation for showing two different behaviors at two different points in time.

Consider average Jane's ability and motivation for a 10-mile run on a Monday noon and on a Saturday early evening. While her motivation for running may be equally high (or low) on both times, her ability for doing so will usually be lower on a Monday noon than on a Saturday late afternoon, since in the earlier case, she will be at work. Now consider her ability and motivation for socializing with her coworkers at these two points in time, for instance by having coffee together. Irrespective of how high her motivation for the inevitable small talk may be, it is safe to assume that usually, Jane's ability for spending some time with her colleagues will be higher during a Monday lunch break than on a Saturday afternoon. It is simply much easier to get together with one's coworkers on a working day than during the weekend.

The FBM adds a third factor to the equation: The aforementioned triggers. According to Fogg, a trigger can be anything that a person can notice and that this person relates to the desired behavior. Ideally, the perception then *triggers* the associated action. In this regard, a trigger can be something that a person sees, hears, feels, tastes, or smells, as long as this stimulus is mentally linked to the

[4] Different to Fogg, I am using a small b to denote a specific person's behavior. Fogg capitalizes the letter.
[5] This is not an equation in a mathematical sense.
[6] I am using δ, derived from Gr. διάστημα (diastima), time interval, as the variable symbol for time instead of the established t to avoid confusion with Fogg's symbol for triggers. This is in line with the fact that δ actually denotes a time interval of variable length, as explained in chapter three.

desired behavior. The 'bing' of a microwave is a trigger, as hearing it will usually make a person get the food waiting inside the oven. The classical knot in the handkerchief, on the other hand, will not always work as a trigger. Although intended to be, the handkerchief's owner may have forgotten about the reason for tying the knot at the time when she notices it. The stimulus is perceived, but it is no more associated with an action.

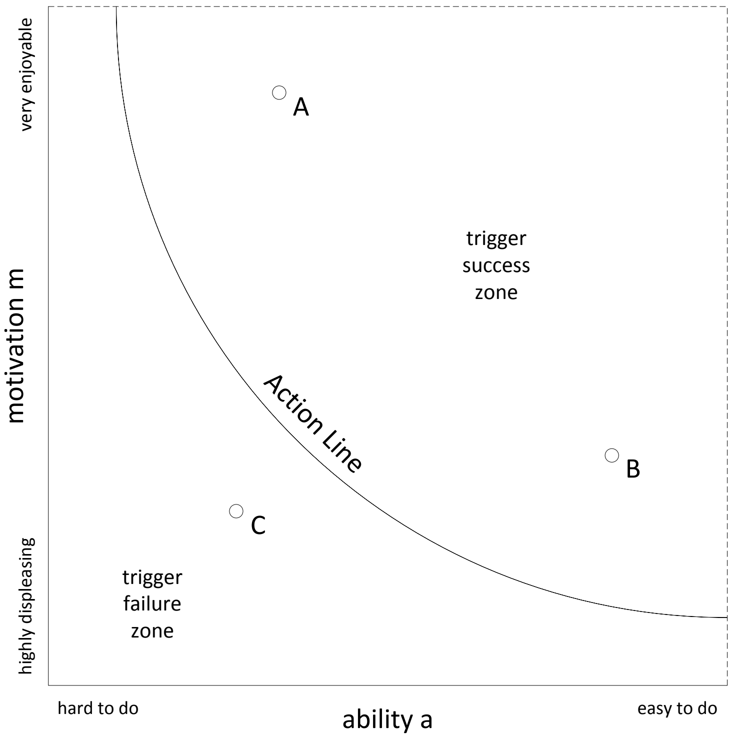


Figure 3: Fogg Behavior Model.

FIGURE NOTES – The Fogg Behavior Model (FBM) shows the interdependence between motivation, ability, and triggers. Adapted from [Fog03].

Fogg further differentiates between hot triggers and cold triggers; the earlier being triggers that occur at a time of high ability for showing the desired behavior, the latter occurring when the perceiving person is currently unable to perform the behavior in question. Fogg summarizes his findings with the words that the key to achieving desired behaviors is to put "*hot triggers in the path of motivated people*" [FHK+10]. Cold triggers on the other hand, triggers occurring when the perceiving person is not able to do as intended, are practically worthless.

Sketched into Figure 3 are the 'Action Line' and the three situations A, B, and C. The Action Line represents the border that separates a sufficiently high combination of motivation and ability from an insufficient amount. Triggers that occur at a point in time when a person's motivation and ability for performing the desired behavior meet above the Action Line will succeed and result in the triggered person behaving in the intended way. Analogously, triggers that occur below the Action Line will fail, and the triggered person will not show the desired behavior. Fogg calls opportune moments in which the combined amount of ability and motivation lies above the activation threshold represented by the Action Line "*kairoi*" [(7)]. The FBM also teaches that ability and motivation can compensate one another to a certain extent. If highly motivated, a person may do something even if she finds the behavior hard or tedious. For instance, offering someone a significant amount of money for running a marathon may result in that person trying to do so, even if she never before ran for more than two miles straight. On the contrary, a high ability may compensate for a low motivation. If something is very easy to do, one may do it despite being not the least bit motivated. Casually voting in an online poll stumbled upon on a website is an example for this effect, as this is something that people do although they neither know the website's owner nor do they really care about the poll's overall result. But because the entire act consists of just a single click with the mouse button – because it is so easy to do – users will occasionally participate in such polls nevertheless.

[7] From Gr. καιρός (kairos), opportunity. Fogg is not the first to apply this ancient Greek term for denoting the 'right time' and indeed, the question of how the term is used correctly has already been a matter of scientific discourse [Mil92].

The points A, B, and C that have been marked in Figure 3 illustrate this interplay of ability and motivation. Starting from the upper left and going clockwise, they represent situations in which the desired behavior is rather difficult for the target person to do, a fact that is compensated by her high motivation (A), a situation in which the target person finds the desired behavior fairly easy to do, although she is not very motivated for behaving in this way (B), and finally a situation in which she is neither particularly able nor motivated to perform the desired behavior (C). An occurring trigger will succeed in the situations A and B, but it will fail in the case of situation C. In this situation, neither ability nor motivation can compensate for one another and a perceived trigger thus cannot initiate the desired behavior, although it may still have a certain effect on the target person. These 'side effects' of triggering pose a significant challenge to all interventive measures. The problem is discussed in more detail in chapters three and four.

We find that there is a clear correspondence between the three types of barriers for physical activity [OB92] and the core elements of the FBM, which reinforces the plausibility of Fogg's model for human behavior. As pointed out in the first chapter, many people are kept from being physically active by a self-perceived lack of ability and/or a lack of motivation. Ability and motivation make up the two axes of the FBM. The third barrier that prevents physical activity, the perceived lack of time, also has a representation in the FBM: Well-timed triggers that reach a person in a situation when both her ability and her motivation are sufficiently high will initiate the desired behavior by grasping an *opportunity* – an opportunity that the triggered person may not even have been aware of herself. Helping a person realize that a desired behavior is both meaningful and possible – intervening at the right time – must be the main competency of interventive measures for behavior change and the number of attempts required before the desired behavior is shown defines the corresponding measure's quality. An accurate measure will await the right situation, a *kairotic* moment, before issuing a (hot) trigger. In contrast, educating people on the benefits of a certain behavior with the intention of increasing its prevalence may not be a successful strategy as such advice will oftentimes turn out to be a cold trigger. The audience, such as a school class, will not be able to act upon the trigger immediately. When they would be able to do so, however, the 'out of sight, out of mind'-effect has taken hold and the previously perceived trigger is long gone and forgotten [8]. In this regard, instead of trying to induce behavior change by spreading long-term knowledge, a more promising approach for overcoming opportunity-related barriers may lie in the creation of measures that are capable of initiating well-timed interventions – hot triggers – that reach people in situations when they can be immediately acted upon. This notion will be picked up later. That aside, it can be subsumed that all three identified types of barriers for physical activity have their representation in the FBM and that the FBM thus appears be a fitting foundation for the conceptualization of technology-based interventive measures that aim to stimulate increased physical activity by helping their users overcome any perceived barriers.

In accordance with the FBM, this work also has three main building blocks: On the one hand, it investigates how a lack of motivation and a lack of ability for physical activity can be counteracted with the help of mobile devices. On the other hand, it analyzes how well-timed triggers can help users overcome their perceived lack of opportunity by pointing out situations in which a few minutes of physical activity are possible. The reminder of this chapter is structured into the same three general categories, but delves into an analysis of how these challenges are already being addressed by science and industry. The described state of the art forms the basis for the development of a new, innovative concept for how to utilize mobile technology to reliably increase the daily amount of people's physical activity. This concept is described in chapters three to six.

[8] It is likely that additional factors besides the 'out of sight, out of mind'-effect prevent people from showing a beneficial behavior. The phenomenon to act against one's better knowledge is called ἀκρασία (akrasia) and is a well-known philosophical problem that has been discussed for millennia [Bau02]. Examples for 'willful wrong doing' are legion and can be found almost anywhere, such as in the domains of personal health [HSS+04], individual mobility [TKV97], and social interaction [DL68], to name just a few. This phenomenon, however, does in no way contradict our understanding of how behavior change works: If a person acts in a certain way although she is aware that another behavior would be more beneficial, then this is simply because either her motivation or her ability for the respective behavior is not sufficiently high.

2.1 User Motivation

The most common conception of the term motivation is that people are focused on gaining pleasure while trying to avoid pain. The notion of this so-called 'hedonic principle' can be traced back at least to Aristoteles and the ancient Greeks [Csi90]. This understanding of motivation can alternatively be formulated as "*people are motivated to approach desired end-states*" [Hig97, p.1282], which implies that motivation is some type of 'energy' that drives people in all their deliberate actions. In other words, motivation consists of an *orientation,* being the direction of actions, and *strength,* being the amount of commitment invested into these actions [RD00]. It should not come as a surprise that such an understanding of motivation spurs some peoples' desires to take advantage of this mechanism: For them, motivation is "*energy to be directed*" and others simply need to be "*fired up*" before being released in the right direction [Hig11, p.18].

Social sciences usually distinguish between multiple kinds of motivation, whereby the most basic distinction is the one between extrinsic and intrinsic motivation [RD00]. Extrinsic motivators are specific outcomes of actions that a person desires and aims to achieve. Such rewards may be material, such as money, or immaterial, such as approval [Fre94]. The required 'energy' for the corresponding activities is then drawn from the perspective of eventually obtaining these rewards. On the contrary, intrinsically motivated persons draw the energy required for the performance of actions from within themselves, because they find the activities interesting or enjoyable. The intrinsically motivated receive no apparent reward [Dec71]. A widespread and often-cited [Meh13, Kon13, Wen15] explanation for intrinsic motivation is the 'Flow Theory' as first described by American-Hungarian psychologist Mihaly Csikszentmihalyi. He points out that people are at their happiest when they "*do feel in control of [their] actions"* and are *"masters of [their] own fate*" [Csi90, p.3]. Being in the flow is thus the state of enjoying an action for its own sake, as it fully involves a person and puts her in a form of trance. Among other things, digital media such as TV or music can induce this flow trance [Nor97] – and digital games.

Sweetser and Wyeth investigated, why digital games are entertaining – intrinsically motivating – to many people and introduced the 'GameFlow Model' [SW05]. Based on the original Flow Theory, this model tries to grasp the core mechanics that make digital games enjoyable, but the authors also point out that their contribution is only meant as a starting point to the development of an understanding on how motivation is being induced by digital games. Based on the findings of Bartle [Bar96], Yee [Yee06], and others, it at least seems safe to assume that not every game mechanic is equally enjoyable to every player. However, what exactly the elements are that motivate people to keep playing a certain game are still difficult to pinpoint. As a consequence, parts of the professional video game industry have resorted to creating games based on fixed 'formulas' that reliably produce results well-received by the majority of their potential customer base instead of experimenting with innovative game designs (9).

Serious games are (mostly digital) games that aim to utilize the intrinsic motivation that well-made games induce by including elements into the gameplay that require the player to behave in a certain way in order to advance in the game [GHW+10]. This required behavior will usually be something "*that would otherwise either be not done or at least with much less enthusiasm*" [Mal14]. In this way, serious games aim to combine the entertaining with the useful and to be "*more than fun*" [SG12]. While the majority of serious games prototypes that come from the scientific community remain accessible to only small groups of users, Japanese game manufacturer Nintendo has already demonstrated twice that serious games can also be commercially successful. The game *Wii Sports* that was released in late 2006 for the game console *Wii* requires users to be physically active in order to play several mini-games such as tennis or golf, whereby the player's movements are tracked by sensor devices that she must hold in her hands while playing [DHK+14]. Ten years later, *Wii Sports* was still the commercially most

[9] The best-known example for this approach is the so-called 'Ubisoft formula' used by French publisher Ubisoft, see https://www.washingtonpost.com/news/comic-riffs/wp/2016/02/26/far-cry-primal-follows-the-tried-and-true-ubisoft-formula-for-success/

successful video game with well over 80 million sold units worldwide [(10)], impressively demonstrating the massive demand for approaches that promise to make physical activity a little more entertaining.

In 2016, Nintendo was able to follow up on its previous success in the serious games niche with the release of the mobile location-based game *Pokemon Go*. To this end, Nintendo partnered with the American Google-spinoff Niantic, a company that had already earned renown in the mobile games market with their previous product, the location-based game *Ingress*. *Pokemon Go* subtly enforces physical activity by requiring the player to reach certain locations or to cover specific distances in order to achieve game-related goals. For example, the 'egg hatching' mechanic of the game demands players to walk or run a minimal distance before they will gain access to a desired game element, while the game utilizes the smartphone's sensors to ensure that the player is not trying to circumvent this requirement by traveling with a car or another vehicle. A mere three weeks after its release, *Pokemon Go* had already reached more than 50 million downloads worldwide and still continued to attract additional users [(11,12)]. It must thus be considered to be the most successful mobile exergame up to date, although certainly not the first of its kind. A good example for an early mobile exergame is the 2012-released game *Zombies, Run!* by UK-based software company 'Six to Start', which combined two game modes: An exercise mode during which the player listens to a series of audio books while running, and an additional strategy-mode, where the player tends to a virtual community [KDH+14]. The connection between the two modes is established by virtual resources that the player automatically gathers during the exercise-mode and that are required for advancing in the strategy-mode.

As both *Pokemon Go* and *Zombies, Run!* can only be meaningfully played when the player is outside and on the move, they can both be considered to be 'pervasive games'. Although the term is heavily debated and there are multiple definitions available [Mon05, HLM+07, Nie07], it should be agreeable to state that pervasive games are digital games that somehow blur the traditional boundaries between a virtual game world and the real world. In other words, pervasive games do not try to block the influences of the real world, but rather embrace them and incorporate them into the gameplay. This sets such games apart from regular digital games played on PCs or TV-based game consoles, where players oftentimes find influences from the real world disturbing and immersion breaking.

The genre of pervasive gaming is usually described with the help of the term 'magic circle', which can be traced back to Dutch philologist Johan Huizinga's monograph 'Homo Ludens': "*All play moves and has its being within a playground marked off beforehand either materially or ideally, deliberately or as a matter of course. Just as there is no formal difference between play and ritual, so the 'consecrated spot' cannot be formally distinguished from the play-ground*" [Hui49, p.10]. Huizinga goes on to list a number of terms to denote this consecrated spot, with 'magic circle' being one of them. Salen and Zimmerman later adopted this term and sharpened its meaning: "*Beginning a game means entering into the magic circle. Players cross over this boundary to adopt the artificial behaviors and rituals of a game. During the game, the magic circle persists until the game concludes. Then the magic circle dissolves and players return to the ordinary world*" [SZ04, p.333].

A good example for this effect is a boxing fight [Cra09]. Within a boxing ring, hitting another person is fully acceptable and even required. However, if the same two people would meet on a public street and started punching one another, this would likely draw a crowd and eventually policemen would stop the combatants. The boxing ring is a magic circle that establishes certain rules which define the game. While the game is ongoing, actions that adhere to the rules are considered to be acceptable. The boxing example also hints at the fact that the magic circle has not only a spatial dimension (the boxing ring), but also a temporal, as the fight only lasts for a predefined amount of time. Finally, the contest also has a social dimension: It is clear, who the fighters are; the rules do not allow one fighter to be replaced by another, let alone a third person to join the ongoing fight. Montola points out, where

[10] http://www.vgchartz.com/gamedb/
[11] https://sensortower.com/blog/pokemon-go-50-million-downloads
[12] Steinmetz et al. point out that there are technical challenges that need to be solved before multiplayer location-based gaming can evolve any further and before more advanced concepts for such games will become practical [SHK+15].

exactly the differences between regular games and pervasive games lie: "*The regular game is played in certain spaces at certain times by certain players [whereby a] pervasive game is a game that has one or more salient features that expand the contractual magic circle of play socially, spatially or temporally*" [Mon05]. As stated before, pervasive games somehow blur the boundaries between the game world and the actual real world, for example by extending the playground beyond a limited area or by allowing people to join and leave the game while it is still ongoing. The EU-funded project IPerG investigated the potentials of this type of game even before the era of mobile devices really kicked off with the release of the first *iPhone* in 2007, and the project consortium created game concepts such as *Love City* and *Rider Spoke* that explored the possibilities of how to best entangle the real world with a (digital) game world [Opp09]. Many more design ideas for pervasive games have emerged since [HLP+12, JS13, KG15].

In the context of this work, pervasive games are an important concept, as they may strengthen intrinsic motivators for the voluntary integration of physical activities into people's everyday lives. As pointed out in the first chapter, knowledge for the benefits of physical activity alone is oftentimes not sufficient. To many people, the external motivator of increasing the odds for staying healthy apparently does not provide enough 'energy' for regular physical activity of sufficient intensity. Trying to create intrinsic motivation instead by making the activity itself more enjoyable may be a more promising approach – and one that has already been shown to be successful by Nintendo's *Wii Sports*, *Pokemon Go*, and other exergames [(13)]. The aforementioned problem of how to reliably create 'good' games, remains, however. As a consequence, Malaka suggests enhancing commercially successful games with exergaming mechanics, practically 'piggybacking' them on tested and proven platforms [Mal14].

In regard to using game mechanics for the purpose of motivating a desired behavior, an alternative to serious games is the gamification approach. The most established definition for gamification is the one coming from Deterding et al., who define gamification as "*the use of game design elements in non-game contexts*" [DDK+11]. There are plenty of examples for gamification in commercial applications, such as *NikeFuel*, a measure used by the company Nike to quantify the amount of physical activity performed by members of their online community *Nike+* [BL13]. *NikeFuel* is a typical case of gamification, as it follows the widespread 'points, badges, leaderboards'-formula that gamification is usually being associated with [HKS14]. This formula, and gamification in general, has gained some infamy for focusing on the creation of (immaterial) extrinsic motivators, although studies show that their installment may actually have a negative effect on the preferable intrinsic motivation [DKR99]. Nevertheless, a systematic survey on gamification by Seaborn and Fels showed a "*positive-leaning but mixed picture of the effectiveness of gamification*" [SF15, p.28] with the actual performance depending on the application domain. It appears that gamification can indeed succeed in creating 'motivational energy' – if it is applied in the right context and not in a way that it feels artificially enforced.

This finding correlates with the success of the quantified-self movement [CLL+14], which is closely related to the principles of gamification. The core idea of quantified-self is that the self-tracking of one's behavior helps with the regulation and ideally the improvement of the same. A widespread example for this idea are smartphone-based activity tracking apps such as *Runtastic*, *RunKeeper*, or *Edomondo* that monitor one's performance during aerobic activities, mainly running and cycling. Such apps have become increasingly popular among amateur athletes in recent years [FM15]. A similar function is provided by wearable activity trackers such as *Jawbone Up*, *Fitbit*, or the *Nike Fuelband* [FRO+15]. These devices, usually worn on the wrist, assess their wearer's overall physical activity with the help of their built-in inertial sensors and quantify this effort, either directly or indirectly via a smartphone application or a website. A very recent addition to the family of wearables is the *Apple Watch* and its *Activity App* [DGH+16]. This application has already been mentioned in the first chapter, and we will come back to it on several occasions.

[13] If only for a limited time. While a number of contributions show the positive short-term effects that exergames can have, e.g. [GHW+10], it still remains to be demonstrated that a long-term adherence to such measures can also be achieved [LSH+13].

2.2 Ability Adaptation

In his 2009-published paper '*A behavior model for persuasive design*', B.J. Fogg speaks about his observation that a person's inclination for showing a specific behavior is closely related to how simple the person finds the behavior. He goes on to list six aspects of simplicity and points out that "*simplicity is a function of a person's scarcest resource at the moment a behavior is triggered*" [Fog09]. In other words, the question of how high a person's ability for a behavior is at a given moment, how simple she finds it, is determined by the one simplicity factor that she lacks the most at this point in time. According to Fogg, the aspects of simplicity are time, money, physical effort, brain cycles, social deviance, and non-routine. The role of time, money, and physical effort in the determination of one's overall ability should be clear. If a person is lacking the time or money that is required for a desired behavior, then this behavior becomes more difficult and thus less likely to be shown. Likewise, as many people try to avoid behavior that is physically demanding, a behavior that requires significant physical effort will usually be considered more difficult or harder to do than one that does not. The factor 'brain cycles' refers to the necessity that a person has to reflect on a subject and possibly must find solutions for more or less difficult problems. In this regard, solving crossword puzzles or playing chess will usually be considered more demanding than watching TV. The final two elements of simplicity according to Fogg are 'social deviance' and 'non-routine'. Both refer to activities contradicting a 'norm', whereby social deviance means that a person is required to act against a social norm and non-routine means that a person acts against her own norm in the sense of having to change the procedures that she has grown accustomed to. Commuting to work with a bicycle instead of using the car for the first time in many years would be a case of a behavior that is 'non-routine'; wearing a full-body chicken suit while doing so might be a rather extreme example for a behavior with a high level of 'social deviance'.

We find that in modern-day societies, 'time' is more often than not the critical element that decides what is considered to be easy or difficult to do. When people speak of the need for "*making things easier*", possibly with the help of technology, they regularly want to see a certain process to be accelerated. This consideration is in accordance with the main reason for physical inactivity, as discussed in the first chapter: More than a third of all interviewees claim that they simply lack the *time* for sports [OB92]. Some other examples for the time as the behavior determining factor include the choice between private car use and public transport, where saving time is considered more relevant than saving money [BC07], the choice for mobile banking instead of traditional banking irrespective of privacy, security, and other concerns [KRW07], or many people's preference for ready-prepared food instead of investing time in the preparation of healthy meals [JD06]. However, although time will oftentimes be the scarcest resource, it is not necessarily so. As Fogg points out, "*each person has a different simplicity profile*" [Fog09]. For instance, a person with a low income will probably value money higher than time and may eventually even sell her car to instead fully rely on public transport. Likewise, a chronically sick person may be willing to invest many hours per week into the preparation of healthy meals. This is a case where a high motivation compensates for a comparably low ability, as although another, more simple-to-do behavior – eating convenience food – would also be possible, the high motivation of wanting to stay as healthy as possible entails the more time-consuming healthy eating. It is not always the easiest alternative that wins the race.

Since computer technology is generally either supposed to support or entertain its users (or both), the examples for devices and programs that are meant to simplify behavior seem to be legion at first sight. The obvious cases include devices and programs to support the writing of texts, to simplify complex calculations, to accelerate long distance communication, and to simplify researching and the distribution of knowledge. But all these domains have one thing in common: Computers more or less replaced all other ways of doing such things. Of all the documents written every day, the vast majority is being written using computers; analogously, almost every communication between two people who are not in shouting distance is done via a computing device. More interesting are those cases where technology is used to simplify behavior that could also be performed without this aid. One such case is the use of navigation systems that simplify planning and orientation while traveling. Even specialized variants of such navigation aids exist that are meant to suit the specific needs of high demanding target groups, such as public transport navigation assistants for the sight or hearing impaired [DMM+14]. Other examples for ability-enhancing computer technologies include Augmented Reality-based guidance tools in machinery maintenance tasks [WBE+13], e-cash systems for making seamless payments in public transportation and supermarkets [HZB+13], and home automation systems that enable their owners to efficiently control the electronic devices in their households [CP10].

The last example is of particular importance, as it hints at the evolutionary goal of computer technology in general, at least according to American computer scientist Mark Weiser and his immensely popular visionary paper '*The computer for the 21st century*' [Wei91]: Eventually, computers will have pervaded every aspect of our daily lives, being embedded into floors, walls, and appliances. They will surround us invisibly and silently and spring to life whenever we ask for – or require – their assistance. But although the appropriateness of Weiser's prediction can already be witnessed, we have not quite achieved truly ubiquitous assistive systems. This is in part because the retrieval of information about the state of the user's environment, about her current activity, and especially about her physical and emotional wellbeing turned out to be a bigger problem than originally expected. Such knowledge is usually being referred to as being contextual information [Dey01] and countless contributions investigate ways of how to gain it with technological means, ranging from physical activity detection with inertial sensors [ABM+10], over situation recognition based on the emittance of smartphone ringtones and notification sounds [DRC+14], to the precise indoor localization of persons, e.g., with the help of regular webcams [BD16]. The immense quantity of such papers that is available on different aspects of the problem of 'context awareness' shows its size and prevalence. Contextual information retrieval has become core to many aspects of modern-day technology.

So far, we have only considered the *simplification* of behavior, which is in line with Fogg's notion of how the occurrence of a desired behavior can be provoked: By reducing the levels of all six aforementioned elements of simplicity, the target person's ability for the desired behavior rises and behaving in the desired way becomes easier for her, which in turn makes it more likely that she actually does behave in this way. However, a person's ability for a desired behavior may also increase (or decrease) by itself, without any external factors artificially simplifying the behavior. If a person regularly uses her bicycle to get to work, this behavior will eventually become normal to her, slowly reducing the degree of this behavior's 'non-routine'. Furthermore, as her body gets used to the strain, the degree of 'physical effort' that is required for commuting with a bike decreases as well, which may in turn reduce the 'time' that she has to invest. Eventually, as she *adapts* to the behavior, her ability for it increases and it gets easier. In some cases, this is a welcomed effect. In others, not so much.

The previously introduced 'Flow Theory' clarifies that enjoyment lies in between boredom and worriedness. The model's creator Csikszentmihalyi elaborates on this and writes that "*activities that reliably produce flow experiences are similar in that they provide opportunities for action which a person can act upon without being bored or worried*" [Csi75, p.49]. Csikszentmihalyi explains this principle with the help of a sketch. An adapted version of the original diagram can be seen in Figure 4. Given three rock climbers A, B, and C, who are of different skill and experience, but who are all trying to take on the same climbing rock. Climber A has the lowest skill of the lot and her experience lies below the climbing route's grade. It is thus likely that rather than looking forward to the climb, she will be worried and possibly even stressed (under the precondition that she is able to correctly assess her own skill level). In contrast, climber C who has the highest skill of the group, a skill level that lies well above the climbing route, will feel bored and annoyed by the necessity of climbing a route below her standard because the other two cannot manage a more challenging ascent. Only climber B, whose skill level perfectly matches the climb ahead, will really enjoy the tour. She is within the 'Flow Channel', the zone where ability and challenge are well-balanced. Csikszentmihalyi points out that "*people in a state of* *worry can return to flow through an almost infinite combination of two basic vector processes: decreasing challenges or increasing skills*" and continues to state that "*conversely, if one is bored one can return to flow either by finding a means to increase environmental challenges or by handicapping oneself and reducing the level of skills*" [Csi75, p.52f]. The 'Flow Theory' thus teaches that it can indeed be desirable to *raise* the physical and/or mental challenge that is imposed by an activity in order to ensure that it remains enjoyable, which in turn means that *adapting* to one's ability does not necessarily imply *simplification*.

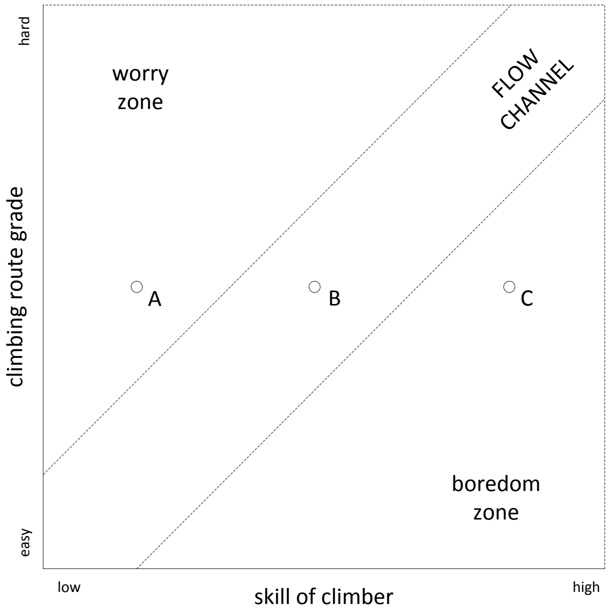


Figure 4: Flow Channel.

FIGURE NOTES – The Flow Channel maintains the balance between worry and boredom, figure adapted from [Csi75].

A domain where this principle immediately becomes obvious is sports and games. When confronted with a task or an opponent, most people prefer a situation where their skill and the challenge that they face are well-balanced in a way that a triumph will convey a sense of actual accomplishment. This holds true for both individual training and competitive sports. The 'Heart Rate Reserve Method', for instance, is an established way of ensuring that exercise intensity matches an athlete's capabilities by modulating the physical strain so that her heart rate is kept within a tight interval in between her resting and her maximum heart rate [PAR+14, p.170]. It is not a coincidence that this approach bears a strong resemblance to the 'Flow Channel'.

The importance of keeping a 'balance' is also encountered in digital games. For instance, Microsoft relies on a method named *TrueSkill* to balance the composition of matches in multiplayer games played on the servers of the *Xbox* game console family. The method is based on Bayesian inference and supposed to deliver better balancing results than the 'Elo rating system' that is used to find balanced opponents for chess matches [HMG06]. Reuter proposes the use of Petri nets to ensure that digital cooperative multiplayer games are well-balanced and pose the same level of challenge to all players, stating that unbalanced games have the potential "*to leave everyone involved dissatisfied*" [Reu16, p.3]. In a similar manner, Wendel proposes the use of an intelligent agent named *GameAdapt* to recognize player behavior and game states in collaborative multiplayer games. When deemed necessary, the agent automatically intervenes by increasing or decreasing the challenge for individual players in order to ensure that game progress remains balanced among all members of a team [Wen15].

Steinmetz and Nahrstedt state that "*adaptive systems monitor the user's activity pattern and automatically adjust the interface or content provided by the system to accommodate (…) changes in user skills, knowledge and preferences.*" [SN04, p.182] Such adaptions to a user's ability are a delicate issue, however, as they harbor the danger of negatively affecting her motivation for the desired behavior. On the one hand, an activity that is (too) difficult for a person to perform will only occur if a sufficiently high motivation compensates for this. On the other hand, however, the 'Flow Model' teaches that finding something *too* easy may have a negative impact on motivation; in this case, simplicity effectively *lowers* the probability of occurrence. The developers of adaptive systems that adjust to the user's behavior must take this interdependence into account.

2.3 Opportunistic Interactions

The studies on the sedentary lifestyle of Australian adults revealed that the main reason for physical inactivity is a lack of opportunity [OB92]. Although it cannot be excluded that a response bias has had an influence on the results, the high amount of participants who gave such answers makes it safe to assume that it is indeed a self-perceived chronic shortage of time that prevents many from being more active. This assumption is supported by several other studies, for instance a more recent one by Welch et al., who found that an alarming 73% of the Australian women perceive time pressure as a barrier to physical activity [WMH+09].

Modern-day societies have high demands on their citizens. Technological and cultural changes such as the Internet and globalization have accelerated the everyday lives of people. The German philosopher Hartmut Rosa uses the term 'social acceleration' to describe the phenomenon that "*the history of modernity seems to be characterized by a wide-ranging speed-up of all kinds of technological, economic, social, and cultural processes and by a picking up of the general pace of life*" [Ros03]. Following up on Rosa's thoughts, Wajcman points out what she calls the 'time pressure paradox': Although technology increases man's efficiency in many of life's aspects such as in the production and transport of wares, this does not "*entail an increase in free time, which in turn would slow down the pace of life*" but rather "*time seems to be increasingly scarce*" [Waj14, p.16]. Computer technology plays not a small part in this acceleration, as its advancements are accompanied by several double-edged developments such as permanent connectivity and information overload. But, as the problem of time scarcity becomes more and more prevalent, a growing number of technological innovations also try to assist the user in coping with this by ensuring that she does things when she wants to – or should.

The majority of these innovations share a common philosophy: That there is a *right time* for interacting with the user. The least sophisticated examples of such timed interactions simply occur at fixed intervals, relying on the flow of time itself as the criterion for when to initiate an interaction. The smartphone application *Plant Nanny*, for example, is meant to ensure the sufficient hydration of its users and allows them to manually define, how many times per day the app should send a reminder that some drinking is due [KOM-B-0540]. In a very similar manner, the *Activity App* of the *Apple Watch* sends a notification to the user every ten minutes before the hour (14), asking her to stand up and move around (15). While this is supposed to help users fight a sedentary lifestyle, the actual realization of this feature can easily lead to irritation and users share stories of all the inappropriate situations in which the *Apple Watch* tried to make them stand up [DGH+16].

There are two alternatives to timed-interactions: Event-based interactions and signal-based interactions. This categorization is based on the findings of Wheeler and Reis, who had analyzed how researchers can self-report relevant information during scientific studies [WR91]. They state that in contrast to the 'interval-contingent recording' approach that demands the aggregate recording of occurred events at the end of each reporting interval, 'event-contingent recording' is done whenever a certain, predefined event takes place, and 'signal-contingent recording' happens when a certain person gives the corresponding command. We find that interactions between users and devices can be grouped into the same three categories. Either the device times the interaction (interval-based) or it awaits the occurrence of specific events (event-based). Finally, the user can always initiate the interaction herself by signaling this intention to the device, usually by a click, a tap, or the utterance of a voice command (signal-based). We also find that event-based interactions can in turn be grouped into those interactions that await events happening inside the device or application, such as the execution of a specific line of code, and those interactions that wait for external events. In the 1990s, an increasing

[14] It must be pointed out that the device does not solely rely on the time to initiate an activity trigger. Rather, the *Activity App* tries to estimate the user's amount and intensity of physical activity during the last hour, based on the device's built-in accelerometer. Only if this estimation value does not exceed a certain threshold, the notification is triggered.
[15] The message says "*Time to stand! Stand up and move a little for one minute*."

interest in finding ways of how to blur the strict borders between computers and the real world and to simplify the flow of information between the two led to a new field of research: Context awareness.

According to Dey [Dey01], the first documented use of the term 'context aware' can be traced back to Schilit and Theimer and their 1994-published paper "*Disseminating active map information to mobile hosts*". They state that "*context-aware computing is the ability of a mobile user's application to discover and react to changes in the environment they are situated in*" [ST94, p.23]. Later, Schilit et al. investigated examples of context-aware computing and found that the three most important aspects of context are "*where you are, who you are with, and what resources are nearby*", but stressing that "*context includes more than just the user's location*" and may also encompass "*lighting, noise level, network connectivity, (…) and even the social situation*" [SAW95]. They also gave two examples of what they called 'context-triggered action applications'. The first example is a UNIX application named *Watchdog* that executes predefined commands when infrared identification tags worn by users [WHF+92] are registered at specific locations. The authors give the example of playing a rooster sound when somebody is in the vicinity of the coffee machine. The second application named *Contextual Reminders* makes use of the same type of identification tags and allows to set reminders that are not only based on time like regular alarms, but additionally also based on the location of oneself and of other (tag wearing) users. As such, *Contextual Reminders* takes two types of information into account when deciding, whether or not to interact with the user, with one of these two being information coming from an external source and being specific to the individual user (namely her current location). *Watchdog* and *Contextual Reminders* may indeed be the first cases of context aware applications.

Two decades lie between these first occurrences of context aware computing and today, and many applications of the event-based interaction paradigm have been created since. IBM's *Intelligent Notification System* of 2001 was meant as a tool against the growing problem of information overflow and supposed to monitor information streams coming from Web services. In case of the occurrence of certain events, such as a stock's price rising above a defined limit, the *Intelligent Notification System* would notify its users [BCE+01]. The system is interesting because of its two-step decision process. The first decision is whether to interact with the user at all. This decision is based on the analysis of the retrieved Web service content, which technically translates to the execution of SQL queries. If certain conditions are found to be met, the system then decides on how to notify the user about this. This is the actual context aware application, as the decision is based on the situation that the user is currently in. However, the authors do not go into much detail on how their *Secure Context Service* gathers the information about the user's contextual state and only list the user's location, her calendar entries, and her "*instant messaging online status*" as possible types of context information to be taken into account.

This lack of diversity is typical for early context aware systems and the majority of them rely either solely or to a large part on the user's location to decide, whether to trigger interactions (16). Some examples include the *stick-e notes* by Brown, digital documents that are automatically opened by mobile devices such as PDAs when the user is at a certain location [Bro96], the *GUIDE system*, an electronic tourist guide for the city of Lancaster (17) [CDM+00], and the *comMotion system* of the Massachusetts Institute of Technology, a speech based tool for defining location-specific reminders [MS00]. It must be noted that many of the papers treating early prototypes of context aware systems express the desire to also be able to use other types of context information besides the user's location. Schmidt et al. even published a (heavily cited) paper titled '*There is more to context than location*', in which they point out that the "*the use of location is dominant [in context aware computing]*" and they proceed to suggest various other types of context information that could be used to enhance applications, such as ambient light or user activity [SBG99]. However, at the time of their writing, technical limitations set a tight

[16] The early fixation on the exploration of application concepts that somehow make use of the user's location can to some extent be explained by the deactivation of the so-called 'selective availability' in the year 2000 – a mechanism, that had artificially reduced the accuracy of the U.S. military owned GPS system for civilian users [KDH+14].
[17] A similar system named *GEIST* was developed for the German city of Heidelberg a few years later [GSH+04].

limit to how sophisticated context aware applications could possibly become. This changed a great deal when sensor-packed smartphones with constant access to Web services became widespread, enabling all kinds of applications based on the occurrence of different types of events external to the device.

While the 2005-introduced and *Nokia Series 60*-based *ContextPhone* platform, being one of the first smartphone-based context aware applications, was still only relying on the user's location, her interaction with apps, and various system states as information sources [ROP+05], a survey by Hoseini-Tabatabaei et al. revealed eight years later how rich the use of context in mobile applications had become since [HGT13]. Table 4 is an incomplete list of the different types of indicators (18) that were reported to be used in context aware applications. The list was assembled by the author of this thesis from twenty-one randomly selected scientific papers published between the years of 1995 (19) and 2015. And although the user's location is still one of the most prevalent types of information used to adapt mobile apps to the user, the consideration of many other indicators such as ambient noise, the user's current activity, and even of the local weather conditions has become a frequent practice.

Especially the detection of user activity has become pivotal for context aware applications. For example, Okoshi et al. rely solely on activity detection (using GPS, inertial sensors, and direct user interactions) to identify situations in which the user is about to change her activity. These so-called 'breakpoints' are then used to deliver all notifications that occurred since the last breakpoint but that were temporarily held back by the smartphone. In an extensive four-week study with 41 participants, Okoshi et al. were able to show that their system indeed had the envisioned effect: It significantly reduced the workload perception of the study participants [ORN+15]. In direct contrast, Mehrotra et al. suggest the use of a much larger set of indicators to predict opportune moments for delivering notifications, among them user activity, user location, ambient sound, ringer mode, and the notification's title [MMH+15]. The inclusion of parts of the notification's content was a novel aspect and an extension to a system that the authors had presented in an earlier paper [PM14], which had touched on another innovative subject: Assessment of the user's emotional state. The inferring of the user's mood and stress level may be considered the supreme discipline of context awareness research, as these factors are among the core determinants of human behavior. However, at the time of this writing, a reliable automatic inference of the emotional state with mobile devices appears to be still years – if not decades – away. Although first considerations were made [LLL+11, LCL+12, LLL+13], the majority of contemporary scientific papers describe systems that rely on a very practical approach when it comes to the assessment of the user's mood: They simply ask.

An important remark that must to be made at this point is that context aware applications are not only encountered on smartphone. For example, Hardy et al. use a desktop PC for gathering information about the user's activity, her schedule, and the local weather and then suggest the one exergame from a list of available (indoor and outdoor) exergames that appears to be the best fit for the current situation [HSG+11]. And as an example from an entirely different domain, consider advanced driver assistance systems that are part of many modern cars. Among other things, they can detect driver drowsiness by monitoring vehicle parameters and driver behavior such as the car's lane position and the frequency of the driver's eye blinking. If such a system detects significant deviations from expected averages, it warns the driver that a break is due [SSM12] and thus tries to influence the driver in a way that she changes her behavior to a more preferable one (such as having a coffee at the nearest gas station). Such *interventive measures*, applications and devices that try to change the user's behavior, are a special type of context aware computing and they are central to this work.

[18] For referring to a specific and well-defined piece of information that a context-aware application is based on, such as the user's location or the ambient temperature, I propose the term 'indicator value', based on [KOM-M-0543]. Indicator values can be both discrete (such as the amount of steps that the user has taken during the day) or continuous (such as the ambient temperature), although the values of indicators of the latter type have to be discretized in some way before they can be used. Indicators are the middlemen between raw sensor data and identified situations. There are a number of alternative terms for this type of information, such as "*cues*" as used by Schmidt et al. [SBG99], "*information category*" as used by Santos et al. [STC+09], or simply "*context*" as used by Hoseini-Tabatabaei et al. [HGT13], among others. For more on indicators, please see chapter four.
[19] The year of the first occurrence of context-aware applications [Dey01].

Table 4: Indicators in Context Aware Systems.

Indicator		Used by
01	User location	[SAW95], [Bro96], [AAH+97], [Dey98], [MS00], [BCE+01], [HHS+02], [MHA04], [ROP+05], [BPT06], [BC08], [CNC10], [BLP+11], [MLC+13], [PM14]
02	Date/time of day	[SAW95], [Bro96], [AAH+97]*, [MS00], [STC+09], [CNC10], [BLP+11], [WRW12], [MLC+13], [PM14]
03	Location of other people	[SAW95], [Bro96]*, [AAH+97]*, [Dey98], [MS00]*, [HHS+02], [ROP+05]*, [BLP+11]†, [PM14]†
04	Program status (e.g., of messenger app)	[BCE+01], [HHS+02], [MHA04], [ROP+05], [MLE+07], [CNC10], [MLC+13], [ORN+15]
05	User posture/activity (e.g., sitting vs. walking)	[SBG99], [SSF+03], [BPT06], [MLE+07], [PM14], [BLP+11]†, [WRW12], [ORN+15]
06	Ambient noise	[SAW95]*, [SSF+03], [MLE+07], [STC+09], [WRW12]
07	Ambient light	[SAW95]*, [SBG99], [SSF+03], [MLE+07], [STC+09]
08	Ambient temperature	[Bro96]*, [SBG99], [MLE+07], [STC+09], [BLP+11]
09	Emotional state of user	[AAH+97]*, [SBG99]*, [MLE+07]*, [BLP+11]†, [PM14]†
10	Indoor/outdoor distinction	[AAH+97]*, [SBG99], [MS00], [MLE+07]
11	Location of objects	[SAW95], [Bro96]*, [Dey98], [HHS+02]
12	News feed content (e.g., stock prices)	[Bro96], [BCE+01], [MLC+13]
13	Conversation detection	[SSF+03], [BPT06], [MLE+07]
14	Local weather conditions (e.g., dry vs. raining)	[SBG99], [BLP+11], [MLC+13]
15	User speed	[BC08], [STC+09]
16	User schedule (i.e., calendar entries)	[BCE+01], [SSF+03]
17	Physiological data of user (i.e., heart rate)	[SBG99]*, [BC08]
18	System state (e.g., battery level)	[Bro96]*, [ROP+05]

An incomplete list of indicators used by (mainly) mobile context aware applications, based on a sample of twenty-one scientific publications from the years 1995 to 2015, items ordered by relevance & bibliography indices ordered chronologically

* use suggested without actual implementation of a detection mechanism
† not automatically detected, but requires manual input from the user

3. General Considerations

The second chapter revealed the existing concepts and approaches for increasing a person's motivation for doing something and for adapting a system to her individual skills or her behavior. We have also looked into sophisticated context aware mechanisms that rely on their knowledge of the current situation to time their interactions with the user; some of these intelligent notifiers are meant for creating awareness for the feasibility or the necessity of a desired behavior. The FBM teaches that all three aspects – motivation, ability, and short-term awareness – must be paid attention to if one aims to create reliable persuasive technologies. This chapter thus investigates, how all elements of the FBM can be unified within a single concept for technology-based approaches that aim to increase the user's amount of daily physical activity – although the considerations made here also hint at the fact that certain qualities of such *interventive measures* may indeed be more important than others.

3.1 Behavior Interventions

The desire to change someone's behavior – either one's own or somebody else's – is the intention of altering the original behavior of this person at a specific point in time in a way that rather than maintaining her original behavior, the person instead shows a different behavior. There are two types of this wish. The first type is focused on a person's original behavior and wants this behavior to end. In this case, the new behavior is more or less irrelevant, as long as it replaces the original behavior. In most cases, terminating an undesired behavior will be possible, at least in principle. The second type of desire for behavior change is focused on a target behavior instead and aims to replace the original behavior with this specific other behavior. Here, the original behavior is usually only as much of a concern as it influences the probability of the target person switching to the desired behavior. This work focuses on the latter case; that is, on the problem of how to replace an arbitrary original behavior with a specific desired behavior.

Just as there are two types of behavior change, there are also two types of interventive measures that intend (or are intended) to change a person's behavior. The first type is the group of *stubborn measures*. As the name implies, such measures do not differentiate between appropriate and inappropriate situations for trying to initiate a specific behavior, but will rather always occur at a fixed time or place. Our everyday lives are full of examples of such stubborn behavior interventions that were either installed by ourselves or by somebody else with the intention of making us behave in a certain way; we usually refer to such measures as 'alarms', 'signals', or 'reminders'. The alarm clock that wakes us in the morning, the red street light that makes us stop at the crossroads, and the verbal "*please mind your step*" reminder at the end of the airport walkway are all stubborn behavior interventions. The alarm clock does not realize that its owner woke up early and is already having breakfast when it goes off. The red light is not aware of trying to stop a racing ambulance on an empty street. And the voice message is undeterred by the fact that most passengers are already minding their step. For the most part, we will simply ignore such stubborn interventions if we consider them to be inappropriate, but we may also try to get rid of them when given the opportunity once their frequent ill-timed intervention attempts have amounted to annoyance. And usually, stubborn measures will not see this coming.

This is where the difference to *discerning* interventive measures lies. Prior to trying to initiate a desired behavior, this type of interventive measure will ponder its chances for succeeding – and it may decide to withhold an intervention attempt if it is given reason to believe that the intervention may not have the desired effect. The discerning measure does this, because it fears the consequences of ill-timed intervention attempts; a good example is a young man patiently awaiting the perfect opportunity for asking his crush for a first date. But not only people are capable of discerning interventions. The driver assistance system mentioned in the second chapter that constantly monitors the driver's behavior in order to detect deviations and that recommends a coffee break if it deems this necessary is an example for a technology-based discerning measure. Such measures try to *await* an appropriate situation before making an intervention attempt. This thesis mainly focuses on this type of behavior intervention measures.

> *Let U be the set of all persons* [20]*, let B be the set of all types of human behavior (being all observable actions and emotions), and let Δ be the ordered set of all points in time. Let $u \in U$ be an arbitrary but fixed person (the 'target person') and let $\delta \in \Delta$ be an arbitrary but fixed point in time ('observed time' or 'δ-period'). The 'behavior reveal function' $\varphi: U \times \Delta \rightarrow B$* [21] *specifies the original, uninfluenced behavior $b \in \mathrm{B}$ of the target person at the observed time:*
>
> $$\varphi(u, \delta) = b$$
>
> *Let $\bar{b} \in B$ be an arbitrary but fixed behavior (the 'desired behavior'). The 'behavior transformation function' $\mu: U \times B \times \Delta \rightarrow B$* [22] *specifies the behavior of the target person after an interventive measure has triggered the person for the desired behavior. We will speak of a 'successful behavior intervention' by the interventive measure, if the following two conditions apply:*
>
> $$(I)\ \varphi(u, \delta) = b \neq \bar{b}$$
>
> $$(II)\ \mu(u, \bar{b}, \delta) = \bar{b}$$

The behavior reveal function φ specifies a person's original and uninfluenced behavior b at a certain point in time δ. If the interventive measure intervenes at this point and issues a trigger with the intention of changing the behavior of this person in a specific way (the required preconditions for successful triggering – that the respective person must perceive the trigger and relate it to the intended target behavior – will henceforth be implied), the person may or may not change her behavior to the desired behavior $\bar{b}$. The behavior transformation function μ specifies the behavior that results from the intervention attempt. If this resulting behavior is the desired behavior $\bar{b}$, then we will speak of a 'successful behavior intervention'.

The parameters of the behavior transformation function μ imply, that the chances for successfully changing a person's behavior are dependent on three factors: First, the target person u, second, the desired behavior $\bar{b}$, and third, the situation that the target person is currently in, represented by the point in time δ. Note that referring to the time δ is sufficient for characterizing the target person's *entire* contextual situation, including her current behavior. This is because the interventive measure – person or device – functions as a mere observer to the behavior of the target person until it influences her by issuing a trigger. If the target person is not triggered and her behavior not influenced, she simply proceeds as she sees fit. There is a resemblance here to watching a play: The progress time determines the setting that the actors are in and how they behave on stage. If uninterrupted by the audience, the actors will proceed with their performance, and their behavior on the stage is decided by the time period that has elapsed since the start of the play. To a passive observer, the current time thus entirely determines the contextual situation of the actors and their behavior, and likewise, from the perspective of an interventive measure, the current point in time determines the situation that the observed person is in and how she is behaving in this situation. This effect becomes especially obvious in retrospect: When looking back at the things that were, the specification of a point in time determines all other aspects of our lives, such as where we were at, who we were with, and how we did behave.

[20] I am using U for *users* instead of P for *persons* to set the stage for a later application of these theories in a technical context.
[21] From Gr. φανερόω (phanero), to reveal.
[22] From Gr. μεταστρέφω (metastrepho), to transform.

Even if an intervention is not successful and the target person does not change her behavior in the desired way (or not at all), it must be assumed that the intervention attempt still somehow influences the target person. Most importantly, the intervention attempt may increase or decrease the target person's motivation and/or her ability to comply to similar intervention attempts in the future. However, this effect will usually be hidden from the observer and only become apparent at a later point in time. We must thus assume that triggering always changes the success probability of similar subsequent triggers occurring within a certain time period and that triggering attempts are never 'for free'. Obviously, these side-effects of triggering are a potential problem. In contrast, the mere consideration of a triggering attempt that is ultimately withheld will not have an effect on the target person, as long as she remains unaware of it. A noteworthy special case of interventions occurs when the target person already shows the desired behavior. In this case, a successful behavior intervention is not possible, as the current behavior cannot be *changed* to the desired behavior. Intervention attempts undertaken nevertheless may instead have significant negative effects. We will get back to this later.

The success of interventions does not only depend on the current situation and the target behavior, but also on the person whose behavior is supposed to be changed. Consider a manager who intends to ask her employees who are just preparing to leave to instead stay late and to finish the task that they have been working on, such as preparing presentation slides for a next day's meeting. Chances are that one of them will agree, while another one will refuse and leave. Although the current situation and the desired behavior are the same for both these persons, there are more subtle differences that also have an effect on the outcome of the intervention attempt. The FBM reveals what these differences are: An intervention will only be successful if both a target person's ability and motivation for the desired behavior are sufficiently high at the time of the intervention attempt.

The 'ability function' $a: U \times B \times \Delta \rightarrow [0,1]$ *determines the ability of an individual u to show the desired behavior* $\bar{b} \in B$ *at the point in time* $\delta \in \Delta$. *Likewise, the 'motivation function'* $m: U \times B \times \Delta \rightarrow [0,1]$ *determines the motivation of an individual u to show the desired behavior* $\bar{b} \in B$ *at the point time* $\delta \in \Delta$. *The dependent 'behavior factors product function'* $BFP: U \times B \times \Delta \rightarrow [0,1]$ *specifies the product of these two functions' results:*

$$BFP(u, \bar{b}, \delta) = a(u, \bar{b}, \delta) \times m(u, \bar{b}, \delta)$$

Let $\tilde{b} \in B$ *be a behavior different to the desired behavior, such that* $\tilde{b} \neq \bar{b}$ *.Given a fixed (but unspecific) 'activation threshold'* $\theta \in [0,1]$ *, we can then define the behavior transformation function* $\mu: U \times B \times \Delta \rightarrow B$ *as follows:*

$$\mu(u, \bar{b}, \delta) = \begin{cases} \bar{b} & if\ BFP(u, \bar{b}, \delta) > \theta; \\ \tilde{b} & else. \end{cases}$$

Maybe the manager in above's example is aware of the fact that one of her employees has a child to look after and is thus not able to stay late. Or maybe the manager already knows from past experience that another employee does not have the motivation for going the extra mile. In both cases, the manager anticipates that the value of the respective employee's *BFP* function for the desired behavior of staying late will not exceed the employee's activation threshold θ, and that she is thus going to refuse the manger's request. The manager would then be well advised to not even ask, as unsuccessful interventions are potentially harmful. Oftentimes, however, she will find herself in situations when she cannot state *a priori*, whether or not a successful behavior intervention is possible, or not. If one cannot safely predict the outcome of an intervention attempt, the only option of finding out is usually to simply give it a try.

Given a target person $u \in U$*, and an observed point in time* $\delta \in \Delta$*. Let* $\bar{b} \in B$ *be the desired behavior. The 'intervention success determination function'* $ISD\colon U \times B \times \Delta \rightarrow \{0,1\}$ *states, whether the product of the target person's motivation and ability for the desired behavior exceeds the activation threshold and thus, whether an intervention attempt will be successful:*

$$ISD(u,\bar{b},\delta) = \begin{cases} 1 & if\ BFP(u,\bar{b},\delta) > \theta; \\ 0 & else. \end{cases}$$

A trigger is essential to the initiation of behavior change. Irrespective of an employee's ability and motivation for staying late, if the manager does not intervene to ask, the employee is likely to pack up and leave. Of course, the employee may also decide herself to stay until the task is finished, without the manager having to explicitly ask for this. In this case, the desired behavior is already achieved and there is no need for a change of behavior. An intervention attempt that is undertaken nevertheless may even be harmful. A manager that urges her employees to stay although they are not indicating that they intend to stop may actually achieve the opposite and find her team leaving within minutes, essentially having reminded them that it is time to call it a day. Analogously, a manager asking an employee to work late although she knows that this employee has a low ability of staying may be considered insensitive; a manager asking an employee with low motivation may be called disrespectful. Ill-timed triggers can do harm, and as such, the act of triggering is a double-edged sword. Ideally, triggers will only be initiated when they have a significant chance of changing the target person's behavior in the desired way. Alas, triggering oftentimes involves guesswork.

Given a target person $u \in U$*, and an observed time* $\delta \in \Delta$*. Let* $\bar{b} \in B$ *be the desired behavior. The 'intervention success confidence function'* $ISC\colon U \times B \times \Delta \rightarrow [0,1]$ *specifies the confidence of an observer – possibly of the interventive measure itself – in an intervention attempt aimed at changing a person's behavior to the desired behavior at the observed time to be successful.*

If $ISC(u,\bar{b},\delta) = 1.0$*, then the observer assumes a 'kairotic situation';*
if $(0.5 < ISC(u,\bar{b},\delta) < 1.0)$*, then the observer assumes a 'parakairotic situation';*
if $ISC(u,\bar{b},\delta) = 0.5$*, then the observer assumes an 'adilotic situation';* (23)
if $(0.0 < ISC(u,\bar{b},\delta) < 0.5)$*, then the observer assumes a 'parachronotic situation';*
and if $ISC(u,\bar{b},\delta) = 0.0$*, then the observer assumes a 'chronotic situation'.* (24)

The observer making assumptions about the success chances of an intervention attempt may be the interventive measure itself, such as the manager. It should be obvious that only *discerning* interventive measures can make such assumptions and indeed, it is this ability combined with a willingness to withhold unpromising interventions that makes the difference between a stubborn and a discerning measure. Note that the 'intervention success confidence function' *ISC* does not state anything about the actual success chance of a behavior intervention attempt, as specified by the (unknown) *ISD* function. *ISC* merely expresses what the observer *believes* this chance to be, which does not necessarily have to be an accurate assumption. A responsible manager who wants to avoid doing damage to the relationship to her employees will estimate the chances for a successful intervention before asking and if she deems herself in a *parakairotic* situation, then she may decide that asking has a point. In doing so, she may possibly be able to convince her employees to stay, but the manager may also discover that her guess is completely off and witness her employees react with outrage. There are plenty of reasons that may lead

23 From Gr. ἄδηλος (adilos), uncertain.

24 I am aware that εὐκαίρως (eukairos) and ἀκαίρως (akairos) would have been better suited as denominators for the two extremes. However, following Fogg's nomenclature, if καίρως (kairos) denotes an opportune moment that must be grasped then χρόνος (chronos) seems to be a fitting antagonist to denote a situation in which time should just be let flown by.

to this kind of misjudgment, but we can identify three main types of problems that can negatively affect a measure's intervention success rate. We will refer to these as the 'three troubles of triggering'.

(1) LACK OF INSIGHT – The observer does not know of a factor that influences the triggered person's ability and/or motivation for the desired behavior at the observed time δ. Due to this, an accurate calculation of the chances for a successful intervention is not possible and any estimation will have to involve guesswork.

(2) LACK OF REASON – Although the observer has access to certain knowledge that is relevant to the outcome of a triggering attempt, this information is essentially ignored.

(3) LACK OF EXPERIENCE – The observer has knowledge of multiple parameters but cannot correctly judge their relevance, especially in regard to the question of how they relate to one another.

Successful behavior interventions manifest as well-timed triggers that reach the target person at a time when the value of her *BFP* function is higher than her activation threshold θ. Ideally, the *ISC* function of a discerning interventive measure will be so accurate that it only differentiates *kairotic situations* from *chronotic situations*, meaning that no uncertainty is involved and that the measure always knows, when to trigger and when not to. In this case, the intervention success confidence function *ISC* perfectly resembles the intervention success determination function *ISD*. Usually, however, the measure will not be able to prognosticate the outcome of a triggering attempt with perfect accuracy. One strategy to solve this problem is to fully investigate the factors that influence the target person's decision, which oftentimes is not possible. A second approach is to try and increase the values of the target person's ability function a, or of her motivation function m, or of both, thus making it easier and/or more desirable for her to show the intended behavior (25). This would have the welcomed side-effect that successful triggering would even become possible in situations where it originally has not been. For some employees, the promise of financial compensation or of a paid time off may go a long way towards increasing motivation. Fostering the ability of an employee, for instance by offering to organize a babysitter, may help as well. Usually, a combined approach that consists of researching the influencing factors on the one hand and of trying to raise the target person's ability and motivation for the desired behavior on the other will be the most promising strategy.

But in any case we find that the trigger is always the linchpin of the behavior intervention. Without a trigger, no behavior change will occur. It thus seems meaningful to focus on the problem of triggering first when trying to initiate behavior change, an insight that is in line with Fogg's conclusion (26). But if both the target person who is supposed to be influenced and the desired behavior which is meant to be initiated are given, solving the problem of successfully changing a target person's behavior translates to solving the problem of identifying appropriate situations that are suited for interventions. In other words: The secret of successful interventive measures lies in their calculation of an accurate *ISC* function that closely resembles the actual *ISD*. A person does this calculation implicitly based on her empathy, her situational awareness, and past experiences. A young man planning to ask his crush for a first date, for example, will usually wait for the perfect moment. This implies that he believes to be able to distinguish between well-suited and ill-suited situations; he may possibly even have an idea of what makes a terrible, an ill-suited, a well-suited, and a perfect opportunity for popping his question. This thesis is focused around the problem how technical systems must be designed so that they can accurately make these kinds of distinctions. Which takes us to the question, what exactly *accuracy* means in this specific context.

[25] There may indeed be another option to increase the probability of successful triggers: To decrease a person's activation threshold Θ, which equals increasing that person's capacity for enthusiasm – a topic well out of scope of this thesis.

[26] "*In our academic and industry work, we've found a specific sequence that works best. And our conclusion may surprise you: When you design for persuasion, you don't start by manipulating motivation. That's what you do last. So what's first? Focus on Triggers first. This is the simplest change, and is often all that is needed*." [Fog10, p.12]

3.2 Accuracy, Effectiveness, Reliability

As pointed out in the previous section, there are two types of interventive measures. The first type intends to prevent a certain behavior from happening, or at least to stop it as soon as it occurs. Such measures can only be successful if they can survey the target person's behavior as exhaustively as possible. Consider a parent trying to prevent her teenage child from smoking. As soon as the child is out of her parents' sight, it may secretly smoke nevertheless, maybe with the conspiratorial support of a cigarette-sharing friend. In this case, the measure fails because it is disrupted and cannot observe all situations that the target person goes through. The second type of interventive measures is those measures that try to enforce a certain behavior. While it may be somewhat less obvious than in the first case, we find that they will also strongly benefit from having constant access to the target person. As an example, consider a loving daughter trying to ensure that her elderly mother who is living independently is drinking enough water. Because she only visits for half an hour in the evenings, there is a limit for how much her mother can drink under her supervision, while many opportunities for drinking during the day are missed as no one is there to enforce the desired behavior (and the old woman herself simply forgets). In addition, the mother may eventually refuse to drink substantial amounts of water in front of her daughter, feeling forced and controlled. Instead, reminding the woman multiple times a day to take small sips would have been the better approach. Opportunities for triggering the desired behavior, situations in which the product of the target person's ability and motivation is sufficiently high, come and go.

Given a target person $u \in U$. Let $\bar{b} \in B$ be the desired behavior. Given furthermore the ordered set of points in time $O \subset \Delta$, a proper subset of Δ which we will refer to as the 'observed period' or as the 'O-period'. Let $i \in \mathbb{N}$ be a natural number and let $\delta_i \in O$ be a point in time within the observed period. The 'total opportunities counter' $TOC: U \times B \times \Delta^{|O|} \to \mathbb{N}$ specifies the total number of opportunities occurring during the observed period at which successfully changing the behavior of the target person to the desired behavior would be possible:

$$TOC(u, \bar{b}, O) = \sum_{i=1}^{|O|} ISD(u, \bar{b}, \delta_i)$$

The *TOC* is a theoretical construct and usually cannot really be calculated. Among other things, its value depends on the exact definition of the 'observed point in time $\delta \in \Delta$'. The implicit understanding of δ contained in the definitions of the behavior transformation function μ and of the intervention success determination function *ISD* is that such a δ-period is long enough to allow the target person to perceive and interpret a trigger. It thus seems meaningful to not go below the minimum of one second for the length of δ. However, a corresponding upper bound for the interval is much harder to find. If we define the length of δ to be, say, one hour, the target person can easily go through a multitude of situations during this time. This is problematic, as different situations may mean different levels of ability and/or motivation for the desired behavior. Within one hour, one can be talking to colleagues in the office, be driving home in a car, be taking a shower, and be watching TV. Behavior interventions aimed at inducing a desired behavior may have very different chances of succeeding in these different settings, depending on the target person's traits and the desired behavior's preconditions. The longer that the time interval δ is, the higher the chances are that a behavior intervention occurring at *some* point during this period will be successful. As such, an overly large interval reduces the expressiveness of parameters such as the *TOC*. We will thus agree on the following convention: The upper bound of the δ-period equals the length of the period that an intervention attempt can be assumed to have a 'reverberant effect' on the target person. A trigger reaching the target person while she is in the office may still lead to the desired behavior half an hour later when she is at home. Specifying how long exactly this 'reverberant effect' lasts requires further investigation, but as we require an upper bound to work with, we will assume that a trigger will usually not change a person's behavior at a point in time later than an hour after its occurrence (although it may influence the person for a much longer period, see below). Hence, an hour shall be the upper limit for δ. For the sake of simplicity, we will also assume that if an intervention attempt is made by an interventive measure, this is always done *right at the beginning* of a δ-period. This is obviously not the case for the majority of interventive measures in general, and especially not for the stubborn ones. Many of them rather happen when the target person reaches a certain location, or when a specific event occurs. Some discerning measures, however, may indeed consider interventions only in fixed time intervals. We will get back to this 'interval-based intervention strategy' – and discuss its alternatives – in chapter four.

When trying to count the total amount of intervention opportunities within a certain time period, we also encounter another problem: The act of triggering itself is likely to affect both the triggered person's ability and her motivation for the desired behavior, regardless of whether the trigger was actually successful. If the daughter in above's example has succeeded in making her mother drink a glass of water, then a second attempt of making her drink another glass just minutes later may not succeed. The success of the first intervention may have lowered the mother's motivation for the desired behavior and thus influences the probability of subsequent successful interventions. Likewise, a successful trigger may also have an effect on the person's ability: If the mother just emptied her last bottle of water, then she will simply not be able to drink a second glass. And even unsuccessful triggers may affect a person, as having been requested to do something before may lower or raise one's willingness to comply with subsequent requests. As stated earlier, we must thus assume that triggering *always* affects a person and that it is never 'for free' in the sense of being without consequences. But here is a dilemma: On the one hand, for a given target person $u \in U$, a given desired behavior $\bar{b} \in B$, and a given point in time $\delta \in O$, we cannot state with absolute certainty *a priori* whether or not at trigger occurring at δ with the intention of changing the target person's behavior to $\bar{b}$ will be successful. On the other hand, actually issuing a trigger to discover its effects is likely to affect the chances of all subsequent triggers during the observed time period and to thus change the total number of triggering opportunities during O. This makes counting the opportunities for successful interventions within a given time period practically impossible, as the attempt of doing so affects this number – the classical observer effect [MDH+08]. We will assume the hypothetical existence of the *TOC* nevertheless, as we require it for another definition.

Given a target person $u \in U$, the desired behavior $\bar{b} \in B$, and the observed period $O \subset \Delta$. Let $i, j \in \mathbb{N}$ be natural numbers and let $\delta_i \in O$ be a single point in time within the observed period. Given furthermore one or more points in time during the observed period at which the interventive measure does NOT have access to the target person: $\tilde{\delta}_j \in \tilde{O}$ with $\tilde{O} \subseteq O$. We can then define the 'missed opportunities quote' $MOQ: U \times B \times \Delta^{|O|} \rightarrow \mathbb{Q}$ which specifies the ratio of triggering opportunities that this interventive measure has missed during the O-period:

$$MOQ(u, \bar{b}, O) = \frac{\sum_{j=1}^{|\tilde{O}|} ISD(u, \bar{b}, \tilde{\delta}_j)}{TOC(u, \bar{b}, O)}$$

Just as the *TOC* that it is based on, the missed opportunities quote *MOQ* is a theoretical construct that cannot actually be calculated. It is, however, the representation of a well-known phenomenon: The "*you should have asked earlier*"-reproach. Vainly requesting a certain behavior from a person may result in that person pointing out that the respective behavior would have been possible at an earlier point in time. Sincere or not, this explanation serves to make clear that the requester has missed an opportunity for a successful intervention. There are three reasons why such an opportune moment may have gone by unused. First, although the interventive measure has been aware of such a moment, she (or it) may have willfully decided against an intervention, for whatsoever reason. Second, the interventive measure may not have recognized the opportunity and thus erroneously decided to wait for a better chance. And finally, the measure may have simply not have had the chance of triggering the desired behavior because she (or it) did not have access to the target person at this specific time. The first two reasons are related to the aforementioned 'three troubles of triggering', that is a lack of insight, a lack of reason, and a lack of experience. The third case, however, is caused by another problem: The disruption of the interventive measure's access to the person who is meant to be triggered. The missed opportunities quote *MOQ* specifies the severity of this problem and the higher that the *MOQ* is for the observed period, the less likely it is for the respective interventive measure to succeed in initiating the desired behavior. A low *MOQ* is thus preferable, which can only be achieved by the interventive measure accompanying the target person throughout the day and throughout all the different settings that she finds herself in. In other words: The measure needs to be *pervasive*.

The 'pervasiveness quote' PQ of an interventive measure specifies the measure's level of access to the target person at the relevant points in time occurring during the observed period O:

$$PQ = 1 - MOQ$$

We will call an interventive measure 'semi-pervasive' during the observed period O, if

$$PQ \geq 0.5$$

Analogously, we will call the measure '(fully-)pervasive' during the observed period O, if

$$PQ = 1$$

According to this definition, a semi-pervasive interventive measure misses at most half of the occurring opportunities for successful interventions and a fully-pervasive interventive measure misses none. Note that this definition of pervasiveness is not focusing on a general availability, but that it rather requires the measure to have access to the target person at the *relevant* points in time. This means that a measure that accompanies a person for large parts of the day but that misses the few minutes during which the person could be successfully triggered for the desired behavior is not even a semi-pervasive measure in the sense of above's definition, as it is not present when it counts. Likewise, a measure that does not have access to the target person for many hours of the day but that is close to her whenever an opportunity for a successful behavior change arises may be fully-pervasive, although it does not have a high presence. A post-it sticker on the refrigerator that states to not forget to buy the milk is not a pervasive interventive measure. While it will be in the vicinity of the message's intended recipient for long hours of the day, it will not be there just when the desired behavior is possible, namely at the supermarket. The target person may still remember the note on the refrigerator when she is at the supermarket, thanks to the intervention's 'reverberant effect', but it is equally likely that she forgets the milk. On the contrary, the tank light of a car is a good example for a pervasive measure that is usually able to remind the driver in time to refuel her vehicle, even if the driver only spends a few minutes of the day in her car.

We find that pervasiveness is not a quality characteristic by itself. Above's definitions of the triggering opportunities counter *TOC* and the missed opportunities quote *MOQ* do not imply actual interventions. They rather only indicate the presence of the corresponding interventive measure during the relevant points in time, which is the basis for being able to make successful interventions. Without this presence, opportunities cannot be identified and grasped. However, an interventive measure with a high pervasiveness quote is not necessarily a *good* interventive measure, and can, in terms of successful interventions, be outperformed by a measure with a very low pervasiveness quote. A high pervasiveness quote simply means that the measure potentially has more opportunities for changing the behavior in the desired way. As an analogy, think of a game of soccer: While one team shots in vain at the opponent's goal over and over again, the other team may shot but once and score. The fact that it has had many more opportunities for scoring a goal may have made it unlikely for the dominant team to lose the game – but it certainly has not rendered it entirely impossible. Likewise, a pervasive interventive measure may have a principle advantage over a non-pervasive one, but that does not necessarily make it the more *reliable* measure.

Given a target person $u \in U$, a desired behavior $\overline{b} \in B$, and an observed period $O \subset \Delta$. Let furthermore be $TOC(u,\overline{b},O) > 0$. We will call an interventive measure that aims to change the target person's behavior to the desired behavior 'reliable during the observed period', if

$$\exists \delta \in O \rightarrow \mu(u,\overline{b},\delta) = \overline{b}$$

Analogously, we will call an interventive measure 'effective during the observed period', if

$$\forall \delta \in O.\,(ISD(u,\overline{b},\delta) = 1) \rightarrow \mu(u,\overline{b},\delta) = \overline{b}$$

Under the condition that a behavior change is possible at some point during the observed period, we find that a *reliable* interventive measure will successfully initiate the desired behavior at least once during this time interval, while an *effective* interventive measure will successfully initiate the desired behavior every single time when the opportunity for such an intervention arises. Naturally, for a given interventive measure the odds for achieving reliability will increase with the length of the observed period while at the same time, increasing this duration also decreases the measure's chances for staying reliable. Pervasiveness is a precondition for effectiveness, as this trait ensures that no opportunities are being missed by the measure. An effective interventive measure is thus always a pervasive measure, but not necessarily the other way around. Reliability, on the contrary, does not require pervasiveness, but the latter will make the achievement of the earlier easier, as the measure will profit from a high amount of opportunities to successfully trigger the intended behavior. Summing up we find that a higher pervasiveness quote *PQ* and a longer observed time period O both increase the odds that an interventive measure will be *reliable* in initiating the desired behavior at least once. Given a sufficient amount of attempts, every interventive measure may eventually be successful. Usually, however, a measure will not be granted a significant number of attempts.

This makes *accuracy* another important characteristic of interventive measures. A *fully-accurate* measure has a perfect success rate of triggering the desired behavior, meaning that whenever a person perceives a trigger initiated by such a measure during the observed period, she will change her behavior in the desired way. In most cases, this will be because the measure is very good in assessing the target person's ability and motivation for the desired behavior and it thus only initiates the behavior when it is certain that the product of ability and motivation – the value of the *ISD* function – exceeds the activation threshold θ [27]. But accuracy is not to be confused with reliability. The latter trait requires a measure to grasp every single opportunity for changing the behavior in the desired way, while accuracy only demands that *when* an intervention attempt is being made, it needs to be successful. Consequently, an accurate measure may let many opportunities go by unused.

[27] Some measures may also be able to successfully initiate the desired behavior with every single intervention attempt because they are capable of raising the target person's ability and/or motivation above the activation threshold. Such measures do not need to await situations that are suited for triggering the target person, as they can rather *create* such situations themselves. As an example, consider a policeman asking a car driver to step out of her vehicle, to close her eyes, and to stand on one leg. The majority of people will not be inclined to show this kind of socially conspicuous behavior voluntarily in public, but the policeman is capable of infusing the required motivation into the target person almost at will.

The 'accuracy quote' AQ of an interventive measure specifies the relative amount of triggering attempts undertaken by this measure during the observed time period O that are successful in initiating the desired target behavior. We will call an interventive measure 'semi-accurate' during an O-period, if

$$AQ \geq 0.5$$

Analogously, we will call the measure '(fully-)accurate' during the observed period, if

$$AQ = 1$$

Accuracy is crucial for many interventive measures, as futile intervention attempts are potentially harmful. Repeatedly urging a person to do something specific in situations when that person is either not sufficiently able or not sufficiently motivated may result in annoyance, which in turn may lead to the respective person ensuring that the measure and the frequent disruptions that it causes are pestering her no longer. An acquaintance frequently asking for a date may be avoided; a mistimed reminder may be turned off. The exact number of improper intervention attempts made by an interventive measure that a person is willing to tolerate may depend on that person's patience and good-naturedness, but it must be assumed that even the most enduring character will eventually try to get rid of a nuisance. Every intervention attempt is thus accompanied by the potential risk of irrevocably annoying the target person. As such, the accuracy quote of any interventive measure should be as high as possible, and – depending on the target person and the desired behavior – it may oftentimes be better to let a potential opportunity for an intervention go by unused than to risk a failed attempt. In this regard, a discerning measure's ability to clearly (and correctly) identify *kairotic situations* may be essentially if the measure is supposed to accompany and influence its target person for a long time.

Up to this point, we have mostly considered people as discerning interventive measures: A manager trying to convince her employees to work overtime, a parent trying to prevent a child from smoking, and a woman trying to ensure the sufficient hydration of her elderly mother. However, discerning behavior interventions can also come from technical devices. A good example is the previously mentioned driver assistance system. Irrespective of whether it is a person or a technology-based system, every discerning interventive measure needs two qualities in order to be effective: It needs to be in the vicinity of the user when opportunities arise, meaning that it needs to be as pervasive as possible, and it needs to be able to distinguish situations that are suited for intervention attempts from those, that are not. We find that, depending on the desired behavior, the pervasiveness quote of smartphones may be higher than that of any other contemporary computational device, as many people keep their smartphones close to them throughout the day: Lying on their desks at the office, tucked away in their pockets and handbags when on the move, and resting on the bedside table during the night [DKH+13]. A smartphone is oftentimes not more than an arm's length away from its owner, regardless of the situation that she is in and as such, it will rarely miss an opportunity for the initiation of a desired behavior [(28)]. In addition, thanks to their internal sensors, smartphones have the means required to monitor various aspects of the user's behavior and of the current state of her environment, which enables them to distinguish different contextual situations. And although the information that can be gathered by smartphones is by far not sufficient for making truly sophisticated decisions as we will discuss in chapter four, they are still more capable of doing so than any other type of mass-market computational device. As such, at the time of this writing, smartphones are the best suited option for the creation of technology-based interventive measures and should thus be the platform of choice for implementations of the concept that is described in chapter four.

[28] The evaluation results presented in chapter seven strongly support the assumption that the majority of users keeps their smartphones close to them throughout the entire day.

3.3 Pervasive Interventions for Physical Activity

Although we have discussed all kinds of behavior interventions on the last pages, there is actually a very specific type of behavior in whose stimulation we are interested in: Physical activity. Since the desired behavior $\bar{b}$ determines the value of the target person's ability function $a(u, \bar{b}, \delta)$, the value of her motivation function $m(u, \bar{b}, \delta)$, and as such the value of the *behavior factors product function BFP*, it has a strong effect on the outcome of any intervention attempt. At a specific point in time δ, the odds for successfully triggering a target person to do one thing may be completely different from the chances of making her do something else. In a meeting room, asking someone to close a window is much more likely to be successful than asking that same person to close her eyes, to stand on one leg, and to sing a song. The success rate of a behavior intervention depends on the person whose behavior is supposed to be changed and on the situation that she is in; but first and foremost, it depends on the desired behavior that the trigger intends to initiate. It can thus not be assumed that an interventive measure that reliably triggers one behavior will have the same success rates in the initiation of another, even if similar, behavior. Usually, highly successful interventive measures will also be highly specialized.

In this regard, the encouragement of physical activity 'in general' may not be a sufficiently specific field of consideration for the design of technology-based interventive measures. Physical activities differ in their intensities, as illustrated by the MET hierarchy [AHH+11], but they also differ in type and number of their preconditions. Many physical activities require the athlete to be at a specific place, to have access to specific equipment and/or infrastructure, and/or to be in the company of a certain amount of other people. Interventions for the last type of activity are especially challenging, as they demand the successful activation of multiple persons at the same time. As a target behavior for technology-based interventive measures, it seems reasonable to rather focus on physical activities that have as few preconditions as possible. In most cases, this should increase the total opportunities counter *TOC* and as such serve to strengthen the chances of successful interventions [(29)]. Table 5 shows a shortlisted selection of medium-intensity and vigorous-intensity physical activities that have only a low number of preconditions and thus seem to be suited for pervasive behavior interventions. One of the activities that has no preconditions at all is 'brisk walking', which can be performed almost anywhere and at any time. With a MET value of 4.3, brisk walking is also well above the lower intensity limit required by the WHO recommendations and so appears to be a good pick for behavior interventions. But the goal of an interventive measure must not only lie in ensuring that an activity is performed sufficiently often, but also sufficiently long.

As described in the introductory chapter, Wen et al. [WWT+11] found that 15 minutes of medium-intensity physical activity per day already have a significant health benefit, resulting in an increase of average life expectancy of 2.55 years for men and 3.10 years for women, while 30 minutes of medium-intensity physical activity per day result in a 4.21 years longer life expectancy for men and a 3.67 years longer life expectancy for women. Longer daily sessions of physical activity and/or an increase of the intensity level will lead to even greater health benefits. On the contrary, less than 15 minutes of activity per day may also entail health benefits, but this has not yet been validated [WWT+11, p.1250]. Because of this, behavior interventions that try to encourage physical activity should aim for the minimum amount of 15 minutes of medium-intensity physical activity per day – any excess of this amount in terms of duration or intensity should be considered an added bonus. Furthermore, studies support the assumption that the effects of multiple short sessions of physical activity are equivalent to

[29] Not necessarily, though. Having few or no preconditions means that the respective physical activity can in principle be performed almost anywhere and at any time. Examples for such activities include various types of own body-weight training (calisthenics) such as squads or push-ups. However, this immediacy only determines a person's *ability* for the behavior. Since the *TOC* counts situations in which a person's product of ability *and* motivation is sufficiently high, a constantly low motivation for a specific type of physical activity, maybe because the respective person considers this activity unbearably boring, may actually result in a lower *TOC* in comparison to another activity with significant preconditions, but one for which the person is highly motivated.

those of single long sessions of their combined duration [WKO+99]. Consequently, instead of trying to activate the user to perform a single activity session of 30 minutes length or more, it may prove advantageous to instead encourage her for doing multiple short sessions distributed over the course of the day, such as two 10 minute episodes of brisk walking.

Table 5: Relevant Physical Activities.

Physical Activity	**Precondition**	**MET**
Tai Chi	None	3.0
Moderate Calisthenics	None	3.8
Brisk Walking	None	4.3
Single Moderate Dancing	Music	5.0
Weight Lifting	Weights	5.0
Jogging	Training Clothes	7.0
Recreational Skating	Inline Skates	7.5
General Bicycling	Bicycle	7.5
Running Up Stairs	Stairway	8.8

A selection of medium-intensity and vigorous-intensity physical activities with no or few preconditions, adapted from [AHH+11]

The considerations laid out in the earlier parts of this chapter revealed that the core element of any interventive measure is the *trigger*. Without triggering a person, the occurrence of the intended behavior cannot be controlled. Of course, the target person may also show the desired behavior without experiencing an external stimulation. Employees may choose to work overtime without their boss explicitly asking for this, just as a driver may choose to refuel her car without perceiving a blinking tank light. In such cases, the interventive measure is simply not needed and any intervention attempts undertaken regardless are unnecessary and potentially even harmful, as they harbor the danger of being deemed inappropriate. However, when the desired behavior is not occurring by itself, a trigger is required to ensure that the target person at least *considers* behaving in the intended way. And while interventive measures may additionally try to increase a person's ability and/or her motivation for the desired behavior in order to raise the odds for successful triggering, this is an optional step of oftentimes obscure effectiveness. Triggering, however, is a must and whether a trigger was successful or not can be assessed almost immediately (30).

Successfully triggering a person for a desired behavior such as brisk walking requires an understanding of the different situations that this person goes through and an understanding of what distinguishes a good – *kairotic* or *parakairotic* – situation from a bad – *parachronotic* or *chronotic* – one. More specifically, it requires the ability of estimating both the target person's ability and motivation for the desired behavior, as, according to the FBM, the product of the ability function and the motivation function determines the success of triggers. We have already found that accurate triggering actually requires not one but three different abilities. First, the interventive measure needs to be able to obtain all parameters that affect the target person's ability or motivation. A measure that is not capable of this suffers from a *lack of insight*. Even more crucial may be the second ability, however, which is to be able to make sense of the gathered information. Simply knowing that a person is 'in a vehicle' does not do any good if this knowledge is not linked to an understanding that one cannot perform certain activities when being 'in a vehicle', such as to stand up and to walk around. If this understanding is missing and the interventive measure thus suffers from a *lack of reason*, then meaningful triggering decisions cannot be expected from it. And finally, if a measure cannot correctly judge the importance of knowledge, then this *lack of experience* may lead to situations in which it erroneously assumes that a certain parameter, such as the target person's current posture, is equally important for the determination of her ability and motivation as another parameter, such as the local weather conditions.

30 As explained earlier, we will assume that the 'reverberant effect' of a trigger will last for no more than an hour, which is very likely to be a simplification of reality. See chapter eight for some more thoughts on this matter.

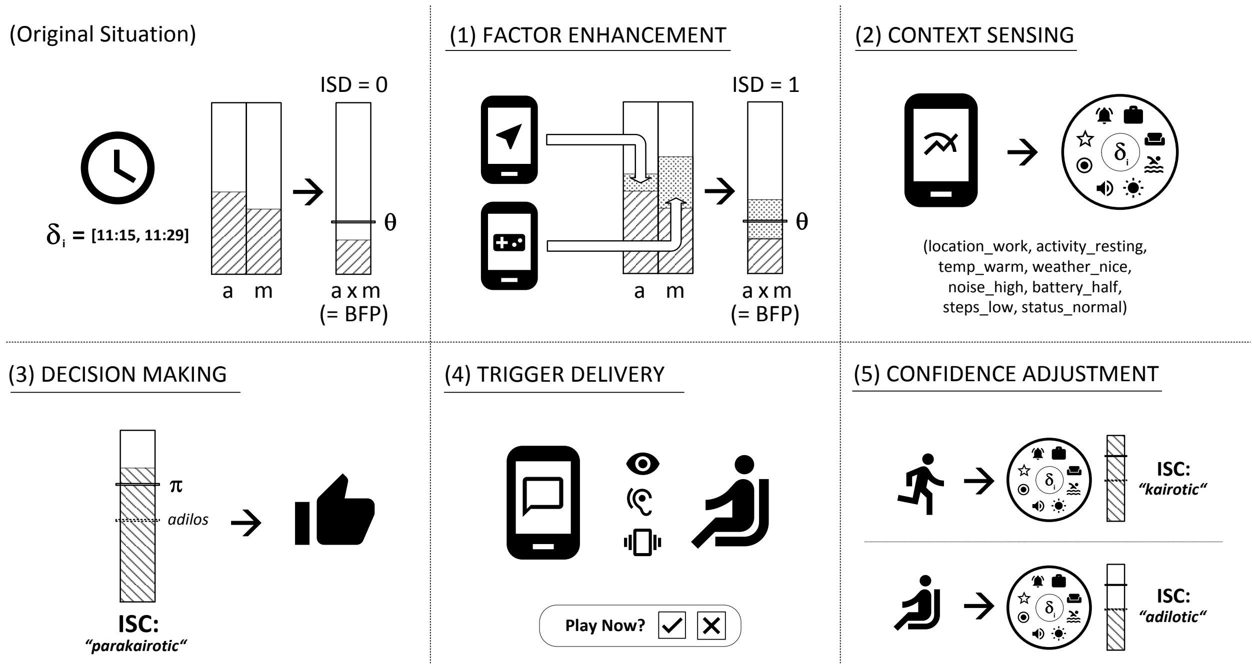


Figure 5: Five-Stage Behavior Intervention Process.

FIGURE NOTES – The figure shows the five stages of the behavior intervention process on the example of a mobile exergaming application meant to foster physical activity. The square in the upper left displays the original situation in which the target person is still uninfluenced by the measure. In stage (1) of the intervention process, the target person's ability and motivation for the desired behavior may or may not have been affected by the measure. In stage (2), the measure gathers information on the current situation, and based on this knowledge, it decides in stage (3), whether or not an intervention attempt seems to be reasonable. If this decision is positive, as in the example, then the actual trigger is delivered using one or multiple communication modalities during stage (4). Finally, the measure observes the outcome of its intervention in stage (5) and learns from it by adjusting its confidence value (ISC value) for being able to successfully trigger this specific user for the desired behavior in the perceived situation.

The difference between the *lack of reason* and the *lack of experience* is that the earlier is a general problem that requires nothing but common sense to resolve. This problem is also not related to specific target persons – no one can stand up and walk around while driving in a car on the highway. In contrast, the *lack of experience* will affect almost any measure until it has been able to gather extensive experience in triggering a specific target person (or a group of target persons) for a specific behavior. Resolving the *lack of experience* is only possible by adapting to an individual and by learning her preferences and priorities, usually through an extended trial-and-error process. On the negative side, this means that for its first number of intervention attempts, any interventive measure will suffer from this problem. On the plus side, however, the *lack of experience* is the only representative of the 'three troubles of triggering' whose severity can be reduced, namely by gathering experience. In other words, a discerning measure may be able to improve its intervention accuracy by learning – if it *can* learn.

Figure 5 depicts the entire five-stage process of a behavior intervention. In the example, the pervasive discerning measure is a smartphone application meant to encourage the desired behavior of 'brisk walking'. However, the same general procedure holds true for all other pervasive discerning measures, irrespective of whether they are human or technological. All of them, being pervasive, will accompany the target person through a number of different situations. Each such situation is characterized by a specific combination of the target person's ability and motivation for the desired behavior. While in some situations her motivation for the behavior may be high, it will be low in others, and the same goes for her ability. According to the FBM, it is the size of the product of these two factors – as specified by the *BFP* function – which decides, whether or not an intervention attempt will be successful in initiating the desired behavior. The problem of an interventive measure lies in not knowing, just how high or low this value is exactly. It may hope to be able to increase the target person's motivation and ability by either trying to create extrinsic motivators, by strengthening intrinsic ones, and/or by somehow simplifying the desired behavior (see the second chapter for more information on these topics). But still, the original problem will remain: The measure will be uncertain about whether or not the current value of the person's *BFP* function exceeds her activation threshold θ and thus, whether successful triggering is even possible in the given situation. In the example, the influence of the smartphone application indeed increases both the user's motivation and her ability for the desired behavior of 'brief walking' and in doing so lifts their product above the activation threshold θ. This means that a trigger initiated by the measure at this time would lead to the desired behavior. Knowledge that the measure does not have.

Instead, it must try to deduce the chances for a successful intervention from contextual information. More specifically, it must observe those 'indicators' of which it assumes that they will have a say in determining the target person's reaction to an intervention attempt. In the example, the application makes use of various internal smartphone sensors and Web services to assemble an 8-tuple of indicator values; this tuple contains the measure's *entire* knowledge of the current situation. Every decision made by the measure will be solely based on such an indicator tuple, although it is likely that other factors exist that influence the user's ability or motivation for the desired behavior as well. This *lack of knowledge* is a very typical problem for technology-based interventive measures, not the least because they have no means of automatically determining the user's emotional state, which in turn is crucial for assessing a person's motivation for many types of behavior. Humans have a significant advantage here in that they can rely on their empathy to take educated guesses on how high or low their opposite's motivation for a specific behavior may be. In contrast, technology-based measures such as smartphone applications must work with what they have got, which more often than not is just a gross simplification of reality [(31)].

Based on the knowledge that it has available, the measure then estimates the success chances of an intervention attempt undertaken in the given situation. The *intervention success confidence function ISC* represents this act of considering the odds for a successful intervention. In the case of a *fully-accurate* measure whose triggers are always successful, the *ISC* function will perfectly resemble the *intervention success determination function ISD* and thus always return either a 1, or a 0. The *ISC* function then only differentiates between *kairotic* situations, in which triggers will succeed, and the opposed *chronotic* situations, in which triggers will fail, and the measure's assumptions will always be correct. In such rare cases, the act of triggering does not involve any kind of uncertainty. But much more often than not, an interventive measure will distinguish between at least five different cases. Based on its contextual knowledge, it may still be convinced that, in a given situation, a trigger must be successful or must fail, which means that the *ISC* function can still assume its extremes. Usually, however, the measure will only be able to state a tendency, namely that a trigger is more likely to succeed than to fail (represented by an *ISC* value above 0.5, but below 1.0), or *vice versa* (represented by an *ISC* value lying in between 0.5 and 0.0). As stated earlier, these situations are called *parakairotic* and *parachronotic*, respectively. Finally, there may also be situations when the interventive measure is undecided, whether or not an intervention attempt will be successful. Such a moment is called an *adilotic situation*, a

[31] We find that there is a certain resemblance here to Plato's famous *Allegory of the Cave* [Baa97]: Just like the prisoners that can only see shadows on the cave wall but not the outside world that casts them, the technology-based measure does not perceive reality, but only a reflection of it that cannot depict all aspects required for truly understanding the situation at hand.

situation of absolute uncertainty (represented by an *ISC* value of 0.5). These situations are the most problematic ones, as they force a measure to entirely resort to guessing. It will then be decided by its 'ideal', whether or not it deems an intervention attempt to be reasonable.

The exact meaning of 'reasonability' is thereby dependent of what the measure aims to achieve. A measure that strives for *full accuracy* may pass on many situations of uncertainty and only issue triggers when it assumes the occurrence of *kairotic* situations. This way, it minimizes the chances for triggering the target person in situations when she will decline the trigger. On the opposite, a measure that needs to be *effective* will usually seize every single opportunity for triggering and not even refrain from *chronotic* situations, as it must fear having made wrong assumptions about the target person's current motivation or ability and that it may thus miss a moment in which the initiation of the desired behavior would have been possible. The triggering strategy of a measure that is supposed to be *reliable* will usually be a compromise lying in between these two extremes.

All three traits, accuracy, effectiveness, and reliability may have their legitimacy, depending on the target person and the desired behavior. A high accuracy is important, if the target person has a low tolerance to ill-timed intervention attempts. As pointed out before, the frequent output of futile triggers may eventually annoy the target person, which in turn may make her get rid of the interventive measure [32]. Likewise, effectiveness is important if the reliable occurrence of a behavior is required. In a way, the third trait reliability is a compromise between the other two, as it allows a measure to pass on situations with a high uncertainty, but also grants it a (limited) number of unsuccessful intervention attempts before it must fear the target person to draw consequences. Reliability merely requires that during the observed time period, at least one intervention attempt will be successful in initiating the desired behavior. As such, the design of *reliable* technology-based interventive measures is the easiest problem of the three and luckily, such measures will also suit the majority of application scenarios. As stated in the first chapter, this thesis focuses on the construction of such reliable interventive measures for the initiation of medium-intensity physical activities, specifically brisk walking.

In the example of Figure 5, the smartphone application assumes a *parakairotic* situation. This means that while based on its understanding of the world and its past experiences with this specific user it leans more towards the assumption that a trigger will indeed be successful in initiating the desired behavior, it is not entirely certain. Still, the degree of its uncertainty is low enough such that it decides in favor of an intervention attempt. Technically, this decision process translates to the *ISC* value that the application attributes to the given situation (more specifically: to the perceived 8-tuple of indicator values) to exceed a certain threshold, the *confidence gate* π [33]. The confidence gate is the border that separates merely considered intervention opportunities from actual intervention attempts. Triggering decisions depend on whether the *ISC* value that a measure attributes to a situation, the assumed chance for a trigger to be successful in this situation, is higher than this confidence gate. If this is the case, then the measure will issue a trigger. If it is not, then the measure is well advised to withhold the attempt. Defining confidence gates for technology-based measures is a non-trivial problem, and one that is always dependent on the corresponding measure's desired primary trait. For a measure that is supposed to be *accurate*, the confidence gate will be high, whereby the confidence gate for measures that need to be *effective* is best set to a low value. The problem of how to define confidence gates for *reliable* measures is a matter that we will return to in the next chapter.

If the triggering decision is positive, as in the provided example, then the measure needs to actually deliver the trigger. As pointed out earlier, a trigger must be both perceivable and clearly associated to the desired behavior. We find, however, that there is a third-requirement specific to technology-based interventive measures: The trigger must be timely. If, for whatsoever reason, there is a delay between the third and the fourth step of the intervention procedure (between *context sensing* and *decision*

---

[32] Albeit this is a mainly psychological phenomenon, it is nevertheless of a certain importance here. We will thus encounter it again in chapters four, seven, and eight.
[33] From Gr. πύλη (puli), gate.

making), or between the fourth step and the fifth (between *decision making* and *trigger delivery*), then this brings the danger of the intervention attempt to have become inappropriate because of a change of situation. If, for example, the measure determines that the target person is at home and sitting on the couch, but delivers the corresponding trigger meant to encourage brisk walking with a couple of minutes delay, then in the meantime the user may have left her house and may already be driving in her car. Such delayed triggers are bound to fail. But alas, although the immediacy of trigger delivery is essential for successful interventions, various reasons can lead to a technical device not being able to guarantee for this.

Computational devices such as smartphones usually have a number of options at their disposal in regard to reaching out to the user, among them text-messages prominently appearing on the device's main screen (the so-called push notifications), sounds being played from the device's loudspeaker, and possibly also haptic feedback, most notably vibration alarms. Multimodal communication is of particular importance for pervasive measures, as the user may have to be contacted in situations where she is unlikely to perceive one of the communication channels, such as sound. Indeed, adjusting the feedback type to better fit the user's current situation was one of the earliest application scenarios of context aware computing [BCE+01]. Creating both effective and appealing methods for human-computer interaction is a research field of its own, however, and the problem is thus largely sidestepped here by simply relying on all available modalities at once when it comes to triggering the user. The challenge of the design and the dynamic adaptation of communication means that are suited for smartphone-based interventions is a problem best addressed by experts of the domain.

The final stage of the behavior intervention process, and a step not actually performed by every measure, is to learn from the user's reaction to the intervention. More specifically, this step involves monitoring the user's behavior [(34)] after she has perceived the trigger and, if adequate, to then adjust one's confidence in whether or not the perceived situation was suited for making interventions with the goal of provoking the desired behavior from the target person. Technically, this means that a measure will either increase or decrease the *ISC* value that it attributes to the indicator value tuple that describes the situation. If the trigger was successful and the user did indeed change her behavior in the intended way, then this should strengthen the measure's confidence in this situation and thus lead to an increase of the *ISC* value that the measure associates to the respective indicator value tuple representing the situation. Analogously, if the user did not change her behavior in the desired way after perceiving the trigger, then this hints at the fact that the situation may actually not be a good opportunity for trying to make the user behave in the desired way and consequently, the measure may want to lower the *ISC* value that it associates to it [(35)]. The exact adaptation strategy, the speed with which the measure changes its established beliefs based on the acquisition of new experience, will thereby depend on the measure's goals and its applied method of learning.

After the completion of the final step, the intervention process starts anew. The following chapter treats three of the five stages of this process: *Context sensing*, *decision making*, and *confidence adjustment*. The conceptualization and prototypical implementation of a reliable triggering mechanic is the main contribution of this work. In contrast, chapter six focuses on the secondary goals of increasing a person's motivation and ability for a few minutes of brisk walking. The stage of *trigger delivery* is largely ignored here and rather left to experts of human-computer interaction.

[34] This may actually be easier said than done. Depending on the desired behavior, determining its occurrence with the means that are available to the measure may not be possible. This is especially true for technology-based measures that must rely on sensors, Web services, and user input for information about external events. If none of these sources provides the desired knowledge then the measure cannot verify that the trigger has been successful.

[35] All these considerations are based on the assumption that human behavior is rational and thus predictable. In other words, they are based on the assumption that given a specific target person and a specific desired behavior, triggering this person in comparable situations will always lead to the same reaction. While this may be the case, it seems just as likely that the model of an entirely rational actor is actually an improper simplification, much like the *homo economics* [HBB+01]. Still, the assumption of reproducible behavior is a necessity for the development of technology-based interventive measures, an issue that will be addressed again in chapter four.

4. User Activation

Based on the findings of the related work as described in the second chapter and the general considerations on the nature of behavior interventions as laid out in the third, this chapter now delves into solving the problems encountered when trying to actually create technology-based mechanisms for overcoming opportunity-related barriers through the production of well-timed triggers. The concept presented here and its evaluation, as described in chapter seven, can be considered to be the main contributions of this thesis.

4.1 Assumptions, Restrictions, Challenges

Chapter three showed that the development of technology-based interventive measures can be a complex and challenging endeavor. Especially the implementation of user triggering mechanisms capable of the production of well-timed notifications for the generation of short-term awareness and for pointing out opportunities for a desired behavior requires finding approaches for overcoming a multitude of technical challenges. Furthermore, triggering mechanisms are difficult to generalize – successful solutions will usually be specialized on the activation of a *specific* user group and on the initiation of a *specific* behavior. We will thus begin our investigation with gathering a list of all employed assumptions and restrictions that sharpen the presented concept's focus. The subsequent sections of the chapter will then treat individual aspects of this triggering mechanism.

(A1) VALIDITY OF THE FBM – All considerations described in this document on how to make a target person show a desired behavior are based on the assumption that the Fogg Behavior Model FBM [Fog03] is correct. This is the axiom from which the rest of this work springs. More specifically, this thesis assumes that a person's motivation and ability for a specific behavior can both be quantified and that the inclination of a person to show this behavior is determined by the question, whether or not the product of the two factors exceeds another value, the activation threshold θ. Starting from this notion, we define the 'behavior factors product function' *BFP* that assumes the value of the aforementioned product of a person's ability and motivation for the desired behavior. From the *BFP* function, we proceed to deduce the 'intervention success determination function' *ISD* that specifies for a given point in time, whether or not the *BFP*'s value lies above the activation threshold. The 'intervention success confidence function' *ISC* of a technology-based triggering mechanism aims to resemble the *ISD* function as much as possible. The question of what must be done to actually achieve this is this chapter's main subject [(36)]. All these considerations are meaningful if and only if the FBM is correct.

(A2) HUMAN BEHAVIOR IS RATIONAL AND PREDICTABLE – This is an admittedly problematic assumption, but a necessary one. Only if a behavior can be reliably reproduced given the exact same circumstances, then measures will be able to improve by adapting to the individual user through learning from her reactions to intervention attempts. Otherwise, if human behavior is *not* entirely predictable, the triggering process will always involve a level of uncertainty, regardless of the amount of information that is available on the current physical and emotional state of the user and on the state of her environment [(37)].

[36] Figure 6 depicts the hypothetical *BFP* function of an office worker who is supposed to be triggered for a 15-minute episode of brisk walking by a smartphone application. The upmost graph shows the *BFP*'s corresponding *ISD* function, and the central and the lower graphs show two possible *ISC* functions of the mobile triggering application, with the upper one considering an intervention attempt every sixty minutes, and the lower one twice as often. See chapter two for more information on the FBM, and chapter three for details on the calculation of the *BFP*, *ISD*, and *ISC* functions.

[37] As discussed in the third chapter, this uncertainty may be inherent to technology-based interventive measures for many more years to come, as some types of relevant information, especially knowledge about the user's emotional state, cannot be assessed automatically. Nevertheless, only if human behavior is predictable then fully-accurate triggering will ever become a possibility.

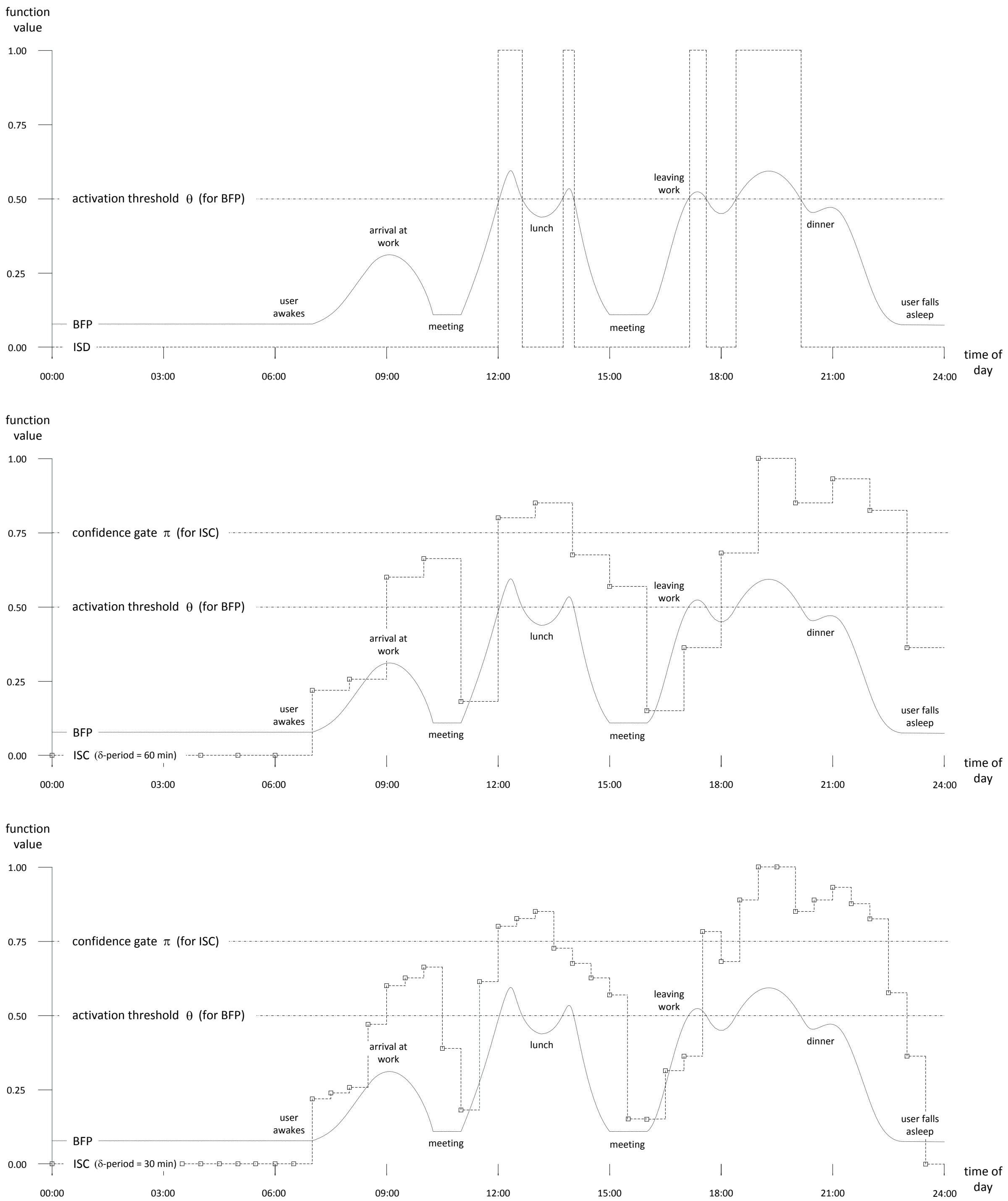


Figure 6: BFP, ISD, and ISC Functions.

FIGURE NOTES – Comparison of a hypothetical BFP function, a corresponding ISD function, and two possible ISC functions. In this example, the interventive measure is a smartphone application, the target person is an office worker, and the desired behavior is 'brisk walking'. The δ–period of the upper ISC function is 60 minutes, the δ–period of the lower one half that value. Note that both ISC functions undershoot and overshoot the BFP function several times (for example at 17:00 hours and at 21:00 hours, respectively), but that they both correctly indicate a kairotic situation at 19:00 hours. Due to its higher duty cycle, the 30-minute ISC function is able to (correctly) indicate a second kairotic situation at 19:30 hours.

(A3) EXISTENCE OF OPPORTUNE MOMENTS – We will assume that there is at least one situation during the observed time period (the *O*-period) in which a successful intervention for making the target person change her current behavior to the desired behavior is possible. In other words: We assume that the *O*-period's total opportunities counter *TOC* is larger than zero, and that there is at least one point in time when the product of the user's motivation and ability for the desired behavior (here: 15 minutes of brisk walking, see restriction R1 below) is higher than the user's activation threshold θ. If this is not the case and if there is not a single such opportunity for achieving the desired behavior, then no interventive measure can ever be successful [38]. This assumption also has implications on the length of the *O*-period. In our specific case, a full day of 24 hours seems to be the most straightforward pick for this value, as the performance of one 15-minute episode of brisk walking per day is both practicable and sufficient for meeting the minimum requirements for health benefits [WWT+11]. Furthermore, the consideration of individual 24-hour periods (more specifically, of the user's awake time during such periods) is in line with other research, as the next assumption reveals.

(A4) THE REVERBERANT TRIGGER EFFECT LASTS FOR ONE HOUR – This specifies the duration after which a measure can safely assume that the desired behavior will no longer occur as a consequence to its most recent intervention attempt. As such, it also defines the maximum length of the δ–period, because in order to keep the user under a constant 'trigger effect', a new trigger must be produced at least once per hour. This amounts to a total of twelve triggers per day when one calculates with a ten hour resting period plus two one-hour 'periods of grace', one right before, and one right after the break. These values are not randomly chosen: We find that the 'once per hour and up to twelve times per day' intervention strategy is being applied by various commercial products and scientific prototypes. Among others, the *Activity App* of the *Apple Watch* tries to make the user stand up for one minute once every hour for twelve hours in a row [DGH+16], while the *NotifyMe* application by Mehrotra et al. sends notifications to the user once every sixty minutes between 08:00 hours and 20:00 hours [MMH+15]. By taking together the assumptions A3 and A4, we can thus state that we consider one-day periods (*O*-periods) made up of 24 one-hour intervals (and possibly more, one hour is just the upper bound of the δ–period).

(A5) USERS ARE TOLERANT TOWARDS ILL-TIMED INTERVENTION ATTEMPTS – Receiving multiple triggers per day from a single interventive measure, as considered in the previous two assumptions, certainly requires some good-naturedness from users. Thin-skinned characters may easily become aggravated by the frequent disruptions and as a consequence, they may try to get rid of the nuisance. This is especially critical for mobile applications, which can be turned off or, even worse, be entirely uninstalled. By making the assumption that the group of target persons is willing to tolerate up to twelve intervention attempts per day, the 'high risk problem' that comes from impatient users is practically sidestepped here. Determining the maximum number of tolerable intervention attempts – and how to reduce the number of ill-timed interventions – must nevertheless remain a central aspect of all considerations on how to create discerning interventive measures, and it will be discussed again later in this chapter (see challenge C2 and section 4.4). The topic also plays a significant role in the interpretation of the evaluation results as presented in chapter seven.

(A6) HIGH PERVASIVENESS OF SMARTPHONE-BASED MEASURES – The most basic requirement for the success of pervasive interventive measures is the existence of a pervasive platform on which they are based on and that accompanies the target persons throughout the day and the majority

[38] Unless, of course, the measure has the means of increasing the target person's motivation and/or ability for the desired behavior, see challenge C9.

of relevant – *kairotic* – situations. We have already found that mobile devices, most notably smartphones, have the highest pervasiveness quotes *PQs* of all contemporary computers. However, it would not be sufficient if such technologies would merely be *more* pervasive than any other device type such as desktop computers or video game consoles. Instead, we need to make the assumption that their pervasiveness is so high that they will miss almost no opportunities for intervening; in other words, we will assume that the missed opportunity quotes *MOQ* of smartphone-based interventive measures will tend towards zero. And although initially this was indeed mainly an assumption [(39)], the evaluation results of the field test strongly support this theory, at least for a specific group of users [(40)].

In addition to these assumptions, we will also have to employ several restrictions. They are necessary for sufficiently narrowing down the problem so that the development of a concrete model becomes viable.

(R1) THE DESIRED BEHAVIOR IS BRISK WALKING – Interventive measures cannot be generalized. As explained in chapter three, the outcome of an intervention attempt, specified by the 'behavior transformation function' $\mu(u, \bar{b}, \delta)$, depends on the individual target user u, the point in time δ when the intervention takes place [(41)], and the desired behavior $\bar{b}$ that the intervention is supposed to initiate. Given a specific target user or a group of users, we can influence two factors: First, the behavior that is supposed to be triggered and second, the point in time when this triggering takes place. Consequently, if we also fixate the point in time δ, then the success chances of an intervention will entirely depend on the desired behavior. We found in chapter two that a variation of the target behavior may lead to vastly differing success rates [(42)]. It thus seems necessary to define a specific target behavior before starting with the conceptualization of a triggering mechanism. For reasons detailed in chapters one and three, this behavior shall be the physical activity of 'brisk walking' – more specifically, a 15 minute episode of brisk walking.

(R2) TARGET USERS ARE YOUNG ADULTS – People are different and one of the most obvious (and inevitably) differences lies in their age. A person's age may have an influence on what motivates her, and on what she finds simple and hard to do. In addition, age may hint at what a person's daily schedule looks like, at what technologies she uses and how, what communication modalities work best for her and how these must be designed, and so forth. In other words: Age can (but most not) have an effect on many of the factors that are relevant for the successful triggering of a target person with a technical device. As such, it seems meaningful to focus on a specific age group and for us this shall be young adults and adults between 20 and 39 years of age [(43)]. This restriction will play a significant role in chapter six, when we consider means of increasing our target persons' motivation and ability for brisk walking.

---

[39] The initial unconfirmed assumption of an overall high pervasiveness of smartphones and mobile devices was based on previous considerations and research [DKH+13, ICS+15].

[40] Question pre14: "*I'm always carrying my smartphone with me and I take it everywhere*", five point disagree-agree Likert scale, $N = 30$, $\mathrm{M} = 4.67$, $\mathrm{SD} = \pm 0.66$.

[41] As explained in chapter three, the specification of a point in time δ is sufficient for the specification of the entire contextual situation of the target person, including her current (uninfluenced) behavior.

[42] Compare, for example, the chances of making a specific person share a coffee with you with the odds of making that same person join you for a 10-mile run, both during a Monday's lunch break.

[43] An important question can be asked here: Whether this really is a relevant target group. It goes without saying that there is little point in the conceptualization of solutions that are trying to solve problems that do not exist. As such, if our target group of young adults does not require measures that help them to increase their levels of daily physical activity, simply because they are already sufficiently active by themselves, then they would have been a bad pick. The study by Haase et al. that was mentioned in the introductory chapter surveyed the sedentary behavior of almost 20,000 university students. They found that only 28% of the male students and 19% of the female students met the recommended levels of leisure time activity [HHS+04], which shows that there is undoubtedly a need for increase here. As such, focusing on young adults as target persons is certainly not a *bad* decision. The question, however, whether they also make up the most urgent age group in regard to the prevalence of physical inactivity and the success chances of technology-based interventions is left open for further discussion.

(R3) THE CONCEPT IS FOR SMARTPHONES – On a sufficiently high abstraction level, the process that needs to be followed in order to make successful behavior interventions is the same for all types of computational devices and resembles the one depicted in Figure 5: First, information about the user and the state of her environment needs to be gathered – the more, the better. Next, based on this information, on its understanding of the world, and on its past experiences, the measure needs to decide, whether or not the undertaking of an intervention attempt is reasonable in the current situation. If this is the case, then the subsequent step is to produce a trigger that can be perceived by the user and that she clearly associates with the desired behavior. As a last step, the measure will ideally also be able to observe the effects of its triggering attempt and to draw adequate conclusions for later intervention considerations. However, this general procedure is not sufficiently specific to allow for an actual implementation, as it leaves too many questions open. A first step to solve this is to narrow down our investigation to a specific type of device, and for several reasons, the choice falls on smartphones. The more of the relevant smartphone characteristics that another device type – possibly one that still needs to evolve – shares, the simpler the transfer of the presented concept to this other device will be [44].

(R4) RELIABILITY IS SUFFICIENT – As discussed in chapter three, a discerning interventive measure can adhere to one of three ideals: It can either try to be accurate, or effective, or reliable. A high accuracy implies that intervention attempts rarely fail, while a high effectiveness means that almost every single opportunity for initiating the desired behavior is successfully grasped. Although possible, it is difficult to achieve both accuracy and effectiveness at the same time, because as soon as uncertainty is involved in the decision making process – which will be the case for the majority of target persons and desired behaviors – the two traits will require contradictory triggering strategies. In contrast, the requirements for reliability are significantly lower, as this characteristic simply demands that the target user is successfully triggered at least once during the *O*-period. Because it is sufficient to initiate the desired behavior of 'brisk walking for 15 minutes' once per day in order to profit from health benefits (see assumption A3), and because in comparison to the other two ideals, 'reliability' is considerably easier to achieve, we will focus on the creation of such a *reliable* intervention mechanism.

Summing up these assumptions and restrictions, we can now be a lot more specific on what this chapter is about: '*The conceptualization of a reliable pervasive smartphone-based intervention mechanism targeted at young adults between the age of 20 and 39 that aims for the successful initiation of a 15-minute episode of brisk walking once per day through the identification of situations most suited for triggering the user, under the assumption that the user of such a mechanism carries her smartphone with her at almost any time and that she is willing to tolerate up to 12 intervention attempts per day.*' A number of challenges can be deduced from this 'mission statement' to whose overcoming the better part of this chapter is dedicated to. In this regard, the following list sets the stage for the further discussion.

(C1) IDENTIFICATION OF KAIROTIC SITUATIONS – Just as there is a main assumption (namely that the FBM is valid, see assumption A1), there is also a main challenge to which all subsequent challenges are subordinate. Accurate triggering demands the ability of being able to reliably distinguish between well-suited and ill-suited situations for reaching out to the user. More specifically: It requires insight into the question, whether the product of the user's motivation and ability for the desired behavior – here 15 minutes of brisk walking – is high enough so that an intervention attempt will be successful. Technically speaking, this is the ability of calculating

[44] The 'relevant smartphone characteristics' in this context are pervasiveness, sensitivity, and computing capability – also see [DKH+13].

an *ISC* function in a way that it perfectly resembles the actual *ISD* function as only then, *kairotic* situations can be identified infallibly and triggering does not involve any uncertainty (also see Figure 6). The implementation of such a mechanism is certainly not impossible – provided that it is given access to all information and resources that it requires. The existence of all other challenges stated below is due to the fact that for the majority of target users and desired behaviors, this cannot be accomplished by today's smartphones or any other type of mobile device (or, as a matter of fact, by *any* kind of contemporary computational device), mainly because such devices lack the means of acquiring and processing all relevant information. This entire chapter is dedicated to the development of adequate strategies for compensating this shortcoming.

(C2) HANDLING HIGH RISK AND LIMITED PREDICTABILITY – In principle, every intervention attempt harbors the danger of annoying the user so much that she will try to get rid of the measure. For smartphone applications, this means that they may be temporarily disabled or, even worse, be entirely uninstalled. It can be assumed that this risk of annoying the user is much higher for unsuccessful intervention attempts than for successful ones, and obviously, the problem grows and shrinks with the user's patience and willingness to tolerate unsuccessful interventions. If we assume that a given user is only willing to tolerate a certain number of intervention attempts per observation period (such as a single day), then an discerning interventive measure will want to make use of these 'free throws' as best as possible. In other words, it will be interested in trying to activate the user only in those situations during the *O*-period in which it believes the chances for a successful intervention to be the highest. The problem lies in the fact that the measure will not know *a priori* at the beginning of the *O*-period, when exactly these situations will occur. Based on past experience, it may have reason to believe that their occurrence is more likely at certain times, than on others. But this does not guarantee that, for example, instead of grasping mediocre opportunities during the morning, waiting for the afternoon is the right thing to do. Just as well, the measure may find that there is not a single opportunity during the second half of the day, not even a mediocre one, because the day evolved differently than the measure had anticipated. In such a case the measure would have to choose between either making ill-timed interventions with a high failure rate, or decide to make no intervention attempts at all. This 'limited predictability' is a significant problem for discerning measures that try to reduce the number of ill-timed intervention attempts by focusing on the most promising situations while at the same time trying to ensure a minimum of successful interventions during an *O*-period. As pointed out before, the concepts discussed on the following pages are all based on the assumption that users are willing to tolerate up to twelve intervention attempts per day, even if all of them are unsuccessful (assumption A5). This largely sidesteps the problem of 'high risk and limited predictability' and it is of course a belittlement. Finding strategies for actually overcoming this challenge instead is undoubtedly essential for the creation of sustainable measures that are not being muted, deactivated, or uninstalled at the first opportunity. We will discuss this problem again later.

(C3) COPING WITH PARTIAL OBSERVABILITY – An interventive measure must be able of determining the user's ability and motivation for the desired behavior. Given the capabilities of today's technology, for the majority of desired behaviors this cannot be done in an automated way. Especially the user's motivation is nigh impossible to assess reliably, as technical devices are simply lacking the means for accurately determining a person's physical and – more importantly – emotional state. A measure thus has two options: To either ask the user [45], or to somehow

[45] Asking the user for a specification of her current motivation and/or ability for the desired behavior, in our case 15 minutes of brisk walking, is no guarantee for receiving the correct information, though. The user may either ignore the request for feedback, unintentionally provide wrong feedback, or simply lie (especially about her current ability if her motivation for the desired behavior is low and she fears that a truthful answer will lead to an intervention attempt by the device).

cope with the limited knowledge that it has available. This in turn means that the developers of such a measure must have provided it with the means to at least acquire all relevant information that *can* be acquired, and at the very least to be able to pick all 'low hanging fruits' (see section 4.3). If a measure is not granted this ability, then the severeness of its *lack of insight* will basically render all other efforts of improving the measure irrelevant.

(C4) HANDLING RESOURCE LIMITATIONS – The problem of suffering from resource limitations is typical for smartphone-based interventive measures. But this is not, as one could assume, mainly a problem of limited memory or processor power, but rather a problem of short battery life. Different to desktop computers and most video game consoles, mobile devices are usually not permanently connected to power outlets but for many hours of the day they rely on their internal battery instead. The acquisition of information that is required by the interventive measure to support its decision making process, however, will to a large part depend on the device's internal sensors and on Web services – and frequently accessing both these information sources will be a drain on the device's battery. Thus, finding an optimal compromise between the *duty cycle* of smartphone sensors/modules on the one hand and the acquisition of required information on the other is important, as a smartphone application that empties the smartphone's battery within the hour will not be well-regarded by the majority of users. On the contrary, Figure 6 shows the necessity of a high duty cycle: A measure's *ISC* function aims to approximate the actual *ISD* function as closely as possible and the measure with the one-hour duty cycle is a lot less flexible than the one with the thirty-minute duty cycle. The higher the measure's duty cycle, the better its *ISC* function can resemble the *ISD* function [(46)].

(C5) HANDLING COLD STARTS – As pointed out before, interventive measures cannot be generalized. They rather depend on the desired behavior that they aim to initiate just as much as they depend on the specific target user who is supposed to be activated. As such, a triggering mechanism that has perfectly adapted to one user and that has acquired a high reliability in that user's activation cannot be expected to achieve similar success rates for another person. Rather, the *lack of experience* will affect the accuracy of every measure when it starts over with a new user until it has been able to sufficiently adapt to her preferences. Due to this limited transferability of triggering expertise, however, 'cold starts' are a significant problem of interventive measures and during their first number of intervention attempts their accuracy will almost inevitably be low. This becomes a real problem when the assumption that users are highly tolerant towards ill-timed interventions proves wrong. In such cases, the measure may not be granted the chance to improve through an adaptation process but it may rather find itself being disposed after a short inspection phase. Strategies for reducing the severity of the 'cold start' problem are very important.

(C6) LEARNING FROM USER BEHAVIOR – It is clear that the intervention mechanism needs to be able to adapt to the individual user through learning, as this is the only way of reducing the *lack of expertise*. This necessity was already pointed out several times. While there are different ways of how this can be achieved, we find that in any case, based on our understanding of how interventive measures in general – and technology-based measures in particular – are organized, learning will always imply that a measure observes the consequences that an intervention

[46] Note that the duty cycle and the accuracy of the *ISC* function do not necessarily go hand in hand. A high duty cycle is just a principle necessity for the *ISC* function to be able to resemble the *ISD* function, at least if confronted with an *ISD* function that frequently alternates such as the one in Figure 6. Only with a high enough duty cycle, the steps of the *ISC* function *can* closely follow those of the *ISD* function. However, the question of whether they *will* do so is decided elsewhere. In the example of Figure 6, both *ISC* functions erroneously state triggering opportunities when there are none and at other occasions fail to point them out when they occur. While this is a hypothetical example, very similar function lines can be expected to be found in real life applications. The question of how closely a measure's *ISC* function is able to resemble the actual *ISD* function is then determined both by the height of the measure's duty cycle and its ability to assess and process the relevant information.

attempt has caused an then adjust its confidence on whether the situation at hand is a *good* situation for making such attempts. In other words, learning means to adjust the *ISC* value of the indicator value tuple that makes up a measure's impression of the current situation.

(C7) RECEIVING FEEDBACK – Dependent on the behavior that a measure aims to initiate, receiving reliable feedback on whether an intervention attempt was successful or not may be problematic. Especially the automatic monitoring of user behavior in order to determine, whether or not the user has changed her behavior in the intended way can be close to impossible, even with pervasive and sensitive devices such as smartphones. Certain activities, such as whether the user has drunk a glass of water, are simply not detectable with contemporary devices [(47)]. As such, the receival of feedback in order to learn from the user's reaction may pose a problem of itself.

(C8) HANDLING BEHAVIOR DRIFT – We have made the assumption that the user's behavior is rational and predictable – but even then it may change over time. Especially a person's motivation for a specific activity may fluctuate. What a person finds interesting on one day may be considered to be boring on the next, and *vice versa*. In part, this can be explained by the Flow Model that was discussed in the second chapter: Once someone gains a certain expertise in doing something, such as a climber that keeps ascending the same rock wall, she may lose interest in this activity and rather start to look for other challenges [Csi75]. As such, it is important for an interventive measure to not stop adapting to a user once a certain level of triggering accuracy has been achieved. Rather, a constant learning and adaption process is required.

(C9) INCREASING MOTIVATION AND ABILITY – Increasing the target user's motivation and/or her ability for the desired behavior is an optional step and not a challenge of user triggering in the actual sense. Nevertheless, such steps can help the user's *BFP* function to exceed the activation threshold a lot more often than it elsewise would and so can create opportunities for successful triggering where there originally had been none.

A non-technical challenge that shall not go unmentioned here is the need for overcoming privacy concerns of potential users. If ignored, then this has the potential of completely foiling all efforts of constructing successful interventive measures. It lies in the nature of context aware applications that they gather information about the user and her environment and indeed, the more information that can be acquired, the better the user experience provided by such applications will usually be. However, many users are very sensitive about the type of personal information that is being collected about them, as well as the *modus operandi* of how this is done. Sometimes, this rejection is so severe that it entirely prevents certain mechanisms, regardless of what their benefits over other approaches would be [(48)]. During the preparation of the field test that is described in chapter seven, we encountered the same problem. Several potential participants of the evaluation withdrew their application when they were informed of the types of information that our application *Twostone-IM* would collect about them [(49)]. Findings ways of informing users without 'scaring them off' seems to be an important aspect of the development of interventive measures.

[47] This statement does of course not hold true in all generality. Already more than a decade ago, scientists used body-worn inertial sensors to create systems capable of reliably detecting activities of daily living, most notably eating and drinking. For example, Amft et al. achieved a detection accuracy of almost 95% with a total of four of such sensors worn on both arms [AJT05]. However, such scenarios are almost always artificial in that they utilize devices that are not found in the average household. Interventive measures that are meant to be employed in real life must be able to make do with the means that are available in real life. And a smartphone tucked away in a user's pocket will not be able to reliably detect, whether the user has just drunk a glass water, brushed her teeth, or waved a friend.

[48] A good example for this phenomenon is the use of low-end off-the-shelf cameras for the creation of indoor localization systems. Although mechanisms based on such devices are cheap and comparably accurate, the majority of users is not willing to tolerate camera-based systems in their apartments and especially not in sensitive areas such as the bedroom or the bathroom – a problem that the author of this work has run into himself [BD13, MDW14, BD16].

[49] In the case of our study, especially the assessment of the user's current activity and of the ambient noise in her surrounding were claimed to be problematic by several potential participants of the evaluation.

4.2 Overall Concept

The considerations made up to this point allow for the deduction of a general procedure for the design of technology-based interventive measures. This section is dedicated to the description of this procedure on an abstract level, but also elaborates on how its individual steps were implemented to solve the problem at hand, namely the construction of a mobile measure for assisting users in increasing their overall levels of leisure-time physical activity. Figure 7 depicts the entire process whose six stages are discussed in the following.

(I.1) SELECT DESIRED BEHAVIOR – The first step in the creation of any technology-based interventive measure is the exact definition of the behavior that the measure is intended to bring forth [50]. It is safe to assume that in most cases, there will already be a rough notion of what the measure is supposed to achieve when one starts with its conceptualization. However, as chapter three clarifies, the selection of a general type of behavior will usually not be sufficient. Instead, narrowing down the focus to a single, very specific activity will help a great deal in the construction of a successful measure. To provide an example, a measure that is meant to ensure its users' sufficient hydration may have undesired side-effects when it makes them drink three bottles of sugary soft-drinks per day rather than three bottles of pure water. As such, in order for the implementation process to lead to a successful measure, its conceptualization should start with a precise formulation of the desired behavior that it is supposed to bring forth. In the case at hand, we narrowed down the desired behavior from originally '15 minutes of medium-intensity physical activity' to '15 minutes of brisk walking' [51]. See restriction R1 for details.

(I.2) SPECIFY TARGET USERS – The specification of the measure's target users is not a mandatory step, but it may help with further design choices. To a certain degree, the target users may indeed be predetermined by the selection of the desired behavior. For the creation of a measure that increases medium-intensity physical activity levels, we decided to focus on young adults, as made clear in restriction R2.

(I.3) SPECIFY PLATFORM – Since we intend to create a technology-based interventive measure, we also need to specify the technological platform that the measure is supposed to run on. Whatever the choice is, it should enable the measure to pervasively accompany the user so that the overall number of *kairotic* situations missed by the measure (the MOQ) is low. As such, the desired behavior itself obviously strongly affects what technological platforms come into question. Furthermore, certain target user groups may also impose limitations on this choice – not all technological platforms are equally distributed among the populace. As stated in restriction R3, for solving the problem at hand we decided for a smartphone-based measure, more specifically for an Android application. As it turned out, this was a choice with a consequence: It limited the amount of potential users of our measure and this led to difficulties during the acquisition of candidates for the field study, as explained in chapter seven.

(II.1) SELECT IDEAL – Once the desired behavior that the measure is supposed to bring forth is clearly defined, the next step consists in the specification of the measure's ideal. This decision is crucial, as it will have an influence on almost all subsequent choices, not the least because a successful

[50] The entire procedure detailed in this chapter is based on the assumption that the respective technology-based interventive measures is supposed to *increase* the prevalence of a desired behavior. As explained in chapter three, the counterpart to this goal, namely the reduction of an activity's prevalence (such as smoking), is not in the focus of this work.

[51] For reasons explained in chapters five and six, we will later change the desired behavior that our measure is supposed to bring forth once more, this time from '15 minutes of brisk walking' to 'playing a round of the mobile exergame *Twostone*'. Changing the desired behavior after the measure's design phase is completed is not without problem and should normally be avoided, unless the two activities are very similar to one another. At least from a physiological perspective, this is the case here.

accurate measure will usually be based on an entirely different intervention strategy than an effective measure: An accurate measure intervenes only when it is certain that its intervention will be successful, which may lead to it not intervening at all for an extended period if no opportune situations arise. In contrast, an effective measure will usually try to intervene as often as possible, as it must ensure that the desired behavior occurs at every given opportunity. For reasons explained by restriction R4, we decide for the implementation of a reliable measure that settles with a single successful initiation of the desired behavior of '15 minutes of brisk walking' once during the observation period.

(II.2) SPECIFY OBSERVATION PERIOD – The definition of the observation period (the *O*-period) is closely related to the selection of the measure's ideal. The *O*-period is the time interval that the measure takes into account when it makes decisions, for instance on how to best distribute a limited number of intervention attempts (see below). In theory, the *O*-period can be infinitely long. Usually, however, it will make sense to limit it to a manageable length. As stated in assumption A3, we have selected an *O*-period of 24 hours for our interventive measure.

(II.3) SPECIFY NUMBER OF ATTEMPTS – Another choice linked to the selection of the measure's ideal (and to the specification of the *O*-period) is the specification of an upper limit for the number of intervention attempts that the measure is allowed to make. This may place severe restrictions on the measure and substantially restrict its liberty of action – which may be meaningful. In theory, this parameter can be left undefined. In the case of accurate and effective measures, this will not be a problem, as these ideals already establish intervention strategies in their own right. However, for reliable measures that only need to initiate the desired behavior once during an *O*-period, a lacking upper limit of the number of attempts may lead to the measure pestering the user with ill-timed interventions to a point at which she decides to get rid of it. In assumption A4, we have decided for an upper limit of twelve attempts per day [52].

(III.1) SELECT INDICATORS – Indicators are the building blocks that make up the measure's understanding of the situation at hand. It depends on the desired behavior – and possibly the target user – which indicators are relevant and it is determined by the technological platform that the measure is based on, which indicators are obtainable. The meaningful selection of indicators is not as simple, as it may initially appear to be, as additional factors such as an indicator's impact on the user's decision and the costs of obtaining it must also be taken into account. Section 4.3 discusses this problem in detail and also states, what indicators our specific solution is based on (and why).

(III.2) HANDLE RESOURCE LIMITATIONS – As already pointed out, the obtaining of indicators is not equally difficult. The most preferable indicators are the 'low hanging fruits' which are both relevant and not very costly to assess. Many other indicators, however, will come at a price. Knowing the user's current GPS position, for instance, is essential for many context aware applications, but leaving a smartphone's GPS module active for a prolonged duration is guaranteed to quickly drain the device's battery. The definition of a duty cycle may help here. At the beginning of each such cycle, all relevant information is obtained and packed into an 'information batch' for further processing. Afterwards, however, the

[52] Triggering attempts serve a dual purpose. On the hand, they are the requirement for successful interventions – if no attempt is made, then the user's behavior cannot be changed. However, they are also a necessity for gaining confidence on which situations are suited for making such attempts, and which are not. A trigger that is limited to a low number of intervention attempts per day will be affected by the *lack of experience* much longer than a trigger that is allowed a significant number of attempts during the *O*-period. Balancing the three parameters 'user annoyance – intervention attempts – learning speed' is a non-trivial problem. Section 4.4 discusses a possible compensation strategy.

measure falls into a dormant state and awaits the end of the current cycle. For discerning interventive measures, this duty cycle resembles the measure's δ-period, as described in chapter three. Figure 6 illustrates, why the selection of the δ-period's length is a delicate matter.

(IV.1) SELECT DECISION MAKING – This step decides the measure's actual intervention strategy. Based on the indicators selected in the previous step, the measure needs to decide for each perceived situation (at the beginning of a δ-period), whether the confidence of the measure that an intervention attempt will be successful in this situation lies above a certain limit, the *confidence gate* π, whose value is determined by the measure's ideal and the number of the allowed intervention attempts in the current O-period. The calculation of the measure's *ISC* function that associates confidence values to situations can be achieved in a number of ways, some of which are discussed in section 4.4.

(IV.2) HANDLE HIGH RISK – Finding a strategy that selects the best opportunities for making intervention attempts during the O-period if the number of allowed attempts is limited is a difficult problem, as explained before.

(IV.3) HANDLE COLD STARTS – Another problem that is closely related to a measure's decision making procedure, the measure must somehow be able to compensate for the 'cold start' problem that inevitably affects it while it tries to adapt to a new user.

(V.1) SELECT LEARNING STRATEGY – A measure needs to be able to learn from its intervention attempts, both the successful and the unsuccessful ones. Technically, this means that the measure adjusts the confidence value that it associates to the perceived situation, but this can be realized in a number of ways, some of which are discussed in section 4.5.

(V.2) HANDLE FEEDBACK RECEIVAL – Being able to learn from the observation whether or not an intervention attempts was successful implies that this information can be obtained. If this is technically difficult (or even impossible), then a workaround must be found.

(V.3) HANDLE BEHAVIOR DRIFT – The user's behavior will change over time, but some learning mechanisms will try to hold on to a once established understanding of the world. This problem must somehow be handled.

(VI.1) CREATE CONSOLIDATING MECHANISM – As pointed out before, this step is optional and not actually part of the creation of an interventive measure – unless the measure is supposed to increase the usage of the consolidating mechanism, of course. However, creating a consolidating mechanism can be beneficial for a variety of reasons. This topic is discussed in the fifth chapter, which also introduces the consolidating mechanisms used for the interventive measure sketched out in this chapter, namely the Android-based exergame *Twostone*.

(VI.2) INCREASE MOTIVATION AND/OR ABILITY – In most cases, the intention behind the employment of a consolidating mechanism will be to create the basis for features that are supposed to somehow increase the target user's motivation and/or her ability for the desired behavior. While the reasoning behind this is understandable, the findings of our two studies, as described in chapters six and seven, point to the problem that such efforts may turn out to be counterproductive. Nevertheless, in some application scenarios the investment of effort into this step may prove worthwhile.

Each of the three remaining sections of this chapter is dedicated to one 'challenge group', namely the selection of indicators, the selection of a decision making procedure, and the selecting of a learning strategy. The fifth and the sixth chapter then focus on the procedure's sixth and optional step.

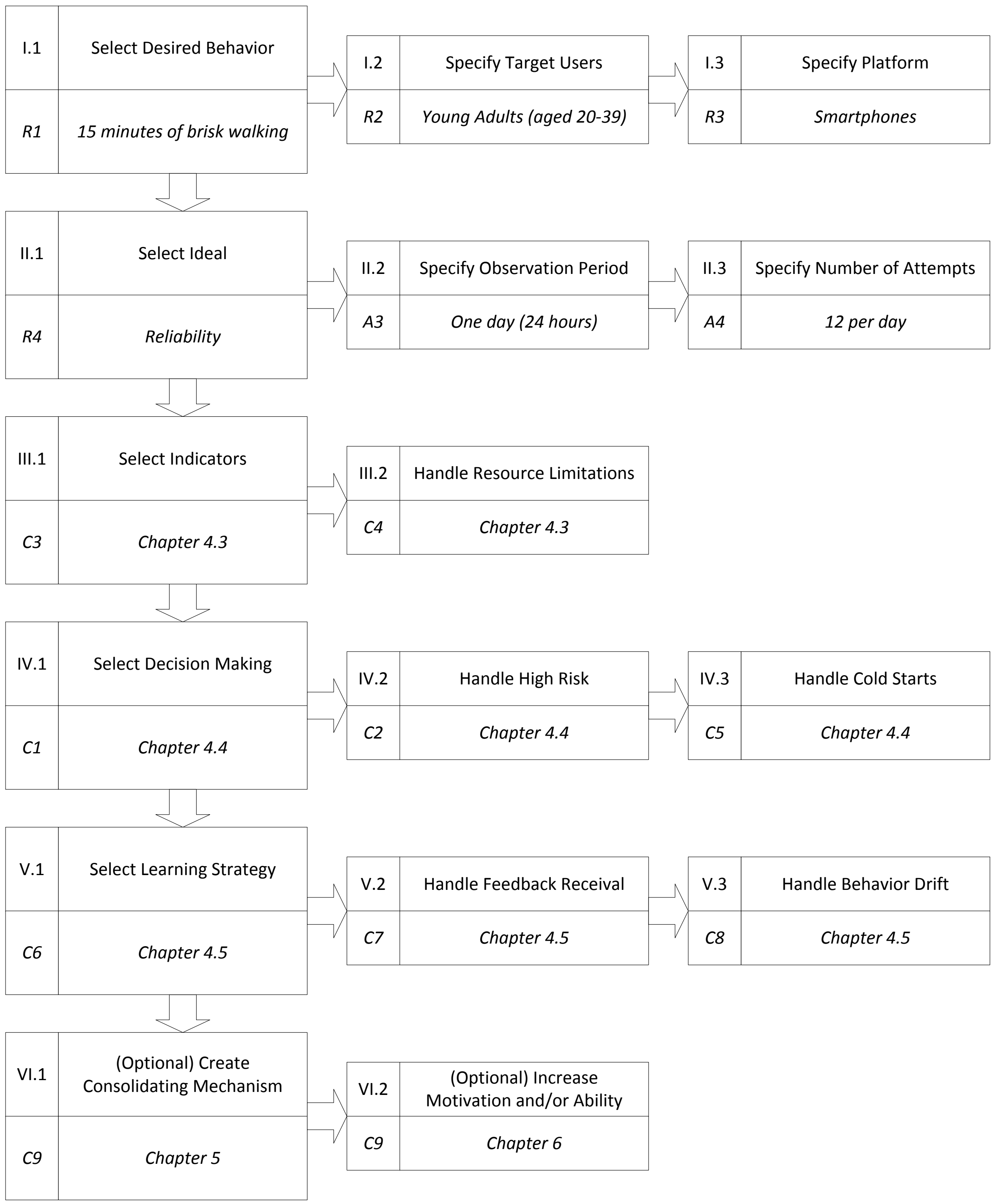


Figure 7: Technology-based Interventive Measure Design Procedure.

FIGURE NOTES – The figure shows the 6-step procedure of designing technology-based interventive measures, as developed in and applied by this thesis. While the upper part of each module specifies the respective procedure step, its lower part states, whether a decision for this step has already made and if so, what it is. Note that such decisions are always either related to an assumption (A), or a restriction (R), while all open problems have been formulated as challenges (C). Challenges are logically grouped together and each such 'challenge group' is addressed by an individual chapter, as its successful overcoming requires a detailed analysis and discussion.

4.3 Context Sensing

From a technological perspective, discerning interventive measures are context aware applications and devices. This means that such mechanisms have a certain amount of knowledge about the state of the user's – or, more specifically, their own – current environment and ideally also about the user herself, such as what she is doing, or how she is feeling. The main problem of discerning measures, and of many other types of context aware applications, is that this knowledge is limited. Such limitations may be voluntary, they may be necessary, and they may be unavoidable.

In the case of technology-based measures, two choices determine, what knowledge is relevant and to what degree such relevant knowledge can be obtained. The specification of the exact activity that a measure is supposed to bring forth will stake out the boundaries that separate the interesting from the irrelevant. For example, a measure that is supposed to ensure that a person consumes a sufficient amount of water will rely on a different set of information to recommend and recognize the activity of drinking than a measure that aims to increase the frequency of short sessions of brisk walking. While in the earlier case, information about the user's current location may be of importance, the local weather may play a role in the latter. In both cases, the outside temperature could be of interest. But none of the two interventive measures is likely to profit from knowing the user's bank account balance. Pinpointing the exact parameters that influence a person's motivation and/or ability for the desired behavior will oftentimes be difficult, however. While some factors, usually those with a high impact on the person's decision, may be rather obvious, the borders between the barely relevant and the mostly irrelevant factors are fluent. Deciding for a set of relevant indicators that have an influence on the user's motivation and ability for the desired behavior and separating those with a high impact from those that are only marginally relevant, is a challenge of its own and, depending on the behavior in question, may require the assistance of domain experts. For instance, only a psychologist or psychotherapist will be able to exactly identify the factors that a measure needs to be aware of for the meaningful recommendation of breaks in order to prevent burnouts, just as the driver assistance systems discussed in chapter two that are meant for the detection of driver drowsiness are based on the expertise of neuropsychologists [JPO+11].

But even if the relevant parameters are clear, it is likely that not all of them will be determinable. Which of the parameters can be obtained is rather dependent on the second design decision, the specification of the technological platform that the measure will be based on. Not every device can obtain each type of information, and, as pointed out in chapter three, especially the assessment of the user's physical and emotional state is technically challenging, even if constant advancements are made in this field (53). As one also needs to take into account several other factors for the selection of a platform besides sensitivity – such as the spread of the platform among the target users and its ability to pervasively integrate itself into their users' everyday lives – it may be necessary to pass on the option of obtaining certain types of information, even if platforms exist that can provide for this. For example, mobile sensor kits are capable of assessing a wide range of bio-signals that could be used for making reliable assumptions about a person's wellbeing. However, such systems are usually both costly and impracticable and as such, they will not be a good pick as a basis for interventive measures that are meant to integrate themselves into the lives of a large number of users. In our specific case, we selected smartphones as the platform for the development of interventive measures and what we thus may have lost in regard to sensitivity we have certainly won in terms of prevalence and pervasiveness.

[53] Scientific research is pushing the boundaries of what is technologically possible in regard to the automatic assessment of a user's health and wellbeing. A parameter that may be suited for making assumptions about the user's emotional stress is the heart rate variability (HRV), the range of the time intervals between a person's heart rate. The assumption is that a high HRV indicates that the respective person is healthy and feeling well, while a low HRV points towards increased stress levels [TAF+12]. While we have found that the current smartphone generations are not suited for correctly assessing HRV [KOM-M-0503], future generations with better cameras and higher frame rates may be able to do so. Once this is the case, the assessment of a user's HRV – and as such, making solid assumption about her level of emotional stress – may be as simple (and accurate) as the measurement of her heart rate by way of a smartphone camera via photoplethysmography [GDG16].

After having separated the obtainable knowledge from the not (automatically) obtainable by choice of platform, another factor comes into play: Cost efficiency. Usually, not all obtainable parameters will also be equally easy to assess. Rather, we can distinguish different categories of information with different levels of accessibility. Almost all types of information that are related to the system's state and currently running applications will be obtainable at 'no-cost', such as the current time. Table 4 that states the most frequently employed indicators in context aware systems names lists four of such inexpensive types of indicators, namely date/time, program status, user schedule, and system state (items 02, 04, 16, and 18). However, it lies in the nature of context aware applications to also require knowledge about events that are occurring outside of the device. Such information can either be gained from internal sensors or, if these cannot provide for it, possibly from external information sources such as the Internet, another device (most notably an external sensors), or the user. The most used type of contextual information is the user's location, which is usually determined by way of the device's built-in GPS module. But since this module is one of the most power consuming components in modern-day smartphones, alongside the Wi-Fi module and the cellular module [DBN12], it may actually be reasonable to pass on the assessment of this type of information if not absolutely necessary. Once again, the distinction between the relevant and the irrelevant plays a key role for make sophisticated design decisions that lead to successful measures [54].

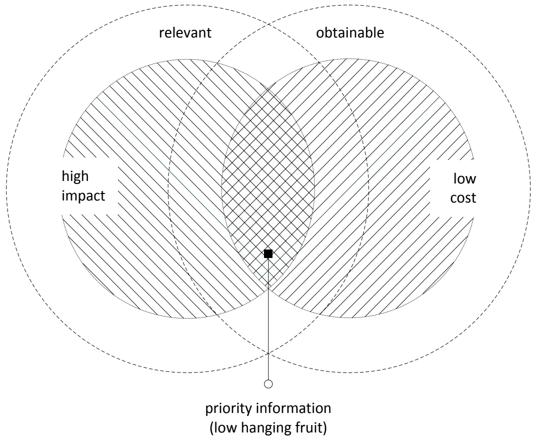


Figure 8: Indicator Types.

FIGURE NOTES – Categories of indicators from the perspective of a discerning measure. The intersection marks the priority information (the low hanging fruits) whose obtaining should be considered mandatory. Information lying outside the two circles is neither obtainable nor relevant and thus entirely negligible.

We thus find that we can distinguish five groups of context information, as can also be seen on Figure 8. As the two major categories, there is information that is *relevant* for the discerning measure, as it determines the user's motivation and/or ability for the desired behavior, and there is information that is (automatically) *obtainable* from measures that are based on the selected platform, either by way of accessing internally managed sources, through the utilization of internal sensors, or from external information sources such as other devices, Web services, or user input. In addition, relevant information may be about parameters with a high impact on the user, or not, and obtainable information may be costly to receive in terms of time, computing resources and – in the case of mobile devices most importantly – battery consumption, or not. The pairing of these different types of information makes the four general categories of indicators, whereby one of these categories is special: Those indicators that are both of high relevance and obtainable at a low cost. These are the '*low hanging fruits*' and their identification is the most crucial aspect of the third stage of the interventive measure design procedure. By contrast, there also exists a fifth category of indicators that can be totally ignored, as they are neither relevant nor obtainable (such as the user's bank account balance in above's example).

[54] In some cases, workarounds may be found that allow for obtaining the same type of information but at a much lower price. Exchanging the GPS information for the SSID (the identification code of the wireless network that the user's device is logged into) is an example for such a workaround and one that is also applied in the context of this work. In other cases, a more efficient implementation may also serve to help. Researchers have found that a large number of context aware applications are realized in a way that they practically waste the smartphone's resources, especially its battery life, without a real need [LXC13].

As pointed out, the distinction between relevant and non-relevant information may require the assistance of domain experts and is in most cases not a technical problem. It is thus not discussed in detail here, although we will briefly come back to it when deciding for the parameters that our application for increasing physical activity will be based on. However, the question of how to obtain such relevant information with technical means and, possibly, how to find ways of reducing the costs of obtaining certain parameters is a fundamentally technical challenge. The first question in this regard is of course, just how expensive a parameter is to assess. While the exact costs will depend on the platform and the means utilized for gathering the respective piece of information, as a rule of thumb it can be assumed that obtaining information about the system's state and currently running applications will always be less costly than employing internal sensors and modules, which, in most cases, will again come at a lower price than gathering information from any external source (excluding user input). However, this rule does not hold true in all cases and especially the complex processing of raw data acquired from internal sensors into high level information by way of machine learning algorithms may often prove to be a very costly endeavor (see below).

In those cases where parameters are identified that are both relevant and costly, one may try to find workarounds for obtaining the same information in another, more cost-efficient way. A good example for such a workaround, and an approach that was also employed for the realization of the interventive measure discussed here, is to not assess a user's location by way of her smartphone's GPS module, but rather by relying on the SSID of the wireless network that she is currently logged in (55). While this may not be a feasible solution for all types of context aware applications and especially not for all types of interventive measures, depending on the desired behavior it may sometimes be sufficient to just know when the user is at home or at work, as it is the case with the measure discussed here (56). Then, passing on the GPS access and instead relying on the SSID of the wireless network to assess location is a much more cost-effective approach. Of course, such workarounds may not always be possible and if a desired behavior requires one or more parameters that can only be obtained in a costly manner, then another strategy is required, namely the employment of a duty cycle.

A duty cycle, which is measured in percent, specifies the relative amount of time that sensors and modules are active and obtaining information. This implies that there are also episodes during which these components are switched off, usually in order to safe battery life and computing resources. During these 'energy saving episodes' the respective device or application is practically deaf and blind and as such cannot make reasonable intervention decisions. Depending on the desired behavior and the measure's intended ideal, these enforced interruptions may pose a significant problem. The duty cycle is closely related to a measure's δ-period, that is the interval between the points in time when the measure makes its decisions on whether or not to intervene. In chapter three, we discussed the example of a one-hour δ-period and found that during such a long time, one can be at work and talk to colleagues, drive home, take a quick shower, and be lying on the couch and watch TV. On the one hand, the fact that the user is undergoing so many different situations while the measure is 'asleep' means that it potentially misses many opportunities for intervening, whereby it mainly depends on the desired behavior, whether or not this is really the case. On the other hand, significantly increasing the duty cycle such that the relevant information is obtained much more frequently may lower the danger of missing relevant situations, but it will also increase the drain on the platform's resources. If this drain is too significant then the measure is likely to not be used in the first place, which in turn means that it

[55] We have certainly not been the first to do this. Among others, Blum et al. used this approach to determine the user's location already a decade ago [BPT06].

[56] Of course, this assumes that all users have access to wireless networks at home and at their working places. While we initially thought so, we were proven wrong by our field study (see chapter seven). We asked users to specify by way of a configuration menu, whether they were currently logged into their wireless network at home or at work. In response, we received feedback from a number of users that were asking us what to do if they had no such wireless access at one or both locations. While this should probably have been anticipated for certain types of professions, the fact that some of our study participants did not have a wireless network at home, considering that the majority of them was young and technically interested, came as a surprise to us. It goes without saying that in this case, the alleged 'workaround' did not meet the expectations.

will not be able to make any interventions at all. Finding a good compromise between information availability (that is, δ-period length) and resource consumption certainly is a non-trivial problem and the quality of a solution will be judged against the desired behavior and the measure's ideal.

Figure 6 shows the implications that the height of the duty cycle has on an interventive measure's *ISC*-function. This function aims to approximate the probability for successful interventions, as specified by the *ISD*-function, which in turn depends on the user's *BFP*-function (see chapter three). The *ISC*-function is always a step function, whereby the length of the individual steps corresponds to the length of the δ-period. As discussed in chapter three, we will assume that intervention decisions are always made at the beginning of such a δ-period. After such a decision is made, the measure falls into a dormant state. This is the only reasonable thing to do, as it lacks updated information on the user's contextual situation. The longer that the δ-period is, however, the more likely the measure is to miss relevant situations. In the example, the *ISC*-function of the measure with the shorter δ-period and the higher duty cycle correctly indicates a *kairotic* situation at 19:30 hours, whereby the measure with the one-hour δ-period practically 'sleeps through' this situation. In the case at hand, we decided for a δ-period length of 15 minutes, a value that was in part motivated by our own considerations on how quickly relevant situations are likely to change, and in part based on the findings of others [57].

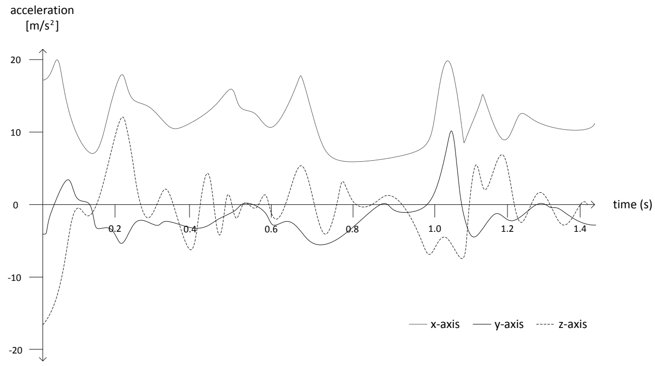


Figure 9: Accelerometer Curves.

FIGURE NOTES – The acceleration curves of a smartphone's 3-axis accelerometer during brisk walking. Figure adapted from [KOM-M-0544].

In most cases, simply obtaining data from internal sensors will not be sufficient as a basis for making decisions, whether or not to trigger the user. Although some sources of information indeed deliver knowledge that can be directly utilized without requiring any kind of further processing, especially hardware sensors will only deliver raw data that must be transformed into meaningful knowledge before it can be used. This transformation is achieved by so-called *detectors*. Detectors can be understood as being the middlemen between raw data and the actual indicator values that the decision making process as described in the next section is based upon. To provide an example, the inference of a user's activity will oftentimes be based on information that is acquired from the smartphone's inertial sensors, in particular its accelerometer. If accessed, this module delivers acceleration curves for all three axes, such as the one shown in Figure 9. In order to transform such raw data into processable indicator values, two types of methods can be employed: Rule-based systems and machine learning algorithms [58]. Rule-based systems are the much more simple approach and rely on the definition of fixed rulesets for the categorization of the obtained raw data. For example, in the case

[57] In their 2010-paper [CNC10], Conti et al. introduce their context information monitoring application *CRePE* (short for Context-Related Policy Enforcing). In regard to the polling of GPS information, they state the following: "*The experiments were conducted without any context active, just to measure the overhead of the GPS polling. We considered the energy consumption observed in 5 consecutive hours, starting with a fully charged battery. We observed that if CRePE requests the current position to the GPS device every 5 minutes, the battery level decreases by 48%, while checking the position every 15 minutes would consume 11% of the battery. The results underline how the energy consumption is one of the main issues in today's solutions for mobile devices. While these results are not negligible, the energy consumption for checking every 15 minutes is quite promising*". Although we finally decided against the utilization of the GPS module for our own approach, we nevertheless hold onto the 15 minute δ-period, as it appeared to be a reasonable compromise. However, only obtaining sensor data in 15 minute intervals and sending the application into a dormant state in between these situations actually led to unforeseen problems on Android devices, as explained in chapter seven.

[58] A very recent and promising trend of machine learning, deep learning, is not considered here, although thoroughly discussed by [KOM-M-0544].

of the accelerometer curves, the oscillation of the y-axis-curve could be used to differentiate between different types of movement such as standing, walking, and running, as it represents the smartphone's vertical movement. Simply by defining intervals for the y-axis values, straightforward recognitions can be achieved [(59)]. The main advantage of such rule-based approaches is their low complexity: For simple problems, they quickly deliver results without requiring much processing power. As such, it makes sense to rely on rule-based systems whenever the complexity of the problem is low. However, their 'straightforwardness' is also the weak point of rule-based systems. Domain experts are required to define the rules, especially for more complex problems; but if an expert does not consider all eventualities, then the system will not work as intended. Furthermore, rule-based systems can never adapt to a changing situations and this inflexibility renders them improper for usage in many cases.

The alternative to rule-based systems are machine learning algorithms. Instead of relying on predefined rules, they try to find patterns in the raw data and aim to infer a model that can later automatically recognize these patterns. The problem with machine learning approaches is twofold. On the one hand, they usually require significant more computing resources for their execution than rule-based systems. In addition, they require training samples that they can learn from and even significant numbers of such samples do not guarantee absolute classification accuracy. Table 6 compares the accuracy of three different machine learning approaches for the categorization of three different types of indicators, whereby the machine learning algorithms are all based on the popular *Weka* toolkit [(60)].

The Android-based context detection framework *ContextRec* was developed by a master's student supervised by the author of this thesis [KOM-M-0544]. *ContextRec* provides for the recognition of several types of indicators, among them activity recognition, conversation detection, and smartphone position, and it achieves an overall accuracy of 88.3%. This framework was used as a basis for the detection of various indicators that the interventive measure discussed here is based on. Table 7 specifies those indicators. Why exactly these indicators were selected is discussed in the next section.

Table 6: Machine Learning Comparison.

Indicator	**Machine Learning Approach**	**Number of Training Samples & Accuracy**					
		5	**10**	**20**	**50**	**100**	**200**
Phone Position	Random Forest	0.57	0.73	0.82	0.91	0.92	0.93
	Support Vector Machine	0.73	0.82	0.89	0.92	0.93	0.93
	Logistic Regression	0.84	0.75	0.86	0.92	0.92	0.93
Indoor vs. Outdoor	Random Forest	0.71	0.74	0.80	0.88	0.90	0.91
	Support Vector Machine	0.76	0.48	0.81	0.87	0.87	0.90
	Logistic Regression	0.87	0.62	0.84	0.87	0.87	0.92
Activity Recognition	Random Forest	0.41	0.40	0.43	0.60	0.65	0.73
	Support Vector Machine	0.35	0.37	0.48	0.57	0.60	0.69
	Logistic Regression	0.20	0.31	0.47	0.57	0.62	0.69

A comparison of the accuracy of different machine learning algorithms with varying training set sizes. The data was assessed during the development of the *ContextRec* framework and all machine learning algorithms are based on the *Weka* suite.

59 Such as in "if ((max(y) – min(y)) >= 2G) then activity = running".
60 The *Weka* software suite, which has been under development for many years by the Machine Learning Group at the University of Waikato, New Zealand, is the most advanced and widespread collection of machine learning algorithms. It is available at no charge at http://www.cs.waikato.ac.nz/~ml/weka/

Table 7: Employed Indicators.

Indicator Name		Raw Data	Detector	Indicator Values		
				Identifier	**Range**	**Crit.**
1	User Location	system state SSID of WiFi	rule-based	home	specific ID	
				work	specific ID	
				other	neither of above	x
2	User Activity	*ContextRec* smartphone sensor accelerometer	machine learning SVM *Weka*	at rest	-	
				moving	-	
				in vehicle	-	x
3	Battery Level	system state	rule-based	high	> 70%	
				medium	30% - 70%	
				low	< 30%	x
4	Ambient Noise	*ContextRec* smartphone sensor microphone	rule-based	high	> 80 dB	x
				medium	40 dB - 80 dB	
				low	< 40 dB	
5	Local Weather	*ContextRec* Web service *openweathermap.org*	rule-based	nice	CLEAR, ATMOSPHERE	
				average	CLOUDLY, DRIZZLE	
				bad	all other	x
6	Outdoor Temperature	*ContextRec* Web service *openweathermap.org*	rule-based	warm	> 25 °C	
				moderate	15 °C - 25 °C	
				cool	< 15 °C	x
7	Steps Counter	*ContextRec* smartphone sensor Android step counter	rule-based	high	> 6.700	x
				medium	3.300 - 6.700	
				low	< 3.300	
8	Notifications Priority Mode	system state	rule-based	normal	all notifications allowed	
				priority	only priority notifications	
				off	no notifications allowed	x

The table lists the indicators that were used as a basis for the interventive measure discussed in this thesis. The right side of the table specifies the indicator values that the raw data as obtained from the smartphone's system state, internal sensors, or Web Services was transferred to by suited detectors. The information was categorized in a way such that each indicator can assume one of three values, plus a fourth undefined state for indicators that cannot be obtained, such as the local weather or the outside temperature when the user has no Internet access. The rightmost column marks those indicator values that were considered to be 'critical', as they denote situations in which a user may not be sufficiently motivated or able for performing the desired behavior of brisk walking. These critical values are of special importance to the measure's decision making procedure, as discussed in section 4.4.

4.4 Decision Making

As explained in chapter three, two general types of interventive measures can be distinguished: Stubborn measures and discerning measures. The main difference between the two lies in the act of decision making. While a discerning measure distinguishes between *kairotic* situations for undertaking interventions attempts and those moments that are not suited for doing so, stubborn measures will never hold back a trigger. They will rather release it in fixed time-intervals, at the occurrence of predefined events, or when being signaled to do so, for instance by the user [(61)].

Of the three named types, the interval-based measures are by far the most simple and straightforward ones to create and as such, also the most common. Contrary to them, both the signal-based and the event-based variants will require information coming from sources outside of the measure and in the case of technology-based measures, the acquisition of such knowledge may be technical challenging and costly. The three different types of stubborn interventions are best explained on the example of a traffic light. The most simple of such appliances switch from green to red and back again in fixed time intervals. This is easier to realize than to additionally supply the device with a manual switch that allows pedestrians to signal that they want to cross the street. The most complex of three variants, however, are event-based traffic lights that switch when they detect the presence of vehicles. This variant requires the installation of inductive loops into the street, which is a costly and time-consuming procedure. A similar increase in cost and complexity when one compares interval-based, signal-based and event-based measures is encountered in the case of interventive measures that are based on mobile devices, such as smartphones. All mobile devices – and all computing devices in general – feature internal clocks that can initiate triggers in fixed intervals. The detection of events external to the device, however, such as the user's arrival at a certain location, requires the utilization of additional modules and sensors, whose powering and processing will consume battery and computing resources. As especially battery life is precious and the primary limitation of mobile devices [SW05], event-based interventions are much more complex and costly to realize than their interval-based counterparts. As such, at the time of this writing, the state of the art still mainly relied on interval-based measures; a good example is the *Apple Watch* and its *Activity App*, as discussed in chapter two.

In the focus of this work, however, are discerning measures that estimate, whether or not a given situation is suited for an intervention attempt prior to actually making one. Such a measure's decision is based on its understanding of the user's current contextual situation; more specifically it is based on its *insight* of the allegedly behavior determining parameters, at least those that it knows of. These parameters have been selected by the designers of the measure because they were deemed to be relevant. Ideally, the designers were correct and all of the selected parameters indeed have a high impact on the user's decision. If in addition no other parameters exist that have a similar high impact on the user and if all this information can be obtained in a cost-efficient manner, then the path for a highly accurate and successful interventive measure is paved. For most desired behaviors and technological platforms, this is a highly unlikely scenario. Rather, the parameters made available to the measure by its designers will consist of a good mix of relevant, somewhat relevant, and possibly even entirely irrelevant ones. Furthermore, some of the relevant indicators will be missing in the list, and resource limitations will not allow for a constant monitoring of the user's contextual state. This leads to the necessity of somehow compensating this *lack of insight* in order to nevertheless enable the making of reasonable decisions.

In principle, we find that for a discerning measure's decision making mechanic, we have the same two categories to choose from as in the case of the construction of detectors. The first option is to base the measure's decisions on fixed rules that specify certain combinations of indicator values that should lead to interventions. This approach will work well for low-complexity problems in which the user's

[61] See chapter two and [WR91] for a more detailed discussion on these three categories.

decision is merely dependent on a very limited amount of parameters [(62)]. The advantage of rule-based systems is that they will not suffer from a *lack of experience* and as such are not faced with the cold start problem (see below). Because their decisions are predetermined, these measures do not need to adapt over time – that they usually *cannot* do this is also their weak point. For complex problems that involve a large number of indicators, the ability to learn through studying the user's reactions to interventions is essential for successful measures. The more indicators that come into play, the more likely it is that target users will differ in their individual preferences. If the occurrence of a behavior depends on a larger number of parameters, then a corresponding interventive measure that is supposed to bring forth this behavior will usually also rely on a significant number of indicators. Such a system should never be rule-based, as it must be assumed that different target users will have different preferences [(63)]. Instead, the decision making procedure of these complex interventive measures should utilize learning algorithms that provide it with the ability to learn from experience and adapt.

Considering all conditions that such a system must satisfy – see challenges C1, C2, and C5 to C8 –, two principle approaches come into question: Reinforcement learning and supervised learning (more specifically: active learning). We will discuss these alternatives in the next section. But regardless of the mechanism employed, it must be capable of stating its confidence for how likely a trigger is to succeed in a situation, rather than to simply differentiate between two classes of situations [(64)]. Only the specification of such a confidence value allows for the definition of different intervention strategies, which has to do with the problem of high risk that was discussed before. The assumption that a user is only willing to tolerate a limited number of intervention attempts may result in a measure choosing to pass on the more doubtful opportunities for triggers in order to not upset the user. Analogously, a user that is known to be more tolerant can be triggered more often, which allows for a more relaxed intervention strategy. For discerning machine learning-based measures, this degree of 'cautiousness' is resembled by the measure's *confidence gate* π – but the specification of this parameter is only possible, if the measure is capable of stating its confidence for a successful intervention in the situation at hand.

As discussed in chapter three, the question of whether or not a user shows a desired behavior in consequence to an intervention attempt depends on the height of the *BFP* function value. The *BFP* function of a person always assumes the product of her (quantified) motivation and (quantified) ability for the desired behavior. Only if this value is sufficiently high and exceeds the activation threshold θ, then the desired behavior will occur. Both the value of the *BFP* function and the height of the threshold θ are not known to an interventive measure, just as it does not know the value of the *ICD* function that states, whether or not the *BFP* function value exceeds θ. A technology-based discerning measure will try to approximate the *ICD* function with its *ICS* function, usually by learning from observations. But since it must be assumed that only a subset of the relevant parameters can be assessed by the measure and that it thus suffers from the problem of partial observability as explained by challenge C3, it can never be certain that triggering the user in the situation at hand will really lead to the desired behavior. Rather, based on experience, it should only state its *confidence* that this will happen. This confidence value resembles the value of the measure's *ISC* function and triggering decisions depend on the question, whether or not it is higher than another value, namely the confidence gate π.

---

[62] As an example of a rule-based interventive measure that would perform well, think of a mobile system that is supposed to remind the user to buy milk when she is at the supermarket. Such a system can be based on just two indicators, each with two values: The user's location with the two values *supermarket* and *else*, and the amount of milk that is still in the fridge, with the values *sufficient* and *lacking*. Only in the case of the situation (*supermarket, lacking*), the measure would trigger a reminder to buy milk. Such a system could be realized with a single rule and it would still perfectly fulfill its purpose. Of course, this hypothetic measure would be based on a mechanism that is capable of somehow assessing the amount of milk that is still left in the fridge, which, irrespective of this idea having been part of ambient assistant living scenarios for many years [GHL05], is still a piece of home automation science fiction.

[63] In the case of the interventive measure for the encouragement of brisk walking that is discussed here, one such example for differing preferences may be related to the 'local weather' indicator, see Table 7. While one user may consider a light drizzle to be bad weather and thus not suited for outdoor activity, another user may find this kind of weather refreshing and will thus be inclined to accept intervention attempts made in such situations. Rule-based systems are not capable of representing such differences in individual preferences.

[64] '*To trigger, or not to trigger*', if you will.

The definition of such a confidence gate fundamentally determines the behavior of discerning measures. For a measure that is supposed to be accurate, the confidence gate needs to be close to its maximum value of 1.0. This means that the measure will only reach out to the user if it is absolutely certain that an intervention attempt will be successful. In contrast, the confidence gate of an effective measure may be set to a much lower value and a corresponding measure will then trigger the user even if it believes in *parachronotic* situations. Defining the confidence gate for a reliable measure is a problem of its own and best considered in dependence of the intended target users. If they are willing to tolerate ill-timed intervention attempts, then the confidence gate can be much more relaxed than otherwise. It may also help to establish an upper limit for the number of intervention attempts that a measure is allowed to make during the *O*-period, as discussed in the context of assumption A4. However, such a limit leads to another problem, the problem of 'limited predictability'. A measure with a limited number of attempts at its disposal will want to make use of these attempts as reasonably as possible and only trigger the user in those situations during the *O*-period that have the highest *ISC* function values. In other words, the measure is interested in determining the global maxima of the *ISC* function. However, since people's lives evolve dynamically, these maxima are not known at the beginning of an *O*-period, but can only be determined in hindsight. In may thus be advisable to analyze a measure's behavior after some time and increase or decrease the confidence value in order to adjust its behavior. Defining the confidence value for reliable measures is thus mostly a try-and-error process. For the measure at hand, we have decided for an initial confidence value of $\pi = 0.75$; by analyzing the results of the field study, it became apparent that this value may have been too relaxed.

Table 7 lists the indicators that we selected as a basis for our interventive measure meant for bringing forth the activity of brisk walking. All indicators were selected because they were believed to provide relevant information on parameters that influence a user's decision, but not all of them are low-hanging fruits: Rather, user activity, ambient noise, local weather, and the outdoor temperature are moderately costly to obtain, as they rely on internal smartphone sensors or external information sources, such as a Web service [(65)]. As pointed out earlier, we were at least able to get rid of the most power-hungry detector, the readout of the GPS module, by replacing it with the assessment of the SSID. The reader may have noticed that none of the indicators is related to the current time or date. This is easily explained: If used in conjunction with other indicators, then time and date usually only provide redundant knowledge. While one can assume that physical activity is easier for a person on a weekend than on a working day, this easiness is not due to the day itself, but rather caused by the fact that the person is at home and has free time at her disposal. Thus, although we heavily relied on time and date for earlier prototypes as can be seen from Table 8, we decided to not make us of them here.

Even if a large number of indicators are available to a measure, it may be reasonable not to employ all of them. This can either be motivated by limited system resources, or because there is reason to believe that not all parameters are as relevant as initially thought. In such cases, it may make sense to willfully reduce the level of insight that a measure has in order to be able to separate helpful from irrelevant knowledge. Such a differentiation between low-, medium-, and high-insight measures may lead to interesting findings. For the problem at hand, we created three variants: A low-insight trigger that only relies on the user's activity and location, a medium-insight trigger that also takes the device's battery level and the ambient noise into account, and finally a high insight trigger that utilizes all indicators specified in Table 7. As all learners need a knowledge base to start (the 'cold start' problem, see challenge C5), we identified critical indicator values for which we assumed that a user will not want to be triggered in the corresponding situations. The *ISC* values of these situations were manually set to a low value, such that the occurrence of triggers in such situations was made unlikely. Based on this knowledge, the measure was then able to adapt to the individual user by way of learning.

[65] The 'step counter' indicator relies on the pedometer built into many Android-based devices. This hardware sensor automatically calculates the user's steps and allows applications to directly access this value without having to analyze the smartphone's accelerometer readings. This is a much more cost-efficient way of obtaining the total number of steps and can be considered 'low cost'.

4.5 Learning Strategy

Obviously, stubborn measures do not learn from their attempts, and neither do rule-based discerning measures. While this means that they do not suffer from the cold start problem, they will also not be able to adapt to the individual user, which is problematic for two reasons. First, it must be assumed that the preferences of users are different and what may be a *kairotic* situation to one person will be a terrible moment for intervening to another. For many types of behaviors, intervention strategies will not be generalizable and must rather be tailored to the individual user. In addition, the attitude of a user may change over time. As such, it is mandatory that measures never stop learning from the user's reactions so that they can handle this behavior drift. Because of these reasons, learning discerning measures are generally preferable to their rule-based counterparts, unless the activity that the measure is supposed to bring forth only depends on a low number of known and observable parameters.

In principle, two different approaches are suited for solving the problem of constructing a learning discerning measure. The first of these two options consists in reinforcement learning [RN16]. Here, a software agent tries to maximize its internally managed score by finding those actions that yield the highest reward, whereby the agent's actions consist in the release of triggers in different types of situations and its reward is the verification that a trigger was successful. The challenge in constructing a good triggering agent lies in finding a reasonable strategy for the *exploration-exploitation-tradeoff*, the balance between triggering the user in promising situations and trying out new and unknown situations. While discerning measures can in principle be implemented as software agents, this is not the best of options for two reasons. First, a measure will usually not need to learn triggering-chains that allow it to concatenate a high number of successful interventions. This, however, is what the agent is trying to do [(66)]. In addition, the high risk problem limits the agent's ability to explore different paths, which reduces the quality of such measures. As such, the alternative to agent-based systems seems to be better suited for solving the problem at hand: The employment of 'classical' supervised learners such as support vector machines, decision trees, or k-nearest neighbor algorithms.

The problem with supervised learners is that they require a large set of pre-classified samples before they are able to make reasonable decisions. As can be seen from Table 6, their classification accuracy significantly increases with the amount of labeled training samples that they can build upon. As most measures must adapt to the individual user, only a small number of indicator tuples will be correctly classified from start. In the case of our specific measure, these may just be the indicators tuples that contain one or multiple of the indicator values that have been labeled as 'critical' in Table 7. The combination of the cold start problem and the high risk problem that limits the number of intervention attempts may result in a measure making many ill-timed intervention attempts during its initial period of application. Even worse, the measure's triggering accuracy will only improve slowly. An active learning strategy may help here. Depending on a selection strategy, active learners pick those samples from the set of unlabeled data that they can learn from the most and present them to the user with a request for labeling [Set10]. Active learners are often used in conjunction with support vector machines (SVMs), such as by Tong and Koller [TK98], Tong and Chang [TC01], and more recently by Kremer et al. [KSI14]. Such SVMs represent samples as feature-vectors in an n-dimensional feature-space. During its training phase, a SVM tries to find a hyperplane that separates the samples of different classes, whereby it aims to find the one hyperplane that maximizes the distance to all vectors. An advantage of SVMs is that once they have been trained, the classification of new samples is a fast and straightforward process, as the SVM 'simply' needs to assess, on which side of the hyperplane the sample is located on. For the interventive measure discussed here, we have decided to rely on a SVM with active learning (also see chapter five). However, in previous approaches we have also experimented with other types of supervised learners (and rule-based systems), see Table 8.

[66] It should be noted that such chains may indeed be of interest to measures that aim to be accurate or effective. However, we will not investigate this approach any further as we focus on reliable measures that are content with a single successful intervention.

A core problem of learning is obtaining feedback that allows for the correct labeling of samples. In the case of interventive measures, this means that only if the desired behavior really occurs in consequence to an intervention attempt, then the confidence value for the situation at hand should be increased. As discussed earlier, obtaining such information through monitoring the user is not always possible. For instance, it is not possible to reliably recognize with a smartphone only that someone has drunk a glass. For this reason, applications that try to bring forth such a behavior, such as *Plant Nanny* (see chapter two), rely entirely on user input. Admittedly, this is not ideal, as the user may provide wrong feedback, but for many types of desired behaviors, no alternatives exist. If the behavior in question can be observed with technical means, however, then an obvious precondition is that the interventive measure must be provided with the ability for doing so, but such monitoring tasks are two-sided swords. While they allow for the verification that the user has really shown the desired behavior, they will also require system resources and as such, they will increase the resource drain on the user's device. It must thus be decided from case to case, whether the verification of a successful intervention is worth the effort [67]. Table 8 states the decision making procedures of three prototypes for interventive measures that have been developed by students supervised by the author of this thesis.

Table 8: Previous Prototypes.

Prototype		**Decision Making**	**Employed Learner**	**Indicator Values**		
				#	**Name**	**Values**
1	[KOM-B-0517]	rule-based	-	1	day of week	Monday, Tuesday, …, Sunday
				2	time of day	30 minute intervals (48 in total)
				3	location (GPS)	home, work, other_1, other_2, else
				4	activity (acc.)	resting, moving, fast, vehicle
				5	local weather (Web)	73 different conditions
				6	outside temp. (Web)	very_cold, cold, average, warm, hot
2	[KOM-M-0535]	machine learning	supervised (decision tree, naïve Bayes, k-NN)	1	day of week	Monday, Tuesday, …, Sunday
				2	time of day	daytime, evening, nighttime
				3	location (GPS)	location_1, location_2
				4	local weather (Web)	nice, average, bad
				5	outside temp. (Web)	cold, average, warm, hot
3	[KOM-M-0543]	machine learning	supervised (decision tree)	1	day of week	Monday, Tuesday, …, Sunday
				2	time of day	8 groups
				3	location (SSID)	up to 10 different locations
				4	activity (acc.)	lying, sitting, walking, running, other
				5	steps counter	< 1.000, 1.000 – 2.000, … , > 10.000
				6	hours since last trig.	1, …, 23

The desired behavior of all three prototypes was medium-intensity physical activity such as brisk walking; none of the prototypes verified intervention success through user monitoring.

[67] There is indeed another benefit that comes from observing the user's response to a trigger: It may allow for the distinction between marginally successful interventions and highly successful ones. If a user is triggered for brisk walking, for example, but only takes a brief five minute walk around the block, then this is another category of 'success' as when she spends an hour outside. This kind of information may be used to optimize a measure's decision making procedure even further.

5. Consolidated Approach

In principle, triggering mechanisms such as the one described in chapter four can be used independently of all other means. Their task is to point out opportunities that are suited for showing a desired behavior, and in that they are entirely self-sufficient. However, the FBM explains that the highlighting of opportunities – the act of triggering – is just one of three aspects that decide whether or not a target person will behave in a desired way. Also raising this person's motivation and/or ability for the behavior in question may make the difference between successful and unsuccessful intervention attempts. In this regard, this chapter investigates the 'embedding application' in which a triggering mechanism based on the concepts as described in the previous chapter was integrated into. Thereby, the application named *Twostone-IM* – which was selected because of its supposed ability for raising peoples' motivation for being physically active – actually consists of two apps to each of which a section of the chapter is dedicated to. While this fifth chapter is mainly focused on the description of *Twostone-IM* on a general level with the goal of providing a good overview alongside pointing out some of the more interesting implementation details, the following sixth chapter explains the elements of *Twostone-IM* that are specifically meant for raising the user's motivation for brisk walking (without overly decreasing her ability for this activity).

5.1 PacStudent and Twostone

One of the earliest location-based games, designed and played at a time when modern-day smartphone technology was still in its infancy, was New York University's *Pac-Manhattan* [(68)]. The game, that enjoyed an astonishing amount of media coverage during the spring of 2004, was strongly based on Namco's arcade classic *Pac-Man*, with the main difference being that in the case of *Pac-Manhatten*, both the protagonist and her opponents were actual people who were running through the streets of Manhatten, New York. Directed via mobile phones by a group of game masters, the ghost players were trying to catch the *Pac-Man* player before she had collected all virtual dots that were spread across the streets of Manhatten – but that were only visible on a computer screen back at the university's lab and thus, only visible to the supervisors [Joh04]. Due to the game's multiplayer aspect and its heavy dependence on all-knowing game masters, it actually resembled a real life version of *Scotland Yard* just as much as the original *Pac-Man* and from today's perspective, the game appears as if it was more of an art project than a scientific prototype. But, nevertheless, there was much fascination for *Pac-Manhatten,* which can be explained by the at that time innovative concept of pervasively fusing digital game mechanics with a real world setting. As pointed out in the second chapter, such *pervasive games* quickly became popular in the first decade of the 21st century and besides trying out many other concepts [(69)], some research groups also kept experimenting with the idea of creating a 'real life version' of *Pac-Man*, such as the University of Singapore with *Human Pacman* [CFG+04], and the University of Lancaster with *PAC-LAN* [RBC+06].

In the lecture series *Urban Health Games* that took place at the Technische Universität Darmstadt between the years of 2013 and 2016, and in whose organization and administration the author of this thesis was involved, students from several disciplines were brought together with the intention of amplifying their creative potential for the conceptualization of pervasive mobile games, specifically for mobile exergames played in an urban context – hence the lecture's name. During the course of three years, a multitude of game prototypes was created by the participating students under the supervision of this thesis' author and others [(70)]. One of these prototypes was the mobile location-based exergame

[68] In late 2016, the game's Web site was still online at http://www.pacmanhattan.com/

[69] See [Opp09] for a good overview of early pervasive game prototypes.

[70] For more information on the *Urban Health Games* lecture series and on some of the game prototypes that it brought forth, please refer to [KDH+13, KKN+14, KDH+14].

PacStudent, which, as its name implies, was yet another interpretation of the '*Pac-Man* in real life'-idea, much like the game prototypes listed above. Different to them, however, *PacStudent* was fully playable by a single player on a smartphone and did not require any additional devices or people. As soon as the player was in the vicinity of a location for which a virtual game map had been created, she was able to start the game, whose interface mainly consisted of a 2D-map of the respective level. The player was then supposed to move along the level's virtual lanes and to collect virtual dots while avoiding virtual, computer controlled opponents, very similar to the original *Pac-Man* (with the notable twist that the player had to move in the real world in order to move her virtual avatar, of course). Our paper [DHB+14] contains some screenshots of *PacStudent*, and also points out two of the game's problems. The first of them consisted in *PacStudent*'s limitation to pre-defined maps; a classical issue of location-based gaming in general [(71)]. *PacStudent* was only playable at a small number of locations distributed across the city of Darmstadt, Germany, among them the public park 'Prinz-Georgs-Garten' whose distinctive grid-like layout allowed for an easy mapping of virtual game lanes to real world walking paths, much simplifying a player's orientation in the real and the virtual world. This was made necessary by the second problem of this early prototype, which was related to the game's user interface. The 2D-map view of the game, displayed on the comparably small screen of a handheld smartphone, was not suited for a moving player. The inevitable shaking of the smartphone's screen while running forced players into regular stops in order to allow them to re-orientate themselves in the virtual game world. However, while players were standing still and looking at the map as displayed on the smartphone's screen, their virtual opponents kept moving. This occasionally led to players getting caught by opponents and losing the game – a frustrating experience.

The intention of finding solutions for these problems sprung two bachelor theses that were supervised by the author of this work [KOM-B-0488, KOM-B-0489]. The integration of their results into the original *PacStudent* prototype led to a new application named *Twostone*. In the following two years, *Twostone* was constantly refined and served as a fundament for a multitude of lab courses and student theses under the author's supervision [KOM-B-0517, KOM-B-0518, KOM-B-0519, KOM-M-0535, KOM-B-0549, KOM-M-0572]. Eventually, a stable release version of the game was uploaded into the Google Play Store and thus made publically available [(72)]. While the work of the author has brought forth a variety of other mobile applications, most notably a handful of mobile exergames, *Twostone* was selected as an integrative solution because it was the most advanced of these prototypes. Indeed, due to the significant amount of student work that went into the game over the course of almost three years, the game can probably be considered to be significantly more advanced than many other scientific prototypes used in lab and field studies. Knowing this is relevant for fully understanding the implications of the evaluation results that will be presented at a later time. However, choosing *Twostone* as an integrative solution was also not without problems. The most obvious issue is that by doing so, the desired behavior that triggers are supposed to initiate was changed from 'brisk walking' to 'playing *Twostone*'. The consequences of this are analyzed in chapters six and seven.

Please see appendix B for screenshots that provide an impression of what *Twostone* looked like at the time of this writing. It is also worth pointing out that both the game and the triggering application described in the next section were only available for Android-based devices. The reason is easily explained: *Twostone* started as a student's project and originally was not meant to be used by a significant number of users. Its evolving into a comparably stable, feature-complete application was a slow process. It goes without saying that in contrast, real life interventive measures should not be limited to a single operating system and instead be made available to the broadest group of users possible. Indeed, *Twostone*'s limitation to a single operating system turned out to be highly problematic when we tried to find participants for our field study, as will be explained in detail in chapter seven.

[71] Typically, location-based games can only be meaningfully played at the locations for which game designers have created content *a priori* – unless such game content is randomly created in the player's vicinity, which is a feasible but usually not an optimal solution. For more on this topic, see chapter six.
[72] The game can be downloaded free of charge from https://play.google.com/store/apps/details?id=de.tu.darmstadt.uhg

5.2 Twostone Interventive Measure

The theoretical concept for a smartphone-based interventive measure as described in chapter four was implemented as a stand-alone Android application by a student as part of his master thesis [KOM-M-0572]. Although this 'triggering application' was intended to be an interventive measure for *Twostone* from start, we decided against a full integration of the two applications. The reason for this is that the triggering application relied on the *ContextRec* framework [KOM-M-0544] for the assessment of several of the indicators that are listed in Table 7. This framework, however, makes use of features of the Android operating system that are only available from Android version 5.0 upwards. As pointed out in the previous section, *Twostone* is publically available from the main online distribution platform for Android applications, the Google Play Store, and compatible with Android version 4.0. As a full integration of the triggering application into *Twostone* would have also increased the operating system requirements of the latter, we decided against this step and instead asked the evaluation participants to install a second application from the Google Play Store (73). Together, *Twostone* and the triggering application made up the *Twostone Interventive Measure* (the *Twostone-IM*) that was evaluated during the field study (74).

As described in chapter four, the triggering application relied on a support vector machine based on the *Weka* suite for its decision making procedure. In addition, it also employed an active learner that, whenever the triggering algorithm decided against an intervention attempt, analyzed whether the situation at hand was considered to be interesting enough to ask the user for feedback nevertheless. We found the entire procedure to be sufficiently performant such that it could entirely be run on a smartphone. Nevertheless, the triggering application uploaded its locally stored database to the *Twostone* server once a day and used this opportunity to check for updated settings, most significantly for a change of phase. In order to be able to compare different variants of the triggering application (stubborn, low-insight, medium-insight, and high-insight, see chapter seven), we made the triggering application's behavior dependent on a database entry. Changing this entry enabled us to make the trigger behave as any of the four types of interventive measures that were supposed to be evaluated during the field study. In addition, this mechanism also allowed us to simultaneously enable and disable interventions for all study participants, which ensured comparable study conditions.

When the triggering application decided for an intervention attempt, it presented a trigger to the user, which was essentially a push-up notification accompanied by a vibration of the device and a beep-sound. Screenshots of such a triggering message can be found in appendix B. Furthermore, as we used active learning in order to improve the accuracy of the discerning measures more quickly, users whose triggering application were set to this mode were also occasionally prompted with learning notifications. While these practically looked like regular triggering messages, the confirmation of such a learning notification did not automatically start *Twostone* with the nearest map, but rather only served as a feedback to the learner that a trigger would have been successful in the given circumstances. The fact that triggering messages and learning notifications were not clearly distinguishable from one another turned out to be problem, as discussed in chapter seven. Another shortcoming of the triggering application was that it did not assess, whether or not the desired behavior had really been performed. Rather, if the user confirmed the trigger, then this was registered as a successful intervention. Analogously, we did not ensure that users did not play *Twostone* unless triggered for doing so, which may have led to some users declining triggers only to play the game a few minutes later nevertheless.

[73] Although not meant for the general public, the triggering application was also uploaded to the Google Play Store. Past experiences had shown that directly sending APKs to study participants oftentimes led to several critical questions in regard to the installation process and data security. Distributing software instead directly via the Google Play Store was found to be a much more uncomplicated and better accepted procedure.
[74] The number of potential study participants that did not have a smartphone with a sufficiently high version of the Android operating system actually made it necessary to quickly find a workaround for the operating system demands of the *ContextRec* detection framework. This is discussed in more detail in chapter seven.

6. Motivation and Ability Increase

While the fifth chapter describes the application *Twostone-IM* from a chronological perspective and as a whole, this chapter is dedicated to an analysis of those features of the application that were specifically meant for increasing a person's motivation and, albeit to a lesser extent, her ability for the desired behavior of brisk walking. Thereby, the first section of this chapter focuses on the motivational elements of *Twostone-IM*, and the second section describes the aspects that were implemented with the goal of somehow adapting to the user's ability (mainly in the sense of simplifying the act of playing). The chapter's final section is an addendum and describes a theoretical concept that is not actually part of the application as it is. The notion of *Sliced Serious Games* originated from a brain storming session with my doctoral advisor Ralf Steinmetz and my superior Stefan Göbel and it is mentioned here because it provides an outlook on where pervasive games in general – and pervasive behavior changing games in particular – may evolve to in the near future.

6.1 Gamification and Game Mechanics

The most straightforward way of creating extrinsic motivation for an activity – besides offering material rewards, most notably monetary compensation – is to quantify one's performance. A logical second step then consists in allowing people to compare their own performances to that of others. And the third escalation of this approach is to abstract from numerical values and to instead award achievements for reaching certain goals, thus replacing sterile numbers with more tangible titles, such as 'Most Valuable Player' or 'Salesperson of the Year'. In a nutshell, this is what gamification is about. And although, as pointed out in the second chapter, the topic has been the subject of much discussion in recent years, this simple formula of 'points, badges, and leaderboards' is known to work, at least in certain fields of application. It has long been known that people will improve their efforts when being supervised (75), and apparently, this effect will also set in if the supervising entity is merely a computer program.

Twostone-IM features two of the three listed gamification mechanics, namely points and a leaderboard. For one, players are awarded a score for each game session that they complete, based on the session's difficulty setting (see next section) and the relative amount of the level that the player was able to complete before she was captured by one of her computer controlled opponents. The game also keeps track of the player's average speed, of the distance that she has covered, and of the time that she has played, and total counts of these values that are being summed up over all game sessions can be reviewed from the player's profile screen. Furthermore, an additional high score screen allows players to compare the total distances in meters that they have covered over the course of all their game sessions to that of other players (76). See appendix B for screenshots of the game's user profile screen, of the leaderboard screen, and of the application in general.

A quick look at *Twostone-IM* reveals that it is not only an activity tracker with gamification mechanics. Indeed, as pointed out in the previous chapter, the application's gamification elements are just an extension to its actual core: At heart, *Twostone-IM* is an exergame, a video game that requires physical activity from its players in order for them to make progress (to 'win' the game, if you will). *Twostone-IM* and its predecessor *PacStudent* are both based on the arcade game classic *Pac-Man* in that players are meant to gather virtual items while evading computer controlled opponents. If a player manages to collect all items, she wins. If she is caught by a ghost while trying to do so, she loses.

[75] The phenomenon that a person will intensify her efforts when aware that she is being watched is named *Hawthorne effect*, derived from the name of the factory in the vicinity of Chicago where it was first discovered [Han67].

[76] There is no compelling reason why players are only allowed to compare their total distances. Since various other values, such as the total score and the total game time, are already being assessed, they could also easily be made available for comparison. An interesting question that arises in this context is, which of such values has the most significant motivational effect on players in average (if there is any at all), and whether multiple leaderboards that allow players to compare their performances based on different factors will increase or possibly even decrease the motivational effect that a single high score list has. However, as this work does not focus on how to increase user motivation for a desired behavior, such questions are out of scope here.

Twostone-IM, being an exergame, shares the principle idea of serious gaming, which is to use video games as motivators for activities "*that would otherwise not be done*" [Mal14]. While this reasoning is easily understood, the scientific perspective demands the same question to be asked here as in the case of gamification and that is whether serious games can really provide for the desired motivational effects. More specifically: Whether such games are equally suited for the motivation of all types of 'serious' behaviors and all types of players. As already pointed out in the second chapter, the creation of 'good games' that a large relative amount of players finds equally enjoyable is a difficult problem that even the professional games industry occasionally struggles with. As such, it is at least questionable whether a game such as *Twostone-IM* that was developed by a small team and in a comparably short amount of time is really suited as a motivator for physical activity.

During a first evaluation of the game, we made an interesting observation. The author of this work and a student investigated several of the application's features under lab conditions, whereby each of the N=23 study participants was asked to fill a pre-study questionnaire, to then play the game for 45 minutes, and to finally fill two post-study questionnaires. The actual purpose of the evaluation was to find out, whether the classification of players into the different player types of an established player model would allow for a reliable prediction of what game elements they would enjoy the most (77). In the following, this evaluation will be referred to as the *lab study,* in contrast to the *field study*, whose results are detailed in chapter seven. 12 of the lab study's 23 participants were male, the majority of them was in their twenties (78), and they claimed to have an average amount of video gaming experience (79). The interested reader is referred to [KOM-B-0549] for more detailed information on the study's organization and for a full list of the evaluation results.

While we were not able to confirm our initial hypothesis – the classification into player types did *not* allow for a reliable prediction of the preferred game elements – the study revealed some unexpected findings. Among many other questions, players were asked to rate their overall impression of *Twostone*, how much they enjoyed being able to compare their performance with that of others via a leaderboard, and how much they enjoyed the game's short story that we had written to give some meaning to the game's name and the role of the player. Taken for itself, none of these results was exceptionally positive or negative (80), but the correlation matrix showed a strong negative and significant correlation between a player's gaming experience and how high she scored the enjoyment that the game and its specific components had brought her (81). In other words: The more experience with video games that a player had, the less she enjoyed playing *Twostone,* and this effect held true for both the game's core mechanics and its added gamification elements. This may point towards a general problem of serious gaming. Once a person gets used to playing video games, her demands on this type of media increase. In times when companies are willing to invest half a billion US Dollars into the production of a single video game [Luc14], this becomes a huge problem for small development teams that can never hope to reach the high degrees of production values that many players have grown accustomed to (82).

---

77 The study was based on the player type model by Yee [Yee06]. All study participants were asked to fill a questionnaire meant for the classification of players into Yee's three overarching components (achievement, social, and immersion) at the beginning of the study. To this end, the original questionnaire by Yee was translated into the German language (with his permission). In addition, participants had to fill the NEO-FFI questionnaire by Costa and MacCrae [CM92] in order to determine their 'Big Five' personality traits (neuroticism, openness to experience, extraversion, conscientiousness, and agreeableness).

78 "*How old are you?",* ratio scale*,* N=23, M=23.6, SD=±5.99.

79 "*I have a lot of gaming experience*", five point disagree-agree Likert scale, N=23, M=3.12, SD=±1.37.

80 "*I enjoyed the application as a whole*", five point disagree-agree Likert scale, N=23, M=3.74, SD=±0.86. "*I enjoyed the game's leaderboard*", five point disagree-agree Likert scale, N=23, M=3.74, SD=±1.29. "*I enjoyed the game's story*", five point disagree-agree Likert scale, N=23, M=2.39, SD=±1.16.

81 Data pair 'experience-overall', correlation coefficient r=-.42, p<.01. Data pair 'experience-leaderboard', correlation coefficient r=-.51, p<.01. Data pair 'experience-story', correlation coefficient r=-.50, p<.01. Another noteworthy correlation was found in the pair 'gender-story': Women enjoyed the game's story significantly more than men, correlation coefficient r=.51, p<.01.

82 Exceptions prove the rule. The game *Minecraft* was originally developed by a single person and became an incredible financial success with a disruptive effect on the entire gaming industry. Indeed, there are a number of examples for small teams that have been able to produce well-received games and this trend sprung a new industry branch, the 'indie games'. However, such games (with the notable exception of *Minecraft*) usually only appeal to the small number of self-proclaimed 'hardcore-gamers', while the vast majority of players prefers the so-called AAA-titles from the likes of *Call of Duty* and *Grand Theft Auto*.

6.2 Pervasiveness and Difficulty Adaptation

The lab study revealed a negative correlation between a player's video gaming experience and her enjoyment of the game *Twostone*. Furthermore, the study only assessed a player's motivation for playing further game sessions after a single, 45-minute test run. It goes without saying that such a set up does not allow for any sound conclusions on whether *Twostone-IM* can create long-term motivation for the 'serious' behavior that it is supposed to bring forth, namely brisk walking or easy running. At this point, it must thus be concluded that no clear statement on whether or not *Twostone-IM* is suited as a motivator for physical activity can be made. It must rather be assumed that if any long-term motivational effects of the game exist that they are likely to be small, at least for the majority of players. Obviously, the creation of extrinsic motivators and/or the strengthening of intrinsic motivation are difficult tasks, which is far from being a novel insight. As mentioned in the third chapter, Fogg suggests that designers of persuasive technologies should focus their attention on the creation of intelligent triggering mechanics first, to then turn to raising ability for the desired behavior, and only consider increasing motivation as the last option. He points out the exact same problem that we experienced during the development of *Twostone-IM*, namely that it is difficult to pinpoint, what exactly makes a motivational mechanic: "*People designing interventions often start by focusing on motivation, believing it is the most effective way to change behavior. However, motivation is the trickiest, most nebulous area. It's harder to measure, and it's hardest to change predictably*" [Fog10, p.12].

Following this notion, the better alternative to trying to raise motivation would be to focus on increasing the user's ability instead – which spawns the question how this could be achieved when, as in the case at hand, the interventive measure is a smartphone application and the desired behavior is brisk walking. However, if we assume that *Twostone-IM* does at least somewhat increase a user's motivation for being physically active, then we would already be on the winning side if we could ensure that the application does not *reduce* the user's ability for the desired behavior, as then, the product of motivation and ability for the desired behavior would still be higher than without the game's influence. On the other hand, a reduction of the user's ability caused by the game that is so significant that it outweighs its positive motivational effects would mean that the application's employment would actually reduce the probability of the desired behavior's occurrence. Indeed, in the case of brisk walking, taking care that an application such as *Twostone-IM* does not artificially impose new restrictions is easier said than done, as we selected brisk walking specifically for being performable anywhere and at any time (see chapter three). The fact that *Twostone-IM* is a location-based game and as such requires users to be at certain locations before they can play does not help with this problem.

When choosing to use *Twostone-IM* as the embedding application for the triggering mechanism, the desired behavior was practically changed from brisk walking to playing a session of the game. While this makes almost no difference from a physiological perspective, it does make a huge difference in terms of organizational requirements. *Twostone-IM*, being a location-based game, can only be played at locations for which game levels have been created *a priori*. In the case of the game's original prototype *PacStudent*, this limited the game's usability to a handful of predefined maps spread across the city of Darmstadt, Germany. In order to play the game, a person first had to travel to one of these locations for which the game's creators had designed a game map and made it publically available. Obviously, this made 'being physically active with *PacStudent*' much more difficult than taking a mere (but brisk) walk around the block. An overall positive effect of such a game is then only brought forth in those rare cases when players are so extremely motivated for playing that they are willing to put up with any hindrances [(83)].

[83] The existence of extreme motivational effects that outweigh all imposed burdens should not be entirely excluded. For example, reports exist of players of the location-based game *Pokemon Go* who are going to extreme lengths for the achievement of certain in-game goals, such as the man who claimed to have lost 12 kilograms by walking close to 230 kilometers in less than three weeks, only for the purpose of advancing quickly in the game [Eas06].

Since the creation of a large enough number of maps by ourselves was not a feasible solution and because we deemed algorithms for the automatic generation of levels in the player's vicinity to be not sufficiently reliable, we decided for providing players with the ability of creating their own levels [84]. Simply drawing the level's lanes on a map seemed to be the most straightforward solution, but this also harbored the danger of users creating overly ambitious or unplayable levels. We thus decided for another approach and to require the player to actually walk the map that she intends to create herself: *Twostone-IM*'s level editor is based on the player's position and while the level creation mode is active, the game automatically connects the user's detected locations to a new map that can then be uploaded to the game's server and shared with other players. This allows players to create new maps in close vicinity to their homes, their working places, or close to any other area that they frequently visit. In addition, a 'quickplay'-button was added to the game that automatically starts the level closest to the player and thus shortens the delay that a player has to wait before she is in-game and (literally) ready to go. Please see [KOM-B-0489] for more information on the game's level editor, its implementation details, and the results of a small usability evaluation with eleven test users.

The level editor simplified playing the game *PacStudent* (and its successors *Twostone* and *Twostone-IM)* by increasing the game's *pervasiveness*. Pervasiveness is an essential trait for interventive measures in general and for behavior changing games in particular – a topic that we will get back to in the next section [85]. As stated before, in order to ensure that the game *Twostone-IM* has an overall positive impact on the probability of its players to be physically active, it is important to reduce its ability impeding effects to an absolute minimum. It was already pointed out that besides it being limited to a handful of game levels, the original *PacStudent* suffered from another problem: Its user interface. Many of the game's early players complained that while walking quickly – necessary for evading the game's virtual ghosts – the inevitably shaking of the screen made one's orientation within the virtual world difficult. Instead of being able to focus on moving quickly, players were rather constantly distracted by their device and forced into regular stops. In their publication '*Considerations for the design of exergames*', Sinclair et al. describe this exact problem, stating that exergames are especially difficult to design since their players move while playing. They come to the conclusion that "*for a player to enter the flow state, they must be able to focus on a narrow field of attention*" and give the game *Dance Dance Revolution* as a good example for an exergame, because its user interface is reduced to an absolute minimum, allowing the player to fully focus on her movements while playing [SHM07].

In the paper [DHB+14], I argue that a problem of the original *PacStudent* was that its interface was lying outside the 'Focus Corridor'. It was too complex for the amount of focus that the player was able to provide while moving, and thus it required the player to stop from time to time and to focus on the smartphone's screen. In order to reduce the problem, we experimented with different interface types

[84] There are three principle options for creating content for location-based games. The first of them is to have the content created manually or semi-manually by professional designers. While this approach is likely to yield the best results in terms of quality, it is also the slowest and least cost-effective and will thus usually limit the game's playability to certain small areas, such as a single city. The second option is to utilize algorithms that automatically create game content in the vicinity of the player. The quality of this content then depends on the quality of the algorithms used, but this approach will usually not lead to satisfactory results, at least not when compared to the content that can be provided by human designers. The third option is to include the community into the process, to enable players to create content for their respective locations, and to allow them to then share this custom made content with other players. Most commercially successful applications such as *Ingress* combine two or all three of these options and for example allow players to make suggestions for new content which is checked and possibly altered by professional designers before being made publically available to the entire player base.

[85] As a side note, several of the student theses supervised by the author of this work have investigated ways of how location-based games can be made more pervasive. In addition to the level editor for the game *PacStudent* as conceptualized and implemented by [KOM-B-0489], the master thesis [KOM-M-0498] looked into options for how to integrate reoccurring urban elements, most notably street name signs, into a game in order to automatically create game content focused around such 'anchors'. In this specific case, an image detection algorithm analyzed the camera feed of the user's smartphone and, if a street name sign was recognized, it created an augmented reality puzzle based on the characters of the respective street name. The player then had a limited amount of time to solve the puzzle and to thus 'conquer' the corresponding street for his team. The benefit of such an approach is that the game content created is guaranteed to be at a location accessible to the player, as well as being meaningfully linked to her surroundings. The thesis [KOM-M-0573] investigated a similar approach: It used car company logos as dynamic anchors for a location-based game for children.

and finally settled for an augmented reality based user interface that allows players to identify the relevant virtual game elements while also being able to keep their eyes on the real world ahead of them. Players can thus maintain high speed without needing to stop and look at the game's 2D-map. However, if desired, they can switch between the two interface modes at any time simply by changing the angle in which they hold their smartphones. See [KOM-B-0488] for more information.

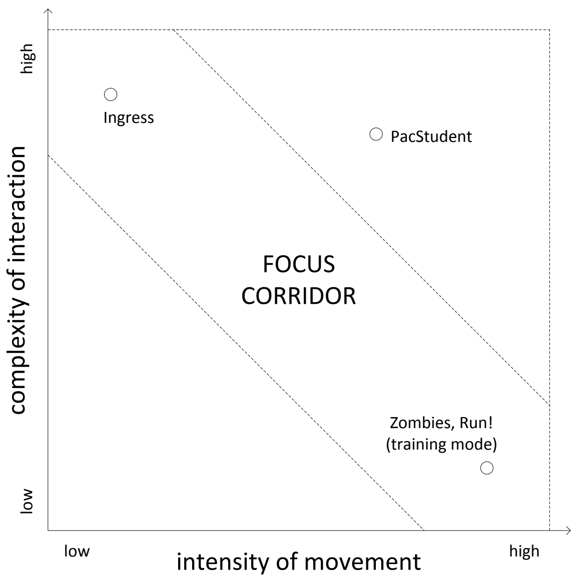


Figure 10: Focus Corridor.

FIGURE NOTES – The Focus Corridor specifies the minimum and maximum complexity that the user interface of an exergame should have without interfering with the player's experience.

Determined to increase the game's overall accessibility, we also experimented with different ways of adapting its difficulty to the player's individual capabilities. The most notable of these approaches was the linking of the player's heart rate to the speed of her computer opponents. At first, we assessed the player's average heart rate during rest and medium-intensity exercise in order to acquire an individual 'base line'. While playing a map of *Twostone*, the player was then given the option to manually indicate via the smartphone's loudspeaker buttons if she wanted the game's difficulty to increase or decrease. We compared three different mechanisms (direct adaptation based on player feedback, an adaptation based on player feedback and player heart rate, and an adaptation based on player feedback, her heart rate, and her average speed) and found that all of our six test users preferred the 'full adaptation mode' that took all three aspects into account over all other versions of the game, especially the one without any options for difficulty adaptation. This may be because the automatic adaptation essentially removed 'idle states' from the game – if the player's speed and/or heart rate dropped, the intelligence and speed of her opponents automatically increased and thus kept her on her feet [86]. While this is a positive effect, it actually contributes to the player's *motivation,* but reduces her ability for playing *Twostone* – playing the game then requires her to bring a heart rate monitor. Because we wanted to avoid making the game even more complicated to play, we decided to remove all adaptation features and to only leave the manual difficulty adaptation via loudspeaker buttons that allows players to dynamically set their opponents' speed and intelligence in five steps. See [KOM-B-0518] for details on this mechanism.

The main question that remains at this point – and a question that is left unanswered – is, whether *Twostone-IM* increases a player's motivation for being physically active to such a high degree that this balances out the game's negative impact on her ability. As we are mainly interested in the effects of different types of user triggers, selecting *Twostone* as an embedding application is a choice as good as any, as it is not the application itself that we are interested in but rather what triggering mechanics will increase the probability of its usage. However, when designing actual interventive measures that are meant to reliably increase the prevalence of physical activity, utmost care must be taken that the application in question really *increases* the probability of occurrence and not *decreases* it by negatively affecting the user's motivation and/or ability for the desired behavior.

[86] The professional games industry refers to the automatic adaptation of a game's difficulty to a player's individual skills as 'rubber banding', and this is the subject of much discussion among game designers, games journalists, and gamers.

6.3 Perspective: Sliced Serious Games

During a discussion with my doctoral advisor Ralf Steinmetz and my superior Stefan Göbel on the topic of pervasive games in general and pervasive behavior changing games in particular, we came up with the notion of *Sliced Serious Games*. As this may well be the next evolutionary step of context-aware mobile games such as *Twostone-IM*, I find the concept worth mentioning here.

Behavior changing games are meant to influence the behavior of a person. It seems reasonable to assume that the effect of such games will be strongest if they accompany their players pervasively throughout the day. However, the constant availability of such games is a double-edged sword. On the one hand, it means that they can permanently influence the player and so hopefully increase the prevalence of a desired behavior (or help decrease the prevalence of an undesired one). On the other hand, this constant availability demands that such games are playable in all the different settings in which a player wishes to play them, regardless of whether she is standing in a driving bus while surrounded by other people or lying on her couch at home. It is these changing contextual situations that distinguish pervasive games from 'regular' video games that are being played at home and in front of a TV screen or a PC monitor. The player of a pervasive game cannot always dedicate her full concentration to the game and the game cannot count on being played for hours in a row. Rather, frequent breaks are the norm for pervasive games and such breaks *slice* the game experience into parts of varying length. Highly adaptive pervasive games must be able to adjust to the different settings that the player is in and they must always expect to be cut in two by an enforced break. *Sliced Serious Games* have this ability.

Developing such highly adaptive games brings challenges on both a technical and a conceptual level. The technical challenges concern the detection of the player's current situation (the classical problem of context aware applications), the preservation of information across different game sessions and, in the case of multiplayer games, the question of how to enable player-to-player communication. While client-server-architectures may be the most obvious choice here, other means of connecting players to one another may provide for more interesting game experiences. Already a decade ago, Sony's mobile video game console PlayStation Portable allowed a player to connect to another player's nearby device via an ad-hoc connection, thus enabling spontaneous game sessions with complete strangers, for instance at an airport. And *Sliced Serious Games* are also challenging on a conceptual level. There is, for example, the question of how game content can be adapted fluently to the situation at hand. We envision an artificial intelligence titled 'game director' that dynamically assembles game content by filling game patterns with predefined game assets (graphics, sounds, texts), and that thus manages to create a game experience best suited to the player's current situation, as detected by the game's underlying context awareness framework. The idea of such a game director is closely related to the principle of procedural content generation, as employed by games such as *Minecraft* or, more recently, *No Man's Sky*. However, different to *Sliced Serious Games*, these examples are not context-sensitive – they do not take the player's situation into account when they expand their game worlds.

The recent success of games such as *Ingress* and *Pokemon Go* demonstrates the demand of users for games that can pervasively integrate themselves into peoples' everyday lives. And although the idea of digital games that blur the borders between the real word and a virtual world can be traced back to a time before the availability of modern-day smartphones, this vision only now starts to become realizable. Highly adaptive *Sliced Serious Games* may be the next evolutionary step of such games and enable users to integrate behavior changing mechanics into their daily routines fluently and in a non-disruptive manner, mainly because such games adapt so well to ever changing contexts. We expect that in the next few years, a fusion of the concepts of contextual awareness, procedural content generation, and behavior change techniques will bring forth this entirely new type of serious gaming [87].

[87] Also see our 2013-published paper on '*Calm Gaming*' [DKH+13].

7. Evaluation

Between August and September of 2016, a field test evaluation of the *Twostone Interventive Measure* (see chapter five) was conducted in the city of Darmstadt, Germany. During a two week period, the thirty participants of the study were frequently prompted by their smartphones with triggering notifications, asking them whether they wanted to play a session of the mobile exergame *Twostone*. The three hypotheses that were supposed to be tested by the evaluation were as follows:

(1) ACCURACY ADVANTAGE OF INTELLIGENT TRIGGERS – The first hypothesis was that an intelligent and adaptive trigger that uses a set of indicators to identify situations suited for intervention attempts would have a higher intervention accuracy (number of successful intervention attempts divided by total number of intervention attempts) than a stubborn trigger trying to activate the user in fixed intervals.

(2) IMPROVEMENT THROUGH ADDITIONAL INDICATORS – The second hypothesis assumed that a high-insight trigger that makes use of a larger number of indicators would have an advantage in terms of triggering accuracy over both low- and medium-insight triggers based on (significantly) lower numbers of indicators.

(3) HIGHER ACCEPTANCE OF DISCERNING MEASURES – The third hypothesis stated that a user's acceptance for and her overall satisfaction with discerning interventive measures that distinguish between opportune and ill-timed interventions would be higher than her acceptance for stubborn mechanisms that do not make this differentiation.

The original plan was to compare all four trigger types as discussed in chapter four, namely a stubborn interval-based trigger and three discerning triggers with varying levels of insight. However, due to difficulties in the acquisition of suited study participants (see section 7.2) the evaluation had to be limited to the comparison of only three triggers instead. In order to ensure sufficiently large study groups of at least ten members each, it was decided to exclude the low-insight trigger from the evaluation in favor of the other three trigger types. Prior to the beginning of the actual evaluation phase, all study participants were supplied with the *Twostone-IM* as described in chapters five and six, consisting of the mobile exergame *Twostone* and an additional application that added a triggering mechanism to the game. Every participant was then assigned to one of three study groups and the participants' triggering applications were remotely set to function as the corresponding type of trigger.

In addition, all participants were asked to specify their resting times in the trigger application's configuration menu. On each day of the evaluation, beginning one hour after their specified wake up time and until one hour prior to their bedtime, the trigger application released trigger notifications, with the exact triggering policy depending on the participant's respective study group. All such triggers appeared as push-up notifications on the user's smartphone screen and were accompanied by a soft vibration of the device and a short beep-sound [88]. If a user chose to confirm the trigger, then a session of *Twostone* was automatically started and the map closest to the user's current position was loaded. If the trigger was denied, then this was registered as an indication that the respective situation was not suited for triggering, at least by the two discerning measures. Finally, triggers that were simply ignored by the user automatically disappeared after a short period. The only rule that participants were given was that they were not allowed to play *Twostone* unless triggered for doing so. They were also expected to not uninstall either of the two applications, although a low number of participants apparently did so for short periods of time, judging from their server-based profiles. Three participants dropped out of the running evaluation entirely, leaving a total of N=27 field study participants.

[88] The push-up notification stated "*Time for Twostone – This is a good opportunity to play Twostone, don't you think?*" and featured a button for confirmation and another one for cancelation. Appendix B contains a screenshot showing such a triggering message.

7.1 Planning and Conduction

The field study of *Twostone-IM* was originally scheduled to begin on Monday, the 15th of August 2016 and planned to last for three full weeks until Sunday, the 4th of September. During a two-week acquisition period, roughly four hundred people were asked to participate, both friends and relatives of the evaluation team [89], as well as students of Darmstadt's two universities. Candidates were either invited personally, or by mail. The invitation flyer that can be found at the beginning of appendix B was attached to all mails in German and English language. All candidates were promised two movie theatre vouchers as an incentive and informed that these would be handed out at the end of the study. The tickets were financed privately by this thesis' author.

A total of fifty-six persons announced an interest in participation by sending a mail to the address specified in the flyer. These people were considered to be the 'potential participants'. To all of them, the first study manual and the pre-study questionnaire were sent, see appendices B and C [90]. On Sunday, the 14th of August, one day prior to the originally planned beginning of the field test, all potential participants were informed that due to technical difficulties [91], the start of the evaluation had to be postponed by a week until Monday, the 22nd of August. Because the planned ending of the study had already been communicated to the study participants and due to other time constraints, the originally envisioned evaluation length of 21 days had thus to be reduced to 14 days.

On Saturday, the 20th of August, the second manual was sent to all potential participants and they were told that only those persons could be included into the study that had worked through both manuals and that had the full *Twostone Interventive Measure* running on their smartphones. At the beginning of the two-week evaluation, a total of thirty-two persons had completed this step [92]. Two of them were excluded from the study, one because of technical problems and a second one for organizational reasons [93]. The remaining N=30 persons were considered to be the actual 'study participants'. The nine pre-study questionnaires received from potential participants who had chosen to not take part in the actual evaluation (including those of the two excluded candidates) were dropped. The evaluation team then manually allocated the study participants to three groups: The 'Babbage group', the 'Von Neumann group', and the 'Turing group'. Membership in the Babbage group meant that the respective person's triggering mechanism was remotely set to function as a stubborn, one-hour interval trigger, while membership in the Von Neumann group or the Turing group implied the provision with a medium-insight or a high-insight trigger, respectively. Due to the low total number of participants, a fourth group that was originally planned – the 'Zuse group' – had to be canceled. Figure 11 depicts the evaluation's originally planned and actual procedures.

---

[89] The team of five that was supervising the evaluation consisted of this thesis' author, a master thesis student under the author's supervision [KOM-M-0572], and three of the author's student assistants (Chris Michel, Gerhard Säckel, and Tobias Welther).

[90] The questionnaires were sent as fillable PDF documents so that they could be answered electronically.

[91] We encountered two main problems during the final testing of the 'Trigger Application' (see chapter five for details). The first problem was related to the 'Doze'-mode that had been newly introduced with the latest update of the Android operating system, Android 6.0 (API level 23). Doze is a feature meant to reduce the battery consumption of applications constantly running in the background, such as the trigger app. This new mechanism affected our application in a way that when the trigger's sleep time exceeded a certain length – apparently five minutes –, then the application did not wake up from its dormant state. We discovered this problem only a few hours before the originally planned start of the evaluation, but were able to fix it in the subsequent days. Our second technical problem that contributed to the decision to delay the evaluation's start was the significant number of interested participants that did not meet the minimum requirements of participation that we had originally asked for. For reasons related to the specific implementation of the context recognition framework (see chapters four and five), the evaluation flyers stated that participants needed an Android-based smartphone running the Android OS in version 5.0 or above. However, the amount of interested participants that did not meet this requirement on the one hand, and the low total number of potential participants on the other, led to the decision to delay the evaluation and to use the acquired time for trying to also make the triggering application work on Android version 4.0 and above, which was achieved. While this decision may have reduced the duration of the evaluation, it also increased the number of participants; this seemed to be a meaningful tradeoff.

[92] We asked for a short feedback by mail when a person had successfully installed and initialized the full *Twostone-IM*.

[93] The participant that had to be ruled out due to technical problems was using an exotic low-end phone that did not feature the same hardware components as most other smartphone models. The second participant was excluded from the study because a number divisible by 3 was required in order to evenly distribute the participants among the three groups. Both participants were informed of the reasons for their exclusion and each was compensated with the two promised movie vouchers.

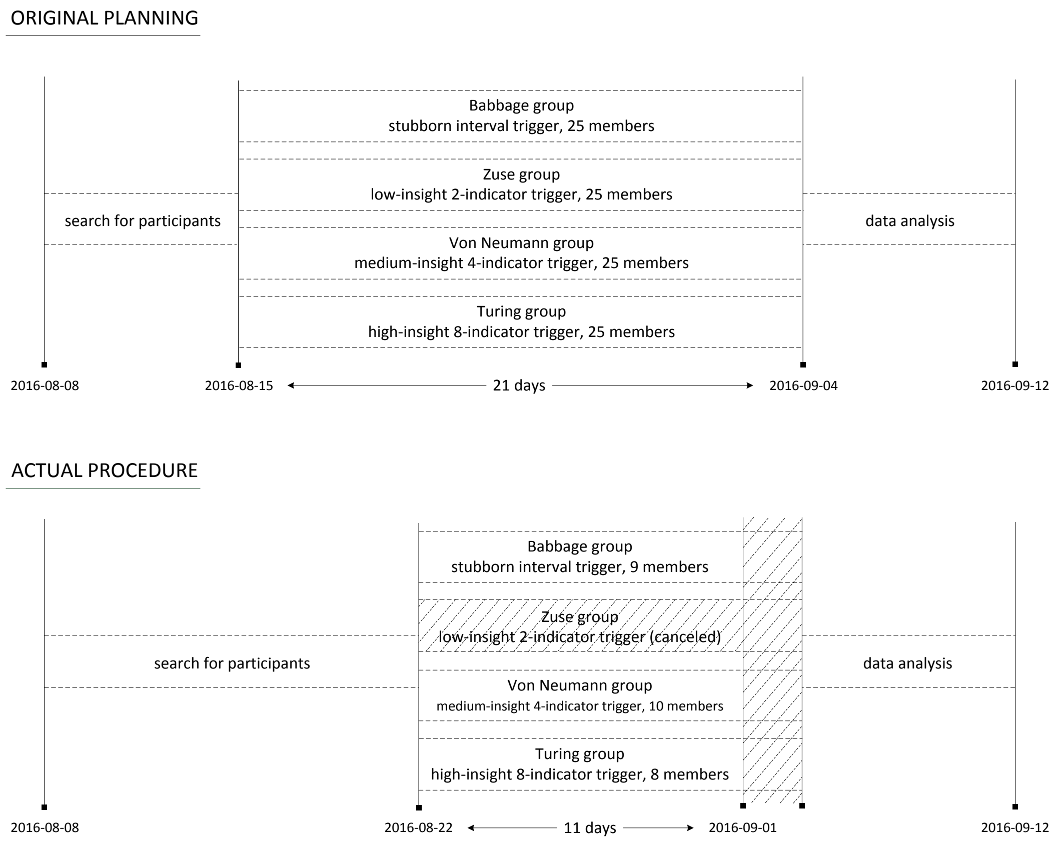


Figure 11: Evaluation Timelines.

FIGURE NOTES – The upper timeline depicts the original planning with four groups and a duration of 21 days, while the lower timeline shows the actual procedure with three groups and a duration of 11 days.

For technical reasons, the allocation of the study participants to the three evaluation groups was not entirely randomized; the exact process is explained in the next section. In the following two weeks, all participants were triggered by their respective interventive measures up to twelve times a day with the *O*-period beginning one hour after a participant's specified wake up time and ending one hour before the start of her specified bedtime. As pointed out previously, the only rule that was given to the study participants was that they were not allowed to play the game *Twostone* unless they had received and confirmed a trigger notification.

Three days before the planed ending of the study, on Thursday afternoon, the application's server experienced a crash. For the time until the server was restored, the study participants were not able to play *Twostone*. After about an hour of the crash, they were informed of the problem by mail. At the evening of the same day, the server was successfully restored and the study participants were once again able to play. Alas, as was later discovered, all data sets that were uploaded after the crash were corrupted and had to be omitted from the analysis. As a consequence, the evaluation was thus effectively shortened by another three days and only eleven days of data could be used.

On the morning of Monday, the 5th of September, the study participants were informed that the evaluation had ended and they were sent the second questionnaire that can be found in appendix B. A total of N=27 study participants returned this second questionnaire within three days. The other three participants (one of them was assigned to the Babbage group, two were members of the Turing group) were considered to be drop-outs and their data sets were excluded from the results.

7.2 Problems and Limitations

The evaluation suffered from several shortcomings and the results presented in the next section need to be understood in this light. The known problems and limitations are as follows.

(1) CHANGE OF DESIRED BEHAVIOR – It was already pointed out that changing the desired behavior from originally '15 minutes of brisk walking' to 'playing a map of *Twostone*' is not without problem. While the actual physical activity remained practically the same and the two behaviors can thus be assumed to have comparable health effects, triggering a user for playing a mobile exergame instead of asking her to simply take a walk around the block may have resulted in *lower* intervention success rates (also see chapter six). This can in part be explained by the quality of the employed application. Although *Twostone-IM* was in development for several years and must be considered to be a fairly advanced prototype for 'scientific standards', we received multiple bug reports and feature requests from our participants. While several users positively remarked on the underlying game mechanic, they also negatively commented on the game's actual realization. The aspects that were most frequently criticized were the usability of the game's level editor, the accuracy of the user's positioning in the virtual world, and the game's overall polishing and stability. In consequence to this early feedback, an updated version of the game was made available prior to the start of the actual two week evaluation. However, based on the initial feedback, we (rightly) assumed the existence of a negative effect on some participants' motivation for playing the game. This, of course, also affected their inclination to accept intervention attempts intended to make them play the game. Furthermore, being a location-based game, *Twostone-IM* requires a map in the player's vicinity in order to be playable. While we asked all participants to create such maps in close vicinity to their homes and their working places in advance of the actual evaluation phase [94], we did not verify the execution of this step. In addition, triggers reaching a user while she was on the move were also likely to be unsuccessful as a confirmation would have required the user to first create a map at her location before she was able to play the game. These problems have negatively affected the users' ability for the desired behavior. In sum, this means that for some users, the employment of *Twostone-IM* is likely to have had a negative effect on the probability of successful intervention attempts. However, due to the lack of a real control group, we were not able to verify this assumption (see below). An additional problem related to the employment of the game *Twostone-IM* consisted in the fact that although we had asked the study participants to not play the game without the receival of a trigger notification, we did not establish means to enforce this restriction. It cannot be excluded that some participants may have played the game although having declined a timely trigger and thus distorted the automatically assessed intervention attempt success rates.

(2) LOW TOTAL NUMBER OF PARTICIPANTS – The total number of N=27 participants is too low to allow for any firm statistical conclusions. The original planning was for a hundred persons to be equally distributed among four groups (see Figure 11), but the actual number of participants even enforced the cancelation of one of these planned groups to ensure to not go below a total of ten persons per group. Four reasons have been identified to have contributed to the difficulties in the acquisition of study participants. The first and possibly the main problem was that the *Twostone-IM* was only available for Android phones – and originally only for Android phones running at least the Android operating system in version 5.0. A significant number of persons contacted the evaluation team and asked for an "*iPhone version*" of the *Twostone-IM*. While the cross development of an additional iOS-based version of the application would have been time and cost consuming, in hindsight it seems as if these efforts would have been worthwhile. With limiting the availability of

[94] See the two manuals in appendix B for the full instructions that were given to the potential participants prior to the beginning of the evaluation.

Twostone-IM to Android-based devices, a significant number of potential candidates were excluded from the study from start. The second problem that may have had an effect on the user acquisition was the long duration of the evaluation in conjunction with a relatively small incentive for participation. While lab-based evaluations usually do not require more than a single hour to participate, we initially asked for three full weeks. This may have put off some of the potential candidates [95]. The third obstacle that we found to interfere with our search for study participants were privacy concerns. We informed all potential participants that we would make use of their smartphone's internal sensors in order to gather information about their contextual situation. We received feedback from a number of candidates that this was the main reason for them to lose their interest for taking part in the study. And finally, the timing of the evaluation phase apparently also contributed to some potential candidates deciding against a participation (see below).

(3) HOMOGENEITY OF PARTICIPANTS – Related to the problem of the low total number of participants is their homogeneity. The vast majority of our participants were between 20 and 39 years of age, had pursued or were currently pursuing university studies, were fairly active, and spent multiple hours per week playing video games. Furthermore, they had a surprisingly high knowledge of technical details, indicating that the majority of them had a technical background [96]. The homogeneity of the participants can be considered a double-edged sword, being both a problem and a benefit. On the one hand, it reduces the validity of statements about potential effects of pervasive smartphone-based interventive measures on the general populace. On the other hand, however, it also increases the validity of statements about this specific subgroup. A thorough analysis of the group of study participants follows in the next section.

(4) NON-RANDOMIZED ALLOCATION OF PARTICIPANTS – For technical reasons, the allocation of the participants to the three trigger groups (see Figure 11) was not entirely randomized. We originally asked for participants owning a smartphone with an Android version higher than 5.0 ('Android Lollipop'). The main reason for this was that the context detection framework that was developed in part by [KOM-M-0544] and discussed in chapter four made use of certain operating system features that were not available in its lower versions. However, we nevertheless received multiple requests for participation by users who used Android in a lower version, and the low overall number of persons interested in participation did not permit us to easily exclude these persons from the evaluation. We thus decided to postpone the start of the evaluation by a week and to use this time to adapt the context recognition framework in order to also make it work on older devices. As an extensive change to the framework would have required more time and testing, we decided for a compromise: All users using a smartphone with Android version 5.0 or below would automatically be assigned to the stubborn trigger group (the Babbage group) that did not rely on the assessment of indicators and thus not on the critical aspects of the underlying context recognition framework. The automatic readout and upload of the users' OS versions into the *Twostone-IM* database allowed us to separate those users with a lower OS version from the rest. A total of seven people had a version equal to or lower than 5.0 and were thus directly assigned to the Babbage group. After this step, the other twenty-three evaluation participants were randomly distributed among the three groups such that at the end of the procedure, each group had exactly

[95] Other comparable field studies promised significantly larger rewards for participation. [ORN+15] rewarded each of the 41 participants of their four week study with 100 U.S. dollars and the chance to win one of two smart watches, [LLL+13] compensated each of their 32 study participants with 75 U.S. dollars and the chance to win an iPad. On the other hand, [MMH+15] managed to recruit 35 participants for their three week study without the provision of any material incentives.

[96] Question pre02: "*How old are you?*", four possible answers with second answer being "*20-39*", N=30, M=2.03, SD=±0.18. Question pre03: "*What's your highest qualification (including currently pursuing)?*", four possible answers with fourth answer being "*Bachelor/Master*", N=30, M=3.77, SD=±0.57. Question pre05: "*How many hours per week do you do sports or exercise?*", four possible answers with third answer being "*1-3 hours*", N=30, M=2.20, SD=±0.85. Question pre06: "*How many hours per week do you spend playing video games?*", four possible answers with third answer being "*4-7 hours*", N=30, M=2.63, SD=±1.30. Question pre17: "*I could explain the difference between an accelerometer and a gyroscope*", five point disagree-agree Likert scale, N=30, M=3.37, SD=±1.77.

ten members. Although we consider this interference to be a minor violation of the rules of good scientific practice, it is still a noteworthy one.

(5) MISSING CONTROL GROUP – Another cutback caused by the low total number of participants was the lack of real control groups, one of which would have been triggered for simply taking a 15 minute walk, a second one which would have received no interventions at all but would have been supplied with the game *Twostone*, and finally a group that would have used neither of the two applications. This would have allowed for more solid assumptions about the relative amount of physical activity that could directly be attributed to the interventive measure (if any at all) and also for making more clear statements about the effects (positive or negative) that the consolidating mechanism *Twostone* had on the study participants. The most preferable study design would have been the alternating assignment of the study participants to the seven different groups [97] based on the latin square principle [Gra48], such that each study participant would have been confronted with each variant. This, however, would have required far more participants and a significantly longer duration.

(6) SHORT DURATION – The short duration of the study is a problem in so far as the two discerning triggering mechanisms were not given the opportunity to completely adapt to their respective users. Rather, especially the full-insight triggers of the Turing group were still learning from the user feedback and adjusting their *ISC*-functions when the evaluation ended. As such, both the accuracy rates of the intervention attempts and the feedback results from the study participants do not entirely reflect the potential quality of this approach.

(7) FRIENDS' BIAS – As pointed out in the first section, the study participants were both students of Darmstadt's two universities and friends and relatives of the study support team. The relation of these groups was about 2:1 in favor of the students. Although all participants were urged to fill both questionnaires based on their true opinions and to react to intervention attempts as normally as possible, the existence of a 'friends' bias' that may have led to more favorable feedback and behavior cannot be fully excluded. However, as we distributed the participants over the three study groups as randomly as possible (see above), the effect of such a bias should be equally high or low in all three groups.

(8) BAD TIMING – During the study preparation we anticipated several timing-related problems that were later confirmed *expressis verbis* by our study participants (and can also be deduced from the evaluation results). To begin with, the evaluation phase overlapped with the exam dates of several students. Some of the potential candidates announced a general interest in participants, but excused themselves due to vacation plans. And finally, at some days during the evaluation phase, the outside temperatures in Western Germany were lying well above average and reached 30 degrees Celsius (86 Fahrenheit) and more. All these factors are likely to have (negatively) affected the motivation and ability of study participants for playing *Twostone* as well as their tolerance towards ill-timed intervention attempts. It must thus be assumed that some of the results presented in the next section would differ if a similar evaluation was conducted at another time of year.

While all these problems limit the significance of the findings presented in the next section, they are not so severe as to render them entirely irrelevant. Rather, the results of the study indicate a clear tendency of the effectiveness of smartphone-based interventions for physical activity, even if only applicable to a very specific group of users.

[97] The seven groups are as follows: The first group being supplied with a stubborn trigger and *Twostone*, three groups being supplied with discerning triggers of varying insight levels and *Twostone*, a control group with discerning interventions of any type but without *Twostone*, its 'counterpart' with *Twostone* but without a trigger, and finally a group with neither of the two apps but means to assess its participant's overall activity levels.

7.3 Results

At the end of the evaluation phase, the automatically logged and uploaded reactions of our study participants to their respective measures' intervention attempts revealed the following numbers [98]. The members of the Babbage group, who had been supplied with a stubborn one-hour trigger, had confirmed a total of 20 interventions and declined 180, which equals an overall accuracy of 0.10. The study participants that belonged to the Von Neumann group who had been supplied with a medium-insight measure [99] had confirmed 33 interventions and declined 152, resulting in an accuracy of 0.18. Finally, the members of the Turing group who had tested a high-insight measure [100] had confirmed 19 and declined 224 triggers, which means an accuracy of 0.08. In regard to the three initially stated hypotheses, this implies that the first hypothesis – discerning measures have an accuracy advantage over stubborn measures – had been partly confirmed by the study, while the second hypothesis – additional indicators increase accuracy – had been disproved.

These results require some explanation. Most notably, the numbers appear to be fairly low. For instance, given that the Babbage group had nine members (one of the initially ten members dropped out of the running evaluation) and was supplied with a stubborn one-hour trigger, then during a period of eleven days a total of roughly 9 (members) * 11 (days) * 12 (triggers per day) = 1,188 intervention attempts should have been made in this group, under the assumption that all participants had defined their sleeping times in a way that they allowed for the initiation of twelve triggers per day. However, only about a tenth of these triggers had been answered by the study participants, and similar proportions are encountered with the other two groups. This phenomenon may have been caused by several different factors. First, it can indeed not be excluded that some study participants defined their 'awake times' so narrowly that these only allowed for a small number of triggers per day. Second, several study participants confessed that their triggering applications were temporarily disabled, either by accident [101] or on purpose. And finally, almost all participants confirmed that they had occasionally ignored intervention attempts and assumed to also have missed a few [102]. Especially the last of the named problems is worthy of investigation and should be addressed by further research. In the context of this work, the question of *how* to trigger the user was essentially ignored, which, according to the evaluation results, may be problematic. But as all of these problems affected all three groups, the numbers stated above can be considered to correctly reflect the accuracy proportions between the three groups – which leads to the question, why the accuracy of the stubborn measure used by the Babbage group was surprisingly high, and why the accuracy of the high-insight measure that the Turing group was supplied with so low.

Figure 12 shows that this difference cannot be accredited to significant differences in the study group compositions, although it must be remarked that by chance, the majority of the female study participants were assigned to the Babbage group, while the Turing group may have only consisted of male participants (one member of this group did not report his or her gender). The leftmost bars of the figure display the mean responses to the question, how much a participant enjoyed playing *Twostone*, a factor that can be assumed to have had a significant influence on the probability of successful

[98] The numbers for discerning measures include confirmed and declined learning notifications.

[99] The medium-insight measure was based on the user's location, activity, the device's battery level, and the ambient noise.

[100] In addition to the indicators employed by the medium-insight measure, the high-insight measure of the Turing group also relied on the detection of the local weather conditions, the outdoor temperature, the number of steps that already taken during the day and whether or not she had enabled her smartphone's do-not-disturb-mode for making intervention decisions.

[101] As pointed out before, the triggering application was implemented to function as so-called 'background application' and thus capable of changing into a dormant state. It was also programmed to automatically start itself when the user turned on her device after it had been switched off. However, we found this functionality not to be reliable on all types of devices; especially Samsung devices are problematic in this regard in that they prevent such background applications to automatically start themselves. About a quarter of our participants used Samsung smartphones and they were informed about this problem. Still, it must be assumed that at least a few of them accidentally left the triggering application disabled after having switched their devices off and on again, until they thought of the evaluation and started the triggering application by hand.

[102] Question pos36: "*I sometimes did not notice the trigger*", five point disagree-agree Likert scale, $N = 27$, M = 3.22, SD = ± 1.53, question pos37: "*I sometimes ignored the trigger*", five point disagree-agree Likert scale, $N = 27$, M = 3.74, SD = ± 1.46.

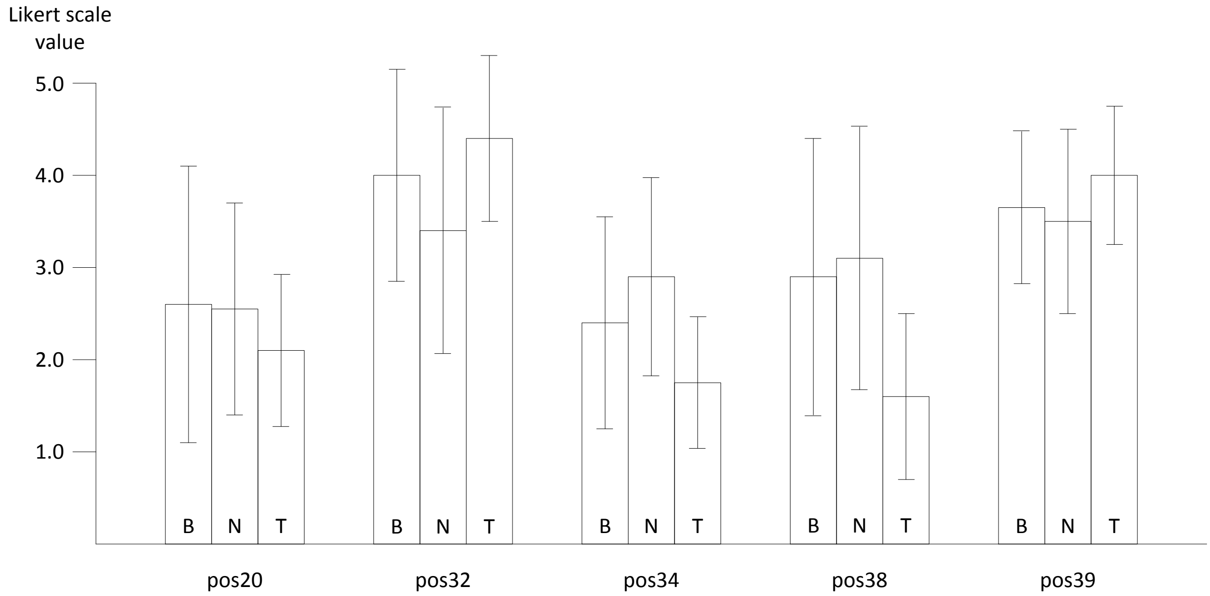


Figure 12: Mean Responses to Selected Trigger-related Questions.

FIGURE NOTES – The figure shows the mean responses to selected questions from the post-study questionnaire, along their associated standard deviations. The values were calculated separately for the members of each of the three study groups, whereby the bars marked with 'B' visualize the answers of the members of the Babbage group (stubborn trigger, N=9), the bars marked with 'N' represent those of members of the Von Neumann group (medium-insight discerning trigger, N=10), and the bars marked with a 'T' show the answers of the members of the Turing group (high-insight discerning trigger, N=8). The five questions were as follows. Question pos20: "I enjoy playing Twostone", question pos32: "The trigger was annoying", question pos34: "The trigger selected good opportunities for playing", pos38: "Without the trigger, I would have played less Twostone", and pos39: "The trigger needs to be more intelligent".

interventions. As can be seen, the differences between the three groups were small here, while in contrast, we find significant differences for the questions of how annoying the intervention attempts were considered by the participants (pos32), and whether they believed that the application selected good opportunities for such interventions (pos34). While the medium-insight trigger of the Von Neumann group performs best for all questions, both the comparably high approval for the stubborn trigger and the low values of the high-insight trigger may be found to be irritating.

One possible explanation for the low performance of the high-insight measure is that it was not given sufficient time to successfully adapt to the respective users. As, due to its larger number of possible indicator value combinations, it had much more situations to learn than the medium-insight trigger, it may still have adjusted its confidence values at the time when the evaluation ended. In addition, its apparent ill-timed interventions may also be attributed to a too relaxed confidence gate (see chapter four). Another explanation for the bad performance of the high-insight measure may be that one or multiple of the additionally employed indicators were not as relevant to the user's decision as assumed. As such, they would only have served to complicate the development of a sound intervention strategy. In contrast, the comparably high approval for the stubborn-trigger may be in part due to its predictability. We had asked our study participants to not play *Twostone* unless having been triggered for doing so. In this regard, users may have utilized the stubborn measure's predictability to their advantage and actively awaited a trigger in order to be able to play the game.

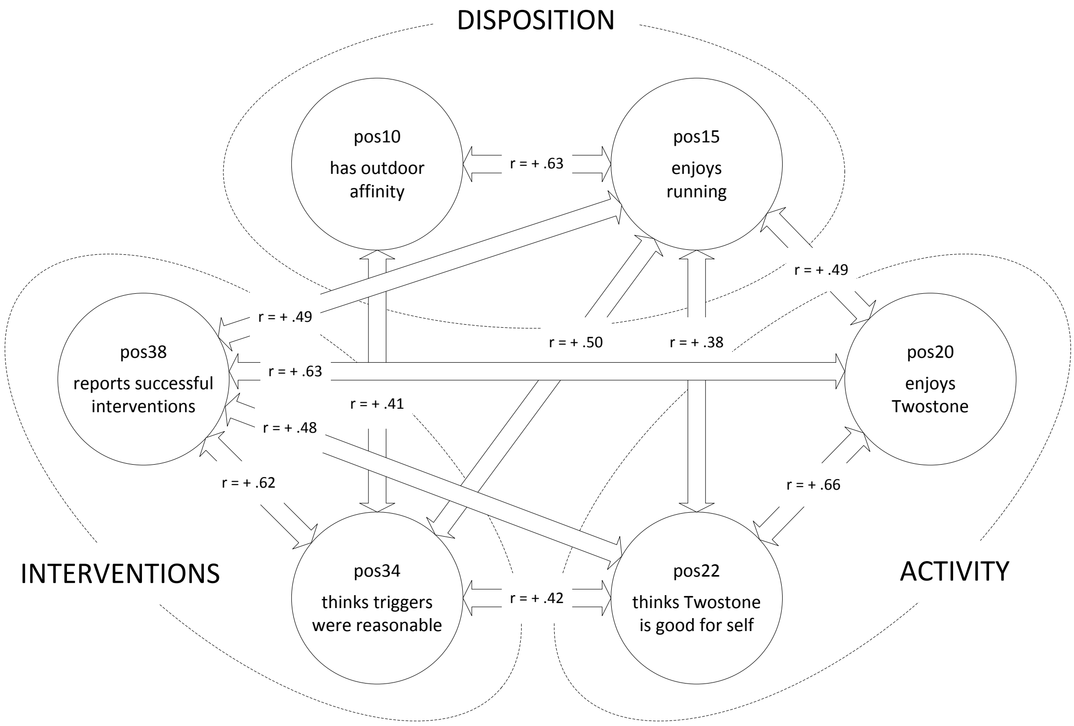


Figure 13: Disposition, Activity, and Interventions.

FIGURE NOTES – The figure shows the interdependence between a person's disposition, the desired behavior, and the success of intervention attempts. Displayed are only statistically relevant correlations (p<.05).

The comparably high approval rates found in the Babbage and the Von Neumann groups make it difficult to draw a clear conclusion on whether or not the third hypothesis – whether discerning measures have higher acceptance rates than stubborn measures – was confirmed by the evaluation. Although the stubborn trigger was indeed found to be slightly more annoying than the medium-insight measure, the results are 'too close to call'. It thus seems reasonable to leave the third hypothesis undecided and open for further investigation.

An interesting finding of the evaluation came from the analysis of the correlation coefficients that quantify the dependencies between the responses to the second questionnaire (see appendix C for both questionnaires and see appendix A for the full listings of all response means and all statistically relevant correlations). Strong and statistically relevant correlation values were found to exist between a person's disposition, in this context whether or not she enjoys being outside and running, her attitude towards the activity in question, here playing the mobile exergame *Twostone*, and the success of interventions. Persons who are likely to enjoy running as a sport are also likely to enjoy playing *Twostone* and either of the two traits increases the probability of successful interventions. This finding is highly satisfactory, as it indicates that interventive measures can increase the prevalence of a desired behavior, at least if the target user is sufficiently motivated for this activity. Combined with the comparably high overall accuracy of the medium-insight trigger, this means that such technology-based discerning measure can be employed to assist users in overcoming perceived barriers for physical activity and to thus increase their chances of staying healthy.

8. Discussion and Outlook

Performing frequent physical activity of sufficient duration and intensity is essential for the preservation of one's health and wellbeing, but large parts of the European and the US-American population perceive barriers that prevent them from doing so. Of the three main types of such barriers, the lack of time appears to be the biggest problem in modern-day societies. Pervasive technology-based interventive measures that are capable of pointing out opportunities for a few minutes of medium-intensity physical activity such as brisk walking may be able to help counteract this problem. In recent years, first commercially available products such as the *Apple Watch* have appeared that aim to do just this. However, as of today, the majority of such devices and applications are little more than stubborn reminders that will simply notify their users in fixed time intervals and usually do not take into account the specific circumstances that the user is currently in. As such, they will also try to reach out to the user when she is neither able nor motivated for performing the respective activity.

This thesis is focused on the development of a theoretical model for a more advanced type of such interventive measures: Discerning measures, context-aware devices and applications that base their decisions of whether or not to trigger the user on a number of contextual parameters and that are capable of withholding intervention attempts if given reason to believe that these will not be successful. In a two-week field study with almost thirty participants, the performance of such discerning measures was compared to the performance of a stubborn, interval-based variant, and the study revealed that discerning measures can indeed achieve higher accuracy rates and approval ratings than their stubborn counterparts – but only, if they are well designed. The theoretical concept presented in this thesis lays the foundation for the creation of such well-designed discerning interventive measures. But while the thesis' main goal, the verification of the model's validity, was achieved, not all of the encountered problems were solved; instead, several additional questions emerged. A good part of the sidestepped problems is thereby related to non-technical questions that demand the attention of domain experts from other fields of study. Among these is the investigation of the number of intervention attempts that a user can be expected to tolerate. The evaluation results of the field study indicate that the upper limit of twelve attempts per day that was employed by the tested prototypes may indeed have been much too high. As a related problem, the question of how triggering notifications should be designed was also largely ignored, which again was proven to be problematic by the study results.

The majority of opportunities for further research, however, are related to the decision making procedure of discerning measures. For example, finding ways of efficiently overcoming the cold start problem that affects all discerning measures is crucial for shortening their adaptation phases. As during this initial period of its employment, a discerning measure still learns from the user's reactions to its intervention attempts, its overall accuracy will be low and as such, the risk of annoying the user to a point that she decides to get rid of the measure will be the highest. Consequently, approaches for shortening this phase to an absolute minimum are highly desirable. Analogously, finding strategies for minimizing the problem of limited predictability by making best use of the limited number of allowed intervention attempts is another pressing issue. If granted only a few 'free throws' during a specific period, how can the measure identify the most promising situations for making use of these? However, for many more years to come, the problem of partial observability will remain to be the main challenge of the design of discerning measures. Given the current state of the art, many of the parameters that will affect a user's ability and/or motivation for a desired behavior such as brisk walking are simply not observable with technical means. While general advances in the field of contextual awareness will also serve to reduce the severity of this problem, finding workarounds that can compensate the lack of insight in other ways promises to improve the overall accuracy of discerning measures much more quickly.

Bibliography

[AAH+97] Gregory D. Abowd, Christopher G. Atkeson, Jason Hong, Sue Long, Rob Kooper, Mike Pinkerton. Cyberguide – a mobile context-aware tour guide. In: Wireless Networks, Vol. 3, No. 5, 1997.

[ABM+10] Akin Avci, Stephan Bosch, Mihai Marin-Perianu, Raluca Marin-Perianu, Paul Havinga. Activity recognition using inertial sensing for healthcare, wellbeing and sport applications – a survey. In: Proceedings of the 23th International Conference on Architecture of Computing Systems, 2010.

[AHL+93] Barbara E. Ainsworth, William L. Haskell, Arthur S. Leon, David R. Jacobs Jr., Henry J. Montoye, James F. Sallis, Ralph S. Paffenbarger Jr. Compendium of physical activities – classification of energy costs of human physical activities. In: Medicine and Science in Sports and Exercise, Vol. 25, No. 1, 1993

[AHH+11] Barbara E. Ainsworth, William L. Haskell, Stephen D. Herrmann, Nathanel Meckes, David R. Bassett Jr., Catrine Tudor-Locke, Jennifer L. Greer, Jesse Vezina, Melicia C. Whitt-Glover, Arthur S. Leon. 2011 compendium of physical activities: A second update of codes and MET values. In: Medicine and Science in Sports and Exercise, Vol. 43, No. 8, 2011.

[AJT05] Oliver Amft, Holger Junke, Gerhard Tröster. Detection of eating and drinking arm gestures using interntial body-worn sensors. In: Procceedings of the 9th IEEE International Symposium on Wearable Computers, IEEE, 2005.

[Bau02] Hans-Ulrich Baumgarten. Acting against better knowledge – on the problem of the weakness of the will in Plato, Davidson, and Kant. In: The Journal of Value Inquiry, Vo. 36, No. 2, 2002.

[Baa97] Bernard J. Baars. In the theater of consciousness – global workspace theory, a rigorous scientific theory of consciousness. In: Journal of Consciousness Studies, Vol. 4, No. 4, 1997.

[Bar96] Richard Bartle. Hearts, clubs, diamonds, spades – players who suit MUDs. In: Journal of MUD Research, Vol. 1, No. 1, 1996.

[BC07] Gabriela Beirao, J.A.S. Cabral. Understanding attitudes towards public transport and private car – a qualitative study. In: Transport Policy, Vol. 14, No. 6, 2007.

[BC08] Fabio Buttussi, Luca Chittaro. MOPET – a context-aware and user-adaptive wearable system for fitness training. In: Artificial Intelligence in Medicine, Vol. 42, No. 2, 2008.

[BCE+01] Vincent Bazinette, Norman H. Cohen, Maria R. Ebling, Guerney D.H. Hunt, Hui Lei, Apratim Purakayastha, Gregory Stewart, Luke Wong, Danny L. Yeh. An intelligent notification system. IBM Research Division, Yorktown Heights, New York, USA, 2001.

[BL13] Ivo Blohm, Jan M. Leimeister. Gamification – Design of IT-based enhancing services for motivational support and behavioral change. In: Business & Information Systems Engineering, Vol. 5, No. 4, 2013.

[BLP+11] Linas Baltrunas, Bernd Ludwig, Stefan Peer, Francesco Ricci. Context-aware places of interest recommendations for mobile users. In: Proceedings of the International Conference of Design, User Experience and Usability '11, Springer, Germany, 2011.

[BPT06] Mark Blum, Alex S. Pentland, Gerhard Tröster. InSense – interest-based life logging. In: IEEE MultiMedia, Vol. 13, No. 4, 2006.

[Bro96] Peter J. Brown. The stick-e document – a framework for creating context-aware applications. In: Electronic Publishing, Vol. 9, No. 1, 1996.

[Cam13] Bill Campbell. Energy balance. In: Sports Nutrition – Enhancing Athletic Performance. CRC Press, 2013.

[CDM+00] Keith Cheverst, Nigel Davies, Keith Mitchell, Adrian Friday. Experiences of developing and deploying a context-aware tourist guide – the GUIDE project. In: Proceedings of the 6th International Conference on Mobile Computing and Networking, ACM, New York, USA, 2000.

[CLL+14] Eun K. Choe, Nicole B. Lee, Bongshin Lee, Wanda Pratt, Julie A. Kientz. Understanding quantified-selfers' practices in collecting and exploring personal data. In: Proceedings of CHI'14, Toronto, Canada, 2014.

[CM92] Paul T. Costa, Robert R. MacCrae. Revised NEO personality inventory (NEO PI-R) and NEO five-factor inventory (NEO FFI) – professional manual. Psychological Assessment Resources Inc., Odessa, USA, 1992.

[CNC10] Mauro Conti, Vu T.N. Nguyen, Bruno Crispo. CRePE – context-related policy enforcement for Android. In: Proceedings of the 13th International Conference on Information Security, Springer, Berlin, Germany, 2010.

[CFG+04] Adrian D. Cheok, Siew W. Fong, Kok H. Goh, Xubo Yang, Wei Liu, Farzam Farzbiz, Yu Li. Human Pacman – a mobile entertainment system with ubiquitous computing and tangible interaction over a wide outdoor area. In: Journal of Personal and Ubiquitous Computing, Vol. 8, No. 2, 2004.

[Cra09] Garry Crawford. Forget the magic circle. University of Salford, Manchester, UK, 2009.

[Csi75] Mihaly Csikszentmihalyi. Beyond boredom and anxiety – experiencing flow in work and play. Jossey Bass Wiley, San Francisco, USA, 1975.

[Csi90] Mihaly Csikszentmihalyi. Flow – the psychology of optimal experience. Haper Perennial, New York, USA, 1990.

[DA00] Anind K. Dey, Gregory D. Abowd. CybreReminder – a context-aware system for supporting reminders. In: Proceedings of the 2nd International Symposium on Handheld and Ubiquitous Computing, Springer, Berlin Germany, 2000.

[DBN12] Soumya K. Datta, Christian Bonnet, Navid Nikaein. Android power management – current and future trends. In: Proceedings of the 1st IEEE Workshop on Enabling Technologies for Smartphone and Internet of Things, IEEE, 2012.

[DDK+11] Sebastian Deterding, Dan Dixon, Rilla Khaled, Lennart Nacke. From game design elements to gamefulness – defining 'gamification'. In: Proceedings of MindTrek'11, ACM, 2011.

[Dec71] Edward L. Deci. Effects of externally mediated rewards on intrinsic motivation. In: Journal of Personality and Social Psychology, Vol. 18, No. 1, 1971.

[Dey98] Anind K. Dey. Context-aware computing – the CyberDesk project. In: Proceedings of the American Association for Artificial Intelligence '98 Spring Symposium on

Intelligent Environments. 1998.

[Dey01] Anind K. Dey. Understanding and using context. In: Personal and Ubiquitous Computing, Vol. 5, No. 1, 2001.

[DKR99] Edward L. Deci, Richard Koestner, Richard M. Ryan. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. In: Psychological Bulletin, Vol. 125, No. 6, 1999.

[DL68] John M. Darley, Bibb Latane. Bystander intervention in emergencies. In: Journal of Personality and Social Psychology, Vol. 8, No. 4, 1968.

[DRC+14] Irina Diaconita, Andreas Reinhardt, Delphine Christin, Christoph Rensing. Bleep! Bleep! Determining smartphone locations by opportunistically recording notification sounds. In: Proceedings of the 11th International Conference on Mobile and Ubiquitous Systems, ACM, New York, USA, 2014.

[Eas16] Sophie Eastaugh. British Pokemon Go player says catching 'em all shed pounds. CNN 2016-07-29, 20016.

[Eva96] James D. Evans. Straightforward statistics for the behavioral sciences. Brooks/Cole Publishing Company, 1996.

[FM15] Florian F. Mueller, Matthew Muirhead. Jogging with a quadcopter. In: Proceedings of CHI'15, ACM, New York, USA, 2015.

[Fog03] B.J. Fogg. Persuasive technology – Using computers to change what we think and do. Elsevier Ltd., Oxford, UK, 2003.

[Fog09] B.J. Fogg. A behavior model for persuasive design. In: Proceedings of the 4th International Conference on Persuasive Technology, ACM, New York, USA, 2009.

[Fre94] Bruno S. Frey. How intrinsic motivation is crowded out and in. In: Rationality and Society, Vol. 6, No. 3, 1994.

[FRO+15] Ty Ferguson, Alex V. Rowlands, Tim Olds, Carol Maher. The validity of consumer-level activity monitors in healthy adults worn in free-living conditions – a cross-sectional study. In: International Journal of Behavioral Nutrition and Physical Activity, Vol. 12, No. 1, 2015.

[FHK+10] B.J. Fogg, Jason Hreha, Robin Krieglstein, Kara Chanasyk, Gaju Krishna. Purple Path Behavior Guide. Stanford Persuasive Tech Lab, USA, 2010.

[GHL05] Alfonso Garate, Nati Herrasti, Antonio Lopez. GENIO – an ambient intelligence application in home automation and entertainment environment. In: Proceedings of the '05 Joint Conference on Smart Objects and Ambient Intelligence, ACM, 2005.

[GHW+10] Stefan Göbel, Sandro Hardy, Viktor Wendel, Florian Mehm, Ralf Steinmetz. Serious games for health – Personalized Exergames. In: Proceedings of the ACM Multimedia '10, ACM, New York, USA, 2010.

[GSH+04] Stefan Göbel, Oliver Schneider, Daniel Holweg, Ursula Kretschmer. GEIST – mobile edutainment with GIS and interactive storytelling. In: Proceedings of the First International Workshop on Ubiquitous GIS, University College of Gävle, 2004.

[GP10] Carlos Gomez, Josep Paradells. Wireless home automation networks – a survey of architectures and technologies. In: IEEE Communications, Vol. 48, No. 4, 2010.

[Gra48] David A. Grant. The latin square principle in the design and analysis of psychological experiments. In: Psychological Bulletin, Vol. 45, No. 5, 1948.

[Han67] Dale L. Hanson. Influence of the Hawthorne effect upon physical education research. In: Research Quarterly of the American Association for Health, Physical Education and Recreation, Vol. 38, No. 4, 1967.

[HBB+01] Joseph Henrich, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, Richard McElreath. In search of Homo Economics – behavioral experiments in 15 small-scale societies. In: The American Economic Review, Vol. 91, No. 2, 2001.

[HGT13] Seyed Amir Hoseini-Tabatabeia, Alexander Gluhak, Rahim Tafazolli. A survey on smartphone-based systems for opportunistic user context recognition. In: ACM Computer Surveys, Vol. 45, No. 3, 2013.

[HHS+02] Andy Harter, Andy Hopper, Pete Steggles, Andy Ward, Paul Webster. The anatomy of a context-aware application. In: Wireless Networks, Vol. 8, No. 2, 2002.

[Hig97] E. Tory Higgins. Beyond pleasure and pain. In: The American Psychologist, Vol. 52, No. 12, 1997.

[Hig11] E. Tory Higgins. Beyond pleasure and pain - how motivation works. Oxford University Press, 2011.

[HLM+07] Steve Hinske, Matthias Lampe, Carsten Magerkurth, Carsten Röcker. Classifying pervasive games – on pervasive computing and mixed reality. In: Concepts and technologies for Pervasive Games – A Reader for Pervasive Gaming Research, 2007.

[HLP+07] William L. Haskell, I-Min Lee, Russell R. Pate, Kenneth E. Powell, Steven N. Blair, Barry A. Franklin, Caroline A. Macera, Gregory W. Heath, Paul D. Thompson, Adrian Bauman. Physical activity and public health – updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. In: Circulation, Vol. 116, No. 9, 2007.

[HLP+12] Inseok Hwang, Youngki Lee, Taiwoo Park, Junehwa Song. Toward a mobile platform for pervasive games. In: Proceedings of MobiGames'12, Helsinki, Finland, 2012.

[HKS14] Juho Hamari, Jonna Koivisto, Harri Sarsa. Does gamification work? – a literature review of empirical studies of gamification. In: Proceedings of the 47th International Conference on System Science, IEEE, 2014.

[HMG06] Ralf Herbrich, Tom Minka, Thore Graepel. TrueSkill – a Bayesian skill rating system. In: Proceedings of Neural Information Processing Systems '06, 2006.

[HSG+11] Sandro Hardy, Abdulmotaleb El Saddik, Stefan Göbel, Ralf Steinmetz. Context aware serious games framework for sport and health. In: Proceedings of Medical Measurements and Applications '11, IEEE, 2011.

[HSS+04] Anne Haase, Andrew Steptoe, James F. Sallis, Jane Wardle. Leisure-time physical activity in university students from 23 countries: Associations with health beliefs, risk awareness, and national economic development. In: Preventive Medicine, Vol. 39, No. 1, 2004.

[Hui49] Johan Huizinga. Homo ludens. Routledge & Kegan Paul, London, UK, 1949.

[HZB+13] Gesine Hinterwälder, Christian T. Zenger, Foteini Baldimtsi, Anna Lysyanskaya, Christof Paar, Wayne P. Burleson. Efficient e-cash in practice – NFC-based payments for public transportation systems. In: Proceedings of the International Symposium on Privacy Enhancing Technologies, Springer, Berlin, Germany, 2013.

[ICS+15] Masa Isakovic, Jaka Cijan, Urban Sedlar, Mojca Volk, Janez Bester. The role of mHealth applications in societal and social challenges of the future. In: 12th International Conference on Information Technology - New Generations (ITNG), IEEE, 2015.

[Jac95] Susan A. Jackson. Factors influencing the occurrence of flow state in elite athletes. In: Journal of Applied Sports Psychology, Vol. 7, No. 2, 1995.

[JD06] Jennifer Jabs, Carol M. Devine. Time scarcity and food choices – an overview. In: Appetite, Vol. 47, No. 2, 2006.

[Joh04] Warren St. John. Quick, after him – Pac-Man went thataway. The New York Times 2014-09-05, New York, USA, 2004.

[JPO+11] Robin R. Johnson, Djordje P. Popovic, Richard E. Olmstead, Maja Stikic, Daniel J. Levendowski, Chris Berka. Drowsiness/alertness algorithm development and validation using synchronized EEG and cognitive performance to individualize a generalized model. In: Biological Psychology. Vol. 87, No. 2, 2011.

[JS13] Klaus P. Jantke, Sebastian Spundflasch. Storyboarding pervasive learning games. In: Proceedings of the International Conference on Advanced ICT for Education, Hainan, China, 2013.

[KRW07] Mirella Kleijnen, Ko de Ruyter, Martin Wetzels. As assessment of value creation in mobile service delivery and the moderating role of time consciousness. In: Journal of Retailing, Vol. 83, No. 1, 2007.

[KG15] Vlasios Kasapakis, Damianos Gavalas. Pervasive gaming – status, trends, and design principles. In: Journal of Network and Computer Applications, Vol. 55, 2015.

[KGS00] Peter T. Katzmarzyk, Norman Gledhill, Roy J. Shephard. The economic burden of physical inactivity in Canada. In: Canadian Medical Association Journal, Vol. 163, No. 11, 2000.

[Kon13] Johannes Konert. Interactive multimedia learning – using social media for peer education in single-player educational games. Doctoral thesis, TU Darmstadt, Germany, 2013.

[KSI14] Jan Kremer, Kim Steenstrup-Pedersen, Christian Igel. Active learning with support vector machines. In: Wiley Interdisciplinary Reviews – Data Mining and Knowledge Discovery, Vol. 4, No. 4, 2014.

[LCL+12] Hosub Lee, Young S. Choi, Sunjae Lee, Ilpyung Park. Towards unobstrusive emotion recognition for affective social communication. In: Proceedings of the 9th Annual IEEE Consumer Communications and Networking Conference, IEEE, 2012.

[LLL+11] Robert Likamwa, Yunxin Liu, Nicholas D. Lane, Lin Zhong. Can your smartphone infer your mood? In: Proceedings of the PhoneSense Workshop, 2011.

[LLL+13] Robert Likamwa, Yunxin Liu, Nicholas D. Lane, Lin Zhong. MoodScope – building a mood sensor from smartphone usage patterns. In: Proceedings of the 11th International Conference on Mobile Systems, Applications, and Services, ACM, New York, USA, 2013.

[LSH+13] Lisbeth H. Larsen, Lone Schou, Henrik H. Lund, Henning Langberg. The physical effect of exergames in healthy elderly – a systematic review. In: Games For Health Journal, Vol. 2, No. 4, 2013.

[Luc14] Victor Luckerson. Activision is reportedly spending half a billion on a single game. TIME Magazine 2014-05-06, New York, USA, 2014.

[LXC13] Yepang Liu, Chang Xu, Shing-Chi Cheung. Where has my battery gone? Finding sensor related energy black holes in smartphone applications. In: Proceedings of the 2013 IEEE International Conference on Pervasive Computing and Communications, IEEE, 2013.

[Mal14] Rainer Malaka. How computer games can improve your health and fitness. In: Proceedings of the International Conference on Serious Games '14, Springer, Berlin, Germany, 2014.

[MHA04] Rainer Malaka, Jochen Häussler, Hidir Aras. SmartKom Mobile – intelligent ubiquitous user interaction. In: Proceedings of the 9th International Conference on Intelligent User Interfaces, ACM, New York, USA, 2004.

[MDH+08] Todd Mytkowicz, Amer Diwan, Matthias Hauswirth, Peter F. Sweeney. We have it easy, but do we have it right? In: Proceedings of the International Parallel and Distributed Processing Symposium '08, IEEE, 2008.

[Meh13] Florian Mehm. Authoring of adaptive single-player educational games. Doctoral thesis, TU Darmstadt, Germany, 2013.

[Mil92] Carolyn R. Miller. Kairos in the rhetoric of science. In: A Rhetoric of Doing – Essays on Written Discourse in Honor of James L. Kinneavy, Southern Illinois University Press, Illinois, USA, 1992.

[MLC+13] Kevin Meehan, Tom Lunney, Kevin Curran, Aiden McCaughey. Context-aware intelligent recommendation system for tourism. In: 2013 IEEE International Conference on Pervasive Computing and Communications. IEEE, 2013.

[MLE+07] Emiliano Miluzzo, Nicholas D. Lane, Shane B. Eisenman, Andrew T. Campbell. CenceMe – injecting sensing presence into social networking. In: Proceedings of the 2nd Conference on Smart Sensing and Context, Springer, Berlin, Germany, 2007.

[MMH+15] Abhinav Mehrotra, Mirco Musolesi, Robert Hendley, Veljko Pejovic. Designing content-driven intelligent notification mechanisms for mobile applications. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, ACM, New York, USA, 2015.

[Mon05] Markus Montola. Exploring the edge of the magic circle – Defining pervasive games. In: Proceedings of Digital Arts and Culture, Copenhagen, Denmark, 2005.

[MS00] Natalie Marmasse, Chris Schmandt. Location-aware information delivery with comMotion. In: Proceedings of the 2nd International Symposium on Handheld and Ubiquitous Computing. Springer, Berlin, Germany, 2000.

[MVS+01] Miguel A. Martinez-Gonzalez, Jose J. Varo, Jose L. Santos, Jokin De Irala, Michael Gibney, John Kearney, J. Alfredo Martinez. Prevalence of physical activity during leisure time in the European Union. In: Medicine and Science in Sports and Exercise, Vol. 33, No. 7, 2001.

[Nie07] Eva Nieuwdorp. The pervasive discourse – an analysis. In: Computers in Entertainment, Vol. 5, No. 2, 2007.

[Nor97] Donald A. Norman. Optimal flow. In: Arts Education Policy Review, Vol. 97, No. 4, 1997.

[OB92] Neville Owen, Adrian Bauman. The descriptive epidemiology of a sedentary lifestyle in adult Australians. In: International Journal of Epidemiology, Vol. 21, No. 2, 1992.

[Opp09] Leif Oppermann. Faciliating the development of location-based experiences. Ph.D. thesis, University of Nottingham, UK, 2009.

[ORN+15] Tadashi Okoshi, Julian Raymos, Hiroki Nozaki, Jin Nakazawa, Anind K. Dey, Hideyuki Tokuda. Reducing users' perceived mental effort due to interruptive notifications in multi-device mobile environments. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, ACM, New York, USA, 2015.

[PAR+14] Linda S. Pescatello, Ross Arena, Debora Riebe, Paul D. Thompson. ACSM's guidelines for exercise testing and prescription. American College of Sports Medicine, Wolters Kluwer, Baltimore, USA, 2014.

[PM14] Veljko Pejovic, Mirco Musolesi. InterruptMe – designing intelligent prompting mechanisms for pervasive applications. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, ACM, 2014.

[PMB99] Michael Pratt, Carol A. Macera, Curtis Blanton. Levels of physical activity and inactivity in children and adults in the United States: current evidence and research issues. In: Medicine and Science in Sports and Exercise, Vol. 31, No. 11, 1999.

[RBC+06] Omer Rashid, Will Bamford, Paul Coulton, Reuben Edwards, Jürgen Scheible. PAC-LAN – mixed-reality gaming with RFID-enabled mobile phones. In: Computers in Entertainment, Vol. 4, No. 4, 2006.

[RD00] Richard M. Ryan, Edward L. Deci. Intrinsic and extrinsic motivations – classic definitions and new directions. In: Contemporary Educational Psychology, Vol. 25, No. 1, 2000.

[Reu16] Christian Reuter. Authoring collaborative multiplayer games – game design patterns, structural verification, collaborative balancing and rapid prototyping. Doctoral thesis, TU Darmstadt, Germany, 2016.

[RN16] Stuart Russell, Peter Norvig. Artificial intelligence – a modern approach. Pearson Education Limited, Harlow, England, 2016.

[ROP+05] Mika Raento, Antti Oulasvirta, Renaud Petit, Hannu Toivonen. ContextPhone – a prototypical platform for context-aware mobile applications. In: IEEE Pervasive Computing, Vol. 4, No. 2, 2005.

[Ros03] Hartmut Rosa. Social acceleration – ethical and political consequences of a desynchronized high-speed society. In: Constellations, Vol. 10, No. 1, 2003.

[SAW95] Bill N. Schilit, Norman Adams, Roy Want. Context-aware computing applications. In: Proceedings of the First Workshop on Mobile Computing Systems and Applications, IEEE, 1995.

[SBG99] Albrecht Schmidt, Michael Beigl, Hans Gellersen. There is more to context than location. In: Computer & Graphics, Vol. 23, No. 6, 1999.

[Set10] Burr Settles. Active learning literature survey. Technical report, University of Wisconsin, USA, 2010.

[SF15] Katie Seaborn, Deborah I. Fels. Gamification in theory and action – a survey. In: International Journal of Human-Computer Studies, Vol. 74, 2015.

[SG12] Ralf Steinmetz, Stefan Göbel. Challenges in serious games as emerging multimedia technology for education, training, sports and health. In: Proceedings of the 18th International Conference on Advances in Multimedia Modeling, Springer, Berlin, Germany, 2012.

[SHK+15] Ralf Steinmetz, Melanie Holloway, Boris Koldehofe, Björn Richerzhagen, Nils Richerzhagen. Towards future Internet communications – role of scalable adaptive mechanisms. In: Proceedings of the 27th Annual Conference on Symbiosis - Synergy of Humans & Technology, Eurographics Association, Goslar, Germany, 2015.

[SN04] Ralf Steinmetz, Klara Nahrstedt. Multimedia applications. Springer, Berlin, Germany, 2004.

[SW05] Ralf Steinmetz, Klaus Wehrle (Eds.). Peer-to-peer systems and applications. Springer, Berlin, Germany, 2015.

[SHM07] Jeff Sinclair, Philip Hingston, Martin Masek. Considerations for the design of exergames. In: Proceedings of the 5th International Conference on Computer Graphics and Interactive Techniques in Australia and Southeast Asia, ACM, New York, USA, 2007.

[SSF+03] Daniel Siewiorek, Asim Smailagic, Junichi Furukawa, Neema Moraveji, Kathryn Reiger, Jeremy Shaffer. SenSay – a context-aware mobile phone. In: Proceedings of the 7th International Symposium on Wearable Computers, IEEE, 2003.

[SSM12] Arun Sahayadhas, Kenneth Sundaraj, Murugappan Murugappan. Detecting driver drowsiness based on sensors – a review. In: Sensors, Vol. 12, No. 12, 2012.

[ST94] Bill N. Schilit, Marvin M. Theimer. Disseminating active map information to mobile hosts. In: IEEE Network, Vol. 8, No. 5, 1994.

[STC+09] Andre C. Santos, Luis Tarrataca, Joao M. Cardoso, Diogo R. Ferreira, Pedro C. Diniz, Paulo Chainho. Context-inference for mobile applications in the UPCASE project. In: Proceedings of the International Conference on Mobile Wireless Middleware, Operating Systems, and Applications '09, Springer, Berlin, Germany, 2009.

[SW05] Penelope Sweetser, Peta Wyeth. GameFlow – A model for evaluating player enjoyment in games. In: Computers in Entertainment, Vol. 3, No. 3, 2005.

[SZ04] Katie Salen, Eric Zimmerman. Rules of play – game design fundaments. MIT Press, USA, 2004.

[TAF+12] Julian F. Thayer, Fredrik Ahs, Mats Fredrikson, John J. Sollers, Tor D. Wager. A meta-analysis of heart rate variability and neuroimaging studies – implications for heart rate variability as a marker of stress and health. In: Neuroscience and Biobehavioral Reviews, Vol. 36, No. 2, 2012.

[TK98] Simon Tong, Daphne Koller. Support vector machine active learning with applications to text classification. In: Proceedings of the 17th International Conference on Machine Learning, Stanford University, USA, 1998.

[TC01] Simon Tong, Edward Chang. Support vector machine active learning for image retrieval. In: Proceedings of the 9th ACM International Conference on Multimedia, ACM, New York, USA, 2001.

[TKV97] Gerard Tertoolen, Dik Van Kreveld, Ben Verstraten. Psychological resistance against attempts to reduce private car use. In: Transportation Research Part A. Vol. 32, No. 3, 1997.

[TOB+02] Stewart G. Trost, Neville Owen, Adrian E. Bauman, James F. Sallis, Wendy Brown. Correlates of adults' participation in physical activity: Review and update. In: Medicine and Science in Sports and Exercise, Vol. 34, No. 12, 2002.

[US08] U.S. Department of Health and Human Services. Physical Activity Guidelines Advisory Committee Report, 2008. Washington, D.C., USA, 2008.

[Waj14] Judy Wajcman. Pressed for time – the acceleration of life in digital capitalism. University of Chicago Press, Chicago, USA, 2014.

[WBE+13] Sabine Webel, Ulrich Bockholt, Timo Engelke, Nirit Gavish, Manuel Olbrich, Carsten Preusche. An augmented reality training platform for assembly and maintenance skills. In: Robotics and Autonomous Systems, Vol. 61, No. 4, 2013.

[Wei91] Mark Weiser. The computer for the 21st century. In: Scientific American, Vol. 256, No. 3, 1991.

[Wen15] Viktor Wendel. Collaborative game-based learning – automatized adaptation mechanisms for game-based collaborative learning using game mastering concepts. Doctoral thesis, TU Darmstadt, Germany, 2015.

[WHF+92] Roy Want, Andy Hoppper, Veronica Falcao, Jonathan Gibbons. The Active Badge location system. In: ACM Transactions on Information Systems, Vol. 10, No. 1, 1992.

[WHO10] World Health Organization. Global recommendations on physical activity for health. WHO Press, Switzerland, 2010.

[WKO+99] Kate Woolf-May, Edward Kearney, Andrew Owen, David W. Jones, Richard C. Davidson, Steve R. Bird. The efficacy of accumulated short bouts versus single daily bout of brisk walking in improving aerobic fitness and blood lipid profile. In: Health Education Research, Vol. 14, No. 6, 1999.

[WMH+09] Nicky Welch, Sarah McNaughton, Wendy Hunter, Clare Hume, David Crawford. Is the perception of time pressure a barrier to healthy eating and physical activity among woman? In: Public Health Nutrition, Vol. 12, No. 7, 2009.

[WR91] Ladd Wheeler, Harry T. Reis. Self-recording of everyday life events – origins, types, and uses. In: Journal of Personality, Vol. 59, No. 3, 1991.

[WRW12] Xinxi Wang, David Rosenblum, Ye Wang. Context-aware mobile music recommendation for daily activities. Proceedings of the 20th ACM International Conference on Multimedia, ACM, New York, USA, 2012.

[WWT+11] Chi Pang Wen, Jackson Pui Man Wai, Min Kuang Tsai, Yi Chen Yang, Ting Yuan David Cheng, Meng-Chih Lee, Hui Ting Chan, Chwen Keng Tsao, Shan Pou Tsai, Xifeng Wu. Minimum amount of physical activity for reduced mortality and extended life expectancy: A prospective cohort study. In: The Lancet, Vol. 378, No. 9798, 2011.

[Yee06] Nick Yee. Motivations for play in online games. In: CyberPsychology & Behavior, Vol. 9, No. 6, 2006.

List of Figures

List of Tables

List of Acronyms

AQ	Accuracy Quote. The relative amount of successful intervention attempts that were made by an interventive measure during a specific time period.	p.30
BFP	Behavior Factors Product function. Specifies the value of the product of the target person's ability and motivation for the desired behavior at an observed point in time. Will assume a value between 0 and 1. Only if the BFP value is larger than the person's activation threshold, then an occurring trigger can successfully initiate the desired behavior. Determines the value of the → ISD function.	p.23
FBM	Fogg Behavior Model. Created by American psychologist B.J. Fogg, explains the three requirements that need to be met in order for a person to show a target behavior (a sufficiently high ability, sufficiently high motivation, and a well-timed trigger). Also see → BFP function.	p.7
ISC	Intervention Success Confidence function. States an observer's confidence in an intervention attempt to be successful in changing the target person's original behavior to the desired behavior at the observed point in time. Aims to approximate the → ISD function.	p.24
ISD	Intervention Success Determination function. States the actual chance of an intervention attempt to successfully change a target person's original behavior to a desired behavior at a specific point in time. Is always either 1 or 0. Dependent on the → BFP function and approximated by the → ISC function.	p.24
MET	Metabolic Equivalent of Tasks. A convention that specifies the intensities of physical activities by assigning 'MET values' to them. The current hierarchy ranges from the least intense physical activity – sleeping – with an associated MET value of 0.9 to the most intense activity – running at 14 mph – with an associated MET value of 23.0. The intensities of all other physical activities lie somewhere in between these two extremes.	p.1
MOQ	Missed Opportunities Quote. Specifies the percentage of triggering opportunities that an interventive measure misses due to a disruption of access to the target person. The MOQ is complementary to the PQ.	p.27
PQ	Pervasiveness Quote. The relative amount of triggering opportunities during which an interventive measure has access to the person whose behavior is supposed to be changed. The PQ is complementary to the MOQ.	p.28
RMR	Resting Metabolic Rate. A special → MET value that specifies the intensity of sitting still. Provides the baseline for defining the relative intensities of other activities.	p.2
TOC	Total Opportunities Counter. The total number of opportunities arising during a specific time period for changing the target person's behavior in the desired way.	p.26
WHO	World Health Organization.	p.1

Appendix A – Evaluation Result Tables

Table 9: Results of the Pre-Study Questionnaire (1/2).

Question			Result		
ID	Text	Options	N	M	SD
pre01	Boy or girl?	2 options (m/f)	28	18 male, 10 female, 2 not reported	
pre02	How old are you?	4 options (low to high)	30	2.03	0.18
pre03	What's your highest qualification (including currently pursuing?)	4 options (low to high)	30	3.77	0.57
pre04	How many hours per week do you spend working or studying?	4 options (low to high)	30	2.40	0.72
pre05	How many hours per week do you do sports or exercise?	4 options (low to high)	30	2.20	0.85
pre06	How many hours per week do you spend playing video games?	4 options (low to high)	30	2.63	1.30
pre07	How many hours per week do you spend watching TV (including Netflix, etc.)?	4 options (low to high)	30	2.73	1.17
pre08	How many hours per week do you spend with other hobbies (excluding TV/PC)?	4 options (low to high)	30	2.67	1.03
pre09	I would like to exercise more, but I simply lack the time.	5 point Likert disagree-agree	30	2.70	1.42
pre10	Honestly: I was never into sports.	5 point Likert disagree-agree	30	2.13	1.46
pre11	I would like to exercise more, but the conditions are not ideal.	5 point Likert disagree-agree	30	3.10	1.30
pre12	If I do sports, I prefer the outside to indoors.	5 point Likert disagree-agree	30	3.33	1.63
pre13	Computers and technical stuff are not my thing.	5 point Likert disagree-agree	30	1.60	1.16
pre14	I'm always carrying my smartphone with me and I take it everywhere.	5 point Likert disagree-agree	30	4.67	0.66
pre15	I occasionally play games on my smartphone.	5 point Likert disagree-agree	30	3.53	1.33
pre16	My life got a lot more hectic during the last ten years.	5 point Likert disagree-agree	30	3.93	1.31
pre17	I could explain the difference between an accelerometer and a gyroscope.	5 point Likert disagree-agree	30	3.37	1.77

(continued on next page)

Table 10: Results of the Pre-Study Questionnaire (2/2).

Question			Result		
ID	Text	Options	N	M	SD
pre18	[How familiar are you with] Twostone	5 options (low to high)	30	1.47	0.82
pre19	[How familiar are you with] Pokemon Go	5 options (low to high)	30	3.30	1.39
pre20	[How familiar are you with] Ingress	5 options (low to high)	30	1.97	0.93
pre21	[How familiar are you with] Zombies, Run!	5 options (low to high)	30	1.60	0.86
pre22	[How familiar are you with] Wii Fit or Wii Sports	5 options (low to high)	30	3.27	0.91
pre23	[How familiar are you with] Xbox Fitness, PlayStation Move Fitness or Zumba	5 options (low to high)	30	1.97	0.67
pre24	[How familiar are you with] Minecraft	5 options (low to high)	30	3.03	1.07
pre25	[How familiar are you with] Grand Theft Auto V	5 options (low to high)	30	2.63	1.19

Table 11: Pre-Study Correlations (1/2).

Strong Correlations				
Item A		Item B		r
pre01	Is female	pre13	Lacks technical interest	+.63
pre04	Works a lot	pre16	Feels stressed	+.65
pre06	Plays a lot of video games	pre13	Lacks technical interest	-.63
pre13	Lacks technical interest	pre17	Has technical knowledge	-.63
Moderate Correlations				
Item A		Item B		r
pre01	Is female	pre06	Plays lots of video games	-.54
pre01	Is female	pre17	Has technical knowledge	-.56

(continued on next page)

Table 12: Pre-Study Correlations (2/2).

Moderate Correlations (cont'd)				
Item A		**Item B**		**r**
pre01	Is female	pre20	Played Ingress	-.43
pre06	Plays a lot of video games	pre17	Has technical knowledge	+.49
pre06	Plays a lot of video games	pre25	Played GTA V	+.42
pre08	Spends time with other hobbies	pre09	Lacks time for sport	-.40
pre09	Lacks time for sport	pre13	Lacks technical interest	+.51
pre09	Lacks time for sport	pre16	Feels stressed	+.55
pre13	Lacks technical interest	pre25	Played GTA V	-.41
pre18	Played Twostone	pre21	Played Zombies, Run!	+.42
pre20	Played Ingress	pre24	Played Minecraft	+.49
Weak Correlations				
Item A		**Item B**		**r**
pre04	Works a lot	pre06	Plays a lot of video games	-.39
pre05	Spends much time doing sport	pre10	Lacks motivation for sport	-.39
pre06	Plays a lot of video games	pre20	Played Ingress	+.36
pre13	Lacks technical interest	pre23	Played PlayStation Move	+.38
pre18	Played Twostone	pre20	Played Ingress	+.38
pre20	Played Ingress	pre21	Played Zombies, Run!	+.37

Table lists only statistically relevant correlations ($p<.05$)
Correlation categories according to [Eva96]

Table 13: Results of the Post-Study Questionnaire (1/3).

Question			Result		
ID	**Text**	**Options**	**N**	**M**	**SD**
pos01	Boy or girl?	2 options (m/f)	25	16 male, 9 female, 2 not reported	
pos02	How old are you?	Ordinal	27	26.41	7.28
pos03	What name did appear in your Trigger App?	3 options	27	9 Babbage, 10 Neumann, 8 Turing	
pos04	I have a lot of video gaming experience.	5 point Likert disagree-agree	27	3.56	1.45
pos05	I'm a sportsperson.	5 point Likert disagree-agree	27	3.11	1.50
pos06	I'm kind of a geek and love to have new technical stuff.	5 point Likert disagree-agree	27	3.70	1.20
pos07	I'm extroverted.	5 point Likert disagree-agree	27	3.00	0.96
pos08	I'm very busy at the moment (with work, exams, etc.).	5 point Likert disagree-agree	27	4.07	1.24
pos09	I try to eat healthy.	5 point Likert disagree-agree	27	3.74	1.10
pos10	I like being outside.	5 point Likert disagree-agree	27	3.74	1.06
pos11	I dislike sweating.	5 point Likert disagree-agree	27	3.22	1.53
pos12	I always have my smartphone with me.	5 point Likert disagree-agree	27	4.59	0.50
pos13	I have enough time for sport.	5 point Likert disagree-agree	27	3.15	1.38
pos14	I like doing sport.	5 point Likert disagree-agree	27	4.04	1.06
pos15	I like to go running.	5 point Likert disagree-agree	27	2.59	1.67
pos16	Sport is a bit complicated for me at the moment (driving to sports range, etc.).	5 point Likert disagree-agree	27	2.96	1.29
pos17	I prefer team sports.	5 point Likert disagree-agree	27	2.63	1.62
pos18	I am a competitive person.	5 point Likert disagree-agree	27	3.00	1.41

(continued on next page)

Table 14: Results of the Post-Study Questionnaire (2/3).

Question			Result		
ID	**Text**	**Options**	**N**	**M**	**SD**
pos19	I should probably do more sport.	5 point Likert disagree-agree	27	3.56	1.22
pos20	I enjoy playing Twostone.	5 point Likert disagree-agree	27	2.48	1.19
pos21	I will keep playing Twostone.	5 point Likert disagree-agree	27	2.07	1.07
pos22	I think Twostone can help me to stay active and healthy.	5 point Likert disagree-agree	27	2.26	1.26
pos23	I think Twostone could help others to stay active and healthy.	5 point Likert disagree-agree	27	3.70	0.95
pos24	Twostone is for kids.	5 point Likert disagree-agree	27	3.41	1.19
pos25	Twostone is for adults.	5 point Likert disagree-agree	27	3.41	0.97
pos26	Twostone is for seniors.	5 point Likert disagree-agree	27	2.26	0.94
pos27	I like playing my own maps.	5 point Likert disagree-agree	27	3.22	1.28
pos28	I like playing maps that were made by others.	5 point Likert disagree-agree	27	2.81	1.11
pos29	Twostone needs to be improved in regard to usability and bugs.	5 point Likert disagree-agree	27	3.67	1.00
pos30	I would probably play a game like Twostone if it was of higher quality.	5 point Likert disagree-agree	27	2.85	1.35
pos31	Playing Twostone is too complicated.	5 point Likert disagree-agree	27	3.11	1.25
pos32	The trigger was annoying.	5 point Likert disagree-agree	27	3.89	1.22
pos33	The trigger was more annoying during the first week.	5 point Likert disagree-agree	27	2.63	1.42
pos34	The trigger selected good opportunities for playing.	5 point Likert disagree-agree	27	2.37	1.11
pos35	The trigger improved over time.	5 point Likert disagree-agree	27	2.96	1.16
pos36	I sometimes did not notice the trigger.	5 point Likert disagree-agree	27	3.22	1.53

(continued on next page)

Table 15: Results of the Post-Study Questionnaire (3/3).

Question			Result		
ID	**Text**	**Options**	**N**	**M**	**SD**
pos37	I sometimes ignored the trigger.	5 point Likert disagree-agree	27	3.74	1.46
pos38	Without the trigger, I would have played less Twostone.	5 point Likert disagree-agree	27	2.59	1.45
pos39	The trigger needs to be more intelligent.	5 point Likert disagree-agree	27	3.70	0.87
pos40	The trigger needs to be improved in regard to usability and bugs.	5 point Likert disagree-agree	27	3.37	1.08

Table 16: Post-Study Correlations (1/6).

Very Strong Correlations				
Item A		**Item B**		**r**
pos05	Has athletic self-perception	pos14	Has sport motivation	+.80
Strong Correlations				
Item A		**Item B**		**r**
pos05	Has athletic self-perception	pos19	Feels lack of exercise	-.69
pos10	Has outdoor affinity	pos15	Enjoys running	+.63
pos13	Has time for sport	pos16	Lacks ability for sport	-.67
pos20	Enjoys Twostone	pos21	Is strongly motivated by Twostone	+.73
pos20	Enjoys Twostone	pos22	Thinks Twostone is good for self	+.66
pos20	Enjoys Twostone	pos27	Enjoys playing own maps	+.63
pos20	Enjoys Twostone	pos38	Reports successful interventions	+.63*
pos23	Thinks Twostone is good for others	pos24	Thinks Twostone is for kids	+.66
pos26	Thinks Twostone is for seniors	pos33	Got used to trigger	+.71†

(continued on next page)

Table 17: Post-Study Correlations (2/6).

Strong Correlations (cont'd)				
Item A		Item B		r
pos34	Thinks triggers were reasonable	pos35	Thinks trigger adapted over time	+.64†
pos34	Thinks triggers were reasonable	pos38	Reports successful interventions	+.62†
Moderate Correlations				
Item A		**Item B**		**r**
pos01	Is female	pos04	Has video game expertise	-.44
pos01	Is female	pos06	Has technical interest	-.49
pos04	Has video game expertise	pos10	Has outdoor affinity	-.48
pos05	Has athletic self-perception	pos11	Has sweating aversion	-.48
pos05	Has athletic self-perception	pos13	Has time for sport	+.57
pos05	Has athletic self-perception	pos16	Lacks ability for sport	-.46
pos05	Has athletic self-perception	pos17	Prefers team sports	+.44
pos05	Has athletic self-perception	pos18	Has competitive character	+.43
pos05	Has athletic self-perception	pos23	Thinks Twostone is good for others	+.43
pos08	Feels stressed	pos13	Has time for sport	-.43
pos08	Feels stressed	pos32	Was annoyed by trigger	+.44*
pos09	Has healthy lifestyle	pos21	Is strongly motivated by Twostone	+.41
pos09	Has healthy lifestyle	pos22	Thinks Twostone is good for self	+.41
pos09	Has healthy lifestyle	pos23	Thinks Twostone is good for others	+.48
pos09	Has healthy lifestyle	pos26	Thinks Twostone is for seniors	+.44

(continued on next page)

Table 18: Post-Study Correlations (3/6).

Moderate Correlations (cont'd)				
Item A		**Item B**		**r**
pos10	Has outdoor affinity	pos11	Has sweating aversion	-.43
pos10	Has outdoor affinity	pos34	Thinks triggers were reasonable	+.41[†]
pos11	Has sweating aversion	pos14	Has sport motivation	-.41
pos11	Has sweating aversion	pos16	Lacks ability for sport	+.44
pos11	Has sweating aversion	pos28	Enjoys playing maps by others	+.43
pos13	Has time for sport	pos14	Has sport motivation	+.55
pos13	Has time for sport	pos19	Feels lack of exercise	-.55
pos13	Has time for sport	pos30	Has interest in quality exergame	+.40
pos14	Has sport motivation	pos15	Enjoys running	+.40
pos14	Has sport motivation	pos16	Lacks ability for sport	-.42
pos14	Has sport motivation	pos17	Prefers team sports	+.46
pos14	Has sport motivation	pos19	Feels lack of exercise	-.58
pos15	Enjoys running	pos16	Lacks ability for sport	-.44
pos15	Enjoys running	pos20	Enjoys Twostone	+.49
pos15	Enjoys running	pos34	Thinks triggers were reasonable	+.50[†]
pos15	Enjoys running	pos38	Reports successful interventions	+.49[*]
pos16	Lacks ability for sport	pos19	Feels lack of exercise	+.55
pos17	Prefers team sports	pos18	Has competitive character	+.59

(continued on next page)

Table 19: Post-Study Correlations (4/6).

Moderate Correlations (cont'd)				
Item A		Item B		r
pos19	Feels lack of exercise	pos38	Reports successful interventions	-.41*
pos20	Enjoys Twostone	pos23	Thinks Twostone is good for others	+.40
pos20	Enjoys Twostone	pos24	Thinks Twostone is for kids	+.54
pos20	Enjoys Twostone	pos26	Thinks Twostone is for seniors	+.40
pos20	Enjoys Twostone	pos31	Thinks Twostone is complicated	-.45
pos21	Is strongly motivated by Twostone	pos22	Thinks Twostone is good for self	+.58
pos21	Is strongly motivated by Twostone	pos23	Thinks Twostone is good for others	+.55
pos21	Is strongly motivated by Twostone	pos24	Thinks Twostone is for kids	+.58
pos21	Is strongly motivated by Twostone	pos26	Thinks Twostone is for seniors	+.51
pos21	Is strongly motivated by Twostone	pos27	Enjoys playing own maps	+.55
pos21	Is strongly motivated by Twostone	pos31	Thinks Twostone is complicated	-.58
pos22	Thinks Twostone is good for self	pos23	Thinks Twostone is good for others	+.55
pos22	Thinks Twostone is good for self	pos26	Thinks Twostone is for seniors	+.52
pos22	Thinks Twostone is good for self	pos31	Thinks Twostone is complicated	-.58
pos22	Thinks Twostone is good for self	pos33	Got used to trigger	+.51†
pos22	Thinks Twostone is good for self	pos34	Thinks triggers were reasonable	+.42†
pos22	Thinks Twostone is good for self	pos38	Reports successful interventions	+.48*
pos23	Thinks Twostone is good for others	pos26	Thinks Twostone is for seniors	+.47

(continued on next page)

Table 20: Post-Study Correlations (5/6).

Moderate Correlations (cont'd)				
Item A		**Item B**		**r**
pos23	Thinks Twostone is good for others	pos31	Thinks Twostone is complicated	-.58
pos23	Thinks Twostone is good for others	pos33	Got used to trigger	$+.57^{\dagger}$
pos23	Thinks Twostone is good for others	pos34	Thinks triggers were reasonable	$+.43^{\dagger}$
pos23	Thinks Twostone is good for others	pos35	Thinks trigger adapted over time	$+.55^{\dagger}$
pos23	Thinks Twostone is good for others	pos38	Reports successful interventions	$+.41^{*}$
pos24	Thinks Twostone is for kids	pos26	Thinks Twostone is for seniors	+.45
pos25	Thinks Twostone is for kids	pos27	Enjoys playing own maps	+.42
pos24	Thinks Twostone is for kids	pos33	Got used to trigger	$+.46^{\dagger}$
pos24	Thinks Twostone is for kids	pos39	Thinks trigger lacks intelligence	$+.53^{\dagger}$
pos26	Thinks Twostone is for seniors	pos28	Enjoys maps made by others	+.52
pos26	Thinks Twostone is for seniors	pos40	Thinks Trigger lacks polish	$+.47^{*}$
pos27	Enjoys playing own maps	pos31	Thinks Twostone is complicated	-.52
pos29	Thinks Twostone lacks polish	pos30	Has interest in quality exergame	+.45
pos29	Thinks Twostone lacks polish	pos31	Thinks Twostone is complicated	-.49
pos29	Thinks Twostone lacks polish	pos39	Thinks Trigger lacks intelligence	$+.41^{\dagger}$
pos29	Thinks Twostone lacks polish	pos40	Thinks Trigger lacks polish	$+.40^{*}$
pos30	Has interest in quality exergame	pos31	Thinks Twostone is complicated	-.45
pos31	Thinks Twostone is complicated	pos35	Thinks trigger adapted over time	$-.47^{\dagger}$

(continued on next page)

Table 21: Post-Study Correlations (6/6).

Moderate Correlations (cont'd)				
Item A		Item B		r
pos33	Got used to trigger	pos34	Thinks triggers were reasonable	+.43†
pos33	Got used to trigger	pos35	Thinks trigger adapted over time	+.55†
pos33	Got used to trigger	pos38	Reports successful interventions	+.49†
pos34	Thinks triggers were reasonable	pos39	Thinks trigger lacks intelligence	-.48†
pos35	Thinks trigger adapted over time	pos38	Reports successful interventions	+.47†
Weak Correlations				
Item A		**Item B**		**r**
pos04	Has video game expertise	pos06	Has technical interest	+.38
pos05	Has athletic self-perception	pos38	Reports successful interventions	+.39
pos06	Has technical interest	pos27	Enjoys playing own maps	+.39
pos08	Feels stressed	pos37	Ignored trigger	+.39*
pos11	Has sweating aversion	pos13	Has sufficient time for sport	-.38
pos15	Enjoys running	pos22	Thinks Twostone is good for self	+.38
pos16	Lacks ability for sport	pos33	Got used to trigger	-.39†
pos16	Lacks ability for sport	pos34	Thinks triggers were reasonable	-.39†
pos20	Enjoys Twostone	pos33	Got used to trigger	+.38†
pos22	Thinks Twostone is good for self	pos24	Thinks Twostone is for kids	+.39

Table lists only statistically relevant correlations ($p<.05$)
Correlation categories according to [Eva96]

* Denotes correlation coefficients with somewhat limited expressiveness due to the aggregation of three trigger groups (concerns pos32, pos37, pos38, pos40)

† Denotes correlation coefficients with highly limited expressiveness due to the aggregation of three trigger groups (concerns pos33, pos34, pos35, pos39)

Table 22: Results of the Post-Study Questionnaire – Babbage only (1/3).

Question			Result		
ID	**Text**	**Options**	**N**	**M**	**SD**
pos01	Boy or girl?	2 options (m/f)	8	3 male, 5 female, 1 not reported	
pos02	How old are you?	Ordinal	9	24.22	3.87
pos03	What name did appear in your Trigger App?	3 options	9	9 Babbage	
pos04	I have a lot of video gaming experience.	5 point Likert disagree-agree	9	3.56	1.24
pos05	I'm a sportsperson.	5 point Likert disagree-agree	9	3.33	1.50
pos06	I'm kind of a geek and love to have new technical stuff.	5 point Likert disagree-agree	9	3.33	1.50
pos07	I'm extroverted.	5 point Likert disagree-agree	9	2.56	1.01
pos08	I'm very busy at the moment (with work, exams, etc.).	5 point Likert disagree-agree	9	3.67	1.32
pos09	I try to eat healthy.	5 point Likert disagree-agree	9	3.56	1.33
pos10	I like being outside.	5 point Likert disagree-agree	9	3.89	1.05
pos11	I dislike sweating.	5 point Likert disagree-agree	9	2.78	1.56
pos12	I always have my smartphone with me.	5 point Likert disagree-agree	9	4.78	0.44
pos13	I have enough time for sport.	5 point Likert disagree-agree	9	3.67	1.00
pos14	I like doing sport.	5 point Likert disagree-agree	9	4.22	0.97
pos15	I like to go running.	5 point Likert disagree-agree	9	2.56	1.74
pos16	Sport is a bit complicated for me at the moment (driving to sports range, etc.).	5 point Likert disagree-agree	9	2.78	1.30
pos17	I prefer team sports.	5 point Likert disagree-agree	9	2.67	1.87
pos18	I am a competitive person.	5 point Likert disagree-agree	9	3.22	1.30

(continued on next page)

Table 23: Results of the Post-Study Questionnaire – Babbage only (2/3).

Question			Result		
ID	Text	Options	N	M	SD
pos19	I should probably do more sport.	5 point Likert disagree-agree	9	3.44	1.24
pos20	I enjoy playing Twostone.	5 point Likert disagree-agree	9	2.67	1.50
pos21	I will keep playing Twostone.	5 point Likert disagree-agree	9	2.22	1.09
pos22	I think Twostone can help me to stay active and healthy.	5 point Likert disagree-agree	9	2.11	1.27
pos23	I think Twostone could help others to stay active and healthy.	5 point Likert disagree-agree	9	3.33	1.12
pos24	Twostone is for kids.	5 point Likert disagree-agree	9	3.33	1.22
pos25	Twostone is for adults.	5 point Likert disagree-agree	9	3.11	1.05
pos26	Twostone is for seniors.	5 point Likert disagree-agree	9	2.44	1.01
pos27	I like playing my own maps.	5 point Likert disagree-agree	9	3.11	1.54
pos28	I like playing maps that were made by others.	5 point Likert disagree-agree	9	2.78	1.20
pos29	Twostone needs to be improved in regard to usability and bugs.	5 point Likert disagree-agree	9	3.67	0.87
pos30	I would probably play a game like Twostone if it was of higher quality.	5 point Likert disagree-agree	9	2.89	1.27
pos31	Playing Twostone is too complicated.	5 point Likert disagree-agree	9	3.56	1.13
pos32	The trigger was annoying.	5 point Likert disagree-agree	9	4.00	1.22
pos33	The trigger was more annoying during the first week.	5 point Likert disagree-agree	9	2.44	1.24
pos34	The trigger selected good opportunities for playing.	5 point Likert disagree-agree	9	2.33	1.22
pos35	The trigger improved over time.	5 point Likert disagree-agree	9	2.67	1.12
pos36	I sometimes did not notice the trigger.	5 point Likert disagree-agree	9	3.56	1.51

(continued on next page)

Table 24: Results of the Post-Study Questionnaire – Babbage only (3/3).

Question			Result		
ID	Text	Options	N	M	SD
pos37	I sometimes ignored the trigger.	5 point Likert disagree-agree	9	3.56	1.74
pos38	Without the trigger, I would have played less Twostone.	5 point Likert disagree-agree	9	2.89	1.54
pos39	The trigger needs to be more intelligent.	5 point Likert disagree-agree	9	3.67	0.87
pos40	The trigger needs to be improved in regard to usability and bugs.	5 point Likert disagree-agree	9	3.00	0.71

Table 25: Results of the Post-Study Questionnaire – Neumann only (1/3).

Question			Result		
ID	Text	Options	N	M	SD
pos01	Boy or girl?	2 options (m/f)	10	7 male, 3 female	
pos02	How old are you?	Ordinal	10	25.10	2.47
pos03	What name did appear in your Trigger App?	3 options	10	10 Neumann	
pos04	I have a lot of video gaming experience.	5 point Likert disagree-agree	10	3.40	1.71
pos05	I'm a sportsperson.	5 point Likert disagree-agree	10	2.90	1.73
pos06	I'm kind of a geek and love to have new technical stuff.	5 point Likert disagree-agree	10	3.70	0.95
pos07	I'm extroverted.	5 point Likert disagree-agree	10	3.30	0.95
pos08	I'm very busy at the moment (with work, exams, etc.).	5 point Likert disagree-agree	10	4.40	1.07
pos09	I try to eat healthy.	5 point Likert disagree-agree	10	4.00	0.67
pos10	I like being outside.	5 point Likert disagree-agree	10	3.60	1.07
pos11	I dislike sweating.	5 point Likert disagree-agree	10	3.10	1.66

(continued on next page)

Table 26: Results of the Post-Study Questionnaire – Neumann only (2/3).

Question			Result		
ID	**Text**	**Options**	**N**	**M**	**SD**
pos12	I always have my smartphone with me.	5 point Likert disagree-agree	10	4.50	0.53
pos13	I have enough time for sport.	5 point Likert disagree-agree	10	3.20	1.55
pos14	I like doing sport.	5 point Likert disagree-agree	10	4.00	1.25
pos15	I like to go running.	5 point Likert disagree-agree	10	3.10	1.66
pos16	Sport is a bit complicated for me at the moment (driving to sports range, etc.).	5 point Likert disagree-agree	10	2.60	1.07
pos17	I prefer team sports.	5 point Likert disagree-agree	10	2.20	1.14
pos18	I am a competitive person.	5 point Likert disagree-agree	10	2.70	1.49
pos19	I should probably do more sport.	5 point Likert disagree-agree	10	3.20	1.48
pos20	I enjoy playing Twostone.	5 point Likert disagree-agree	10	2.60	1.17
pos21	I will keep playing Twostone.	5 point Likert disagree-agree	10	2.00	0.82
pos22	I think Twostone can help me to stay active and healthy.	5 point Likert disagree-agree	10	2.50	1.18
pos23	I think Twostone could help others to stay active and healthy.	5 point Likert disagree-agree	10	3.90	0.74
pos24	Twostone is for kids.	5 point Likert disagree-agree	10	3.50	1.18
pos25	Twostone is for adults.	5 point Likert disagree-agree	10	3.30	1.06
pos26	Twostone is for seniors.	5 point Likert disagree-agree	10	2.30	1.06
pos27	I like playing my own maps.	5 point Likert disagree-agree	10	2.90	1.10
pos28	I like playing maps that were made by others.	5 point Likert disagree-agree	10	2.60	1.17
pos29	Twostone needs to be improved in regard to usability and bugs.	5 point Likert disagree-agree	10	3.60	1.07

(continued on next page)

Table 27: Results of the Post-Study Questionnaire – Neumann only (3/3).

Question			Result		
ID	Text	Options	N	M	SD
pos30	I would probably play a game like Twostone if it was of higher quality.	5 point Likert disagree-agree	10	2.90	1.45
pos31	Playing Twostone is too complicated.	5 point Likert disagree-agree	10	3.00	1.25
pos32	The trigger was annoying.	5 point Likert disagree-agree	10	3.40	1.35
pos33	The trigger was more annoying during the first week.	5 point Likert disagree-agree	10	2.90	1.66
pos34	The trigger selected good opportunities for playing.	5 point Likert disagree-agree	10	2.90	1.10
pos35	The trigger improved over time.	5 point Likert disagree-agree	10	3.30	1.25
pos36	I sometimes did not notice the trigger.	5 point Likert disagree-agree	10	3.30	1.64
pos37	I sometimes ignored the trigger.	5 point Likert disagree-agree	10	3.80	1.40
pos38	Without the trigger, I would have played less Twostone.	5 point Likert disagree-agree	10	3.10	1.45
pos39	The trigger needs to be more intelligent.	5 point Likert disagree-agree	10	3.50	0.97
pos40	The trigger needs to be improved in regard to usability and bugs.	5 point Likert disagree-agree	10	3.70	1.25

Table 28: Results of the Post-Study Questionnaire – Turing only (1/3).

Question			Result		
ID	Text	Options	N	M	SD
pos01	Boy or girl?	2 options (m/f)	7	7 male, 1 not reported	
pos02	How old are you?	Ordinal	8	30.50	12.00
pos03	What name did appear in your Trigger App?	3 options	8	8 Turing	
pos04	I have a lot of video gaming experience.	5 point Likert disagree-agree	8	3.75	1.49

(continued on next page)

Table 29: Results of the Post-Study Questionnaire – Turing only (2/3).

Question			Result		
ID	Text	Options	N	M	SD
pos05	I'm a sportsperson.	5 point Likert disagree-agree	8	3.13	1.36
pos06	I'm kind of a geek and love to have new technical stuff.	5 point Likert disagree-agree	8	4.13	1.13
pos07	I'm extroverted.	5 point Likert disagree-agree	8	3.13	0.83
pos08	I'm very busy at the moment (with work, exams, etc.).	5 point Likert disagree-agree	8	4.13	1.36
pos09	I try to eat healthy.	5 point Likert disagree-agree	8	3.63	1.30
pos10	I like being outside.	5 point Likert disagree-agree	8	3.75	1.16
pos11	I dislike sweating.	5 point Likert disagree-agree	8	3.88	1.25
pos12	I always have my smartphone with me.	5 point Likert disagree-agree	8	4.50	0.53
pos13	I have enough time for sport.	5 point Likert disagree-agree	8	2.50	1.41
pos14	I like doing sport.	5 point Likert disagree-agree	8	3.88	0.99
pos15	I like to go running.	5 point Likert disagree-agree	8	2.00	1.60
pos16	Sport is a bit complicated for me at the moment (driving to sports range, etc.).	5 point Likert disagree-agree	8	3.63	1.41
pos17	I prefer team sports.	5 point Likert disagree-agree	8	3.13	1.89
pos18	I am a competitive person.	5 point Likert disagree-agree	8	3.13	1.55
pos19	I should probably do more sport.	5 point Likert disagree-agree	8	4.13	0.64
pos20	I enjoy playing Twostone.	5 point Likert disagree-agree	8	2.13	0.83
pos21	I will keep playing Twostone.	5 point Likert disagree-agree	8	2.00	1.41
pos22	I think Twostone can help me to stay active and healthy.	5 point Likert disagree-agree	8	2.13	1.46

(continued on next page)

Table 30: Results of the Post-Study Questionnaire – Turing only (3/3).

Question			Result		
ID	Text	Options	N	M	SD
pos23	I think Twostone could help others to stay active and healthy.	5 point Likert disagree-agree	8	3.88	0.99
pos24	Twostone is for kids.	5 point Likert disagree-agree	8	3.38	1.30
pos25	Twostone is for adults.	5 point Likert disagree-agree	8	3.88	0.64
pos26	Twostone is for seniors.	5 point Likert disagree-agree	8	2.00	0.76
pos27	I like playing my own maps.	5 point Likert disagree-agree	8	3.75	1.16
pos28	I like playing maps that were made by others.	5 point Likert disagree-agree	8	3.13	0.99
pos29	Twostone needs to be improved in regard to usability and bugs.	5 point Likert disagree-agree	8	3.75	1.16
pos30	I would probably play a game like Twostone if it was of higher quality.	5 point Likert disagree-agree	8	2.75	1.49
pos31	Playing Twostone is too complicated.	5 point Likert disagree-agree	8	2.75	1.39
pos32	The trigger was annoying.	5 point Likert disagree-agree	8	4.38	0.92
pos33	The trigger was more annoying during the first week.	5 point Likert disagree-agree	8	2.50	1.41
pos34	The trigger selected good opportunities for playing.	5 point Likert disagree-agree	8	1.75	0.71
pos35	The trigger improved over time.	5 point Likert disagree-agree	8	2.88	1.13
pos36	I sometimes did not notice the trigger.	5 point Likert disagree-agree	8	2.75	1.49
pos37	I sometimes ignored the trigger.	5 point Likert disagree-agree	8	3.88	1.36
pos38	Without the trigger, I would have played less Twostone.	5 point Likert disagree-agree	8	1.63	0.92
pos39	The trigger needs to be more intelligent.	5 point Likert disagree-agree	8	4.00	0.76
pos40	The trigger needs to be improved in regard to usability and bugs.	5 point Likert disagree-agree	8	3.38	1.19

Appendix B – Evaluation Instructions

Request for Participation

Hi and thanks for taking an interest in our study. :)

Knowledge about the connection between physical activity and health is widespread. Still, more than a third of the European populace is not being sufficiently active. And this, although a mere 15 minutes of activity per day can increase one's life expectancy for up to three years, a goal achievable without having to shed even a single drop of sweat.

For some time now there has been an ongoing discussion on whether location-based games such as *Ingress* or *Pokemon Go* have the potential of helping people to achieve the recommended minimum amount of daily physical activity. We have developed our own version of such a game: *Twostone*.

In a three week study that starts on Monday, the 15th of August and goes until Sunday, the 4th of September, we would like to find out what effects this game and its "intelligent user trigger" really have. Participants of the evaluation should own an Android-based smartphone, ideally running at least Android OS version 5.0. All participants are required to fill two short questionnaires, one at the beginning of the three week study and one at the end. Additionally, they are asked to fully initialize the game once prior to the evaluation's start. All other steps are fully optional. As a small token of appreciation, all participants are handed two movie vouchers for any Kinopolis cinema at the end of the evaluation – as well as that warm feeling of having contributed to mankind's scientific advancement. ;)

If you want to support us, please send a short mail until Friday the 12th of August to this email-address:

twostonestudy@gmail.com

Simply write your name and what kind of smartphone you own. On Saturday, the 13th, we will send two documents to all participants: A manual on how to install and initialize the game, as well as the first of a total of two questionnaires.

Thanks a bunch!
Tim, Jens, Chris, Gerhard & Tobi

TECHNISCHE UNIVERSITÄT DARMSTADT

KOM – Multimedia Communications Lab

KINOPOLIS

Manual I

Thank you for your participation in our study. You will need an Android smartphone. If you have multiple of such devices at your disposal, please select the one that you carry with you the most.

This evaluation is about physical activity. Based on long and extensive studies, scientists found that a mere 15 minutes of medium-intensity physical activity per day will increase one's life expectancy by up to three years. Every additional minute of activity adds to the positive effect on one's health. But although the knowledge about the connection between physical activity and health is fairly widespread, about a third of the European populace is still not sufficiently active. Many types of diseases are the consequence: Hypertension, diabetes, and even cancer. A multitude of different and in part contradictory theories exist, why people will not move, even against their better knowledge, but there is one thing that everybody can agree on: People won't do things that they don't enjoy. For this reason, scientists are looking into ways of how physical activity can be made a little more entertaining. One approach lies in the so-called exergames, sometimes also referred to as fitness games. Such games require the player to move while she is playing them. Many of you probably remember this concept from the *Wii* and in recent years, multiple exergames have also been released for smartphones. The insanely successful *Pokemon Go* is just one example.

At the TU Darmstadt, we have developed our own version of such a mobile exergames, the Android-based game *Twostone*. In this study, we would like to investigate what kind of effects this game has. Alas, we cannot tell you what exactly we are looking for, as this is likely to influence your behavior. And we need you to act all natural, as if there was nobody watching.

Speaking of "nobody watching": Please be informed that we are tracking all kinds of information about you. This ranges from your location over the speed with which you are moving to your smartphone's battery level and how loud it is in your vicinity (note that we are just monitoring the decibel level and not recording noise or speech). And this is something that we are doing not just while you are playing, but regularly. This may seem a little off-putting at first, but please be assured that preserving your privacy is our utmost concern. To begin with, we are utilizing your data only on an abstracted level. For instance, instead of assessing "*she is in a train*", we will rather abstract this information to "s*he is in a vehicle*". More importantly, however, we will ensure that any link between your identify and your data is being removed. Among other things, we will delete all your emails after the study has ended, we will remove all your data from our servers, and we will erase the list of study participants. Not fishing for compliments here, but we are sure that neither Google nor Facebook would do such things. If you are still concerned about your privacy, please let us know.

One more thing: Please bear with us. We are a small team and although we have been working hard on both Twostone and the reminder application (we will get back to this one later), there will be crashes and other problems. In addition, your battery may drain a little faster than what you are used to. If you encounter such problems, try to ignore them. However, if things don't seem to work at all, just let us know via twostonestudy@gmail.com and we will try to come up with a solution as fast as possible.

Enjoy!
Tim, Jens, Chris, Gerhard & Tobi

Study Plan

A few words regarding our study planning. Today, on Saturday the 13th of August, you should have received two documents: This manual and in addition to it, a first questionnaire. Please follow the instructions in this manual, fill out the questionnaire and send it back to us. Here is what is going to happen next:

- Sunday, the 14th of August: You will receive another manual. Please work through this second manual as soon as possible, but DO NOT start with the second manual before you haven't finished this first one.

- Monday, the 15th of August: The actual beginning of the study. It will last 20 days, until Saturday the 3rd of September. Your only task during this time is to keep the reminder application running (more on that tomorrow) and to play Twostone whenever you feel like it.

- Sunday, the 4th of September: We are going to send to you a second questionnaire. Please also fill this second questionnaire as soon as possible and then send it back to us. As a small token of appreciation, everyone who completed the study and returned the second questionnaire will be handed two Kinopolis movie vouchers. To obtain them, simply get in touch with the person who asked you to participate in this study.

- You can have a look at our results! If you take an interest in what we have found (and what we may publish in a scientific paper), simply let us know. We well then send you a document with all our findings as soon as we are done with our analysis.

Installation and Initialization of Twostone

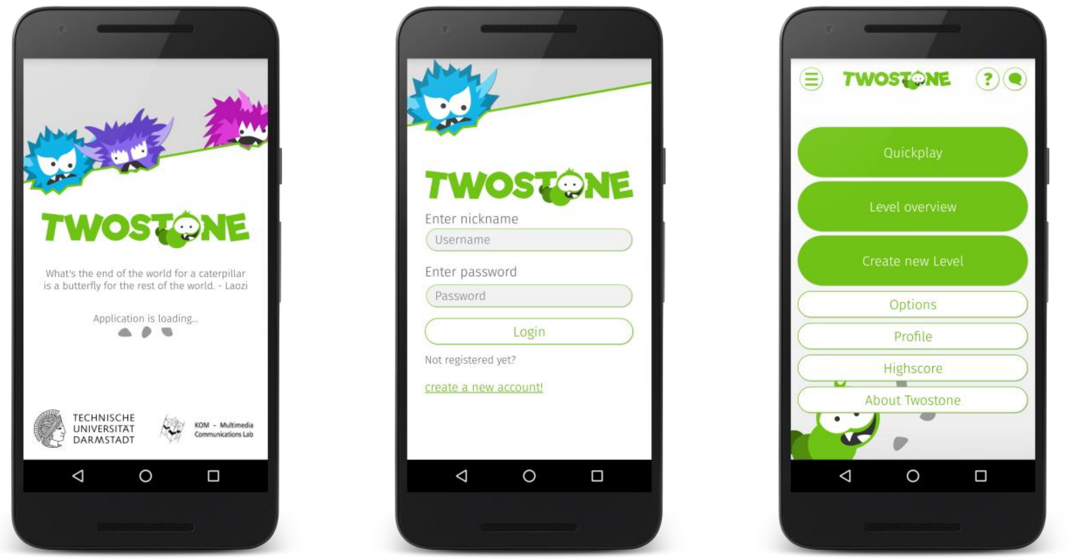
Today's manual is about installing and initializing Twostone. The second part of the manual, which we are going to send you tomorrow, is about another application that we need you to install. Please follow the instructions below and in the second manual in their exact order and without skipping any.

1. If you already had Twostone installed on your device: Please delete it.

2. Install the latest version of Twostone. To do this, either follow the link below on our smartphone, or go to the Google Play Store application on your phone and look for „Twostone".

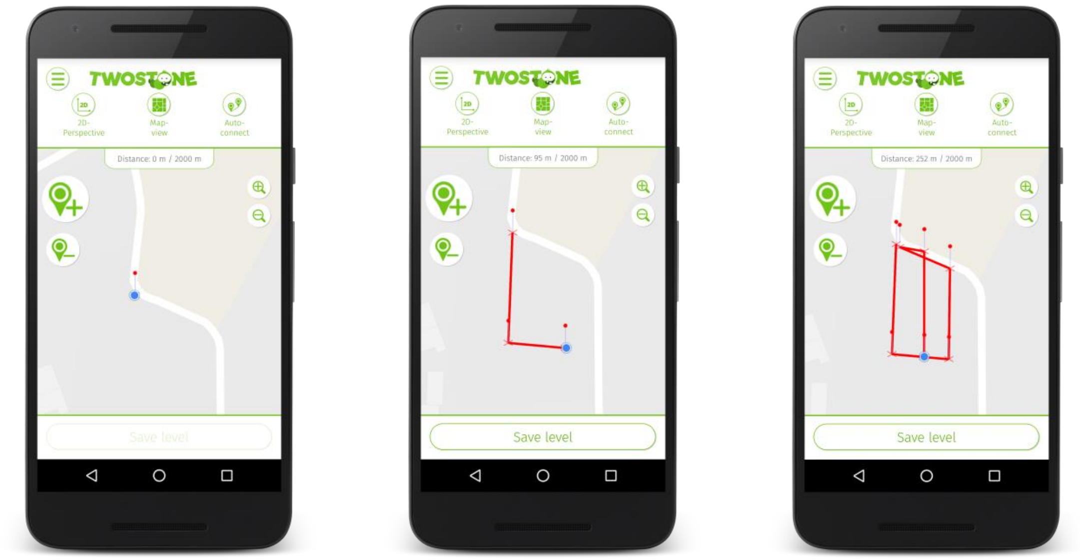
   https://play.google.com/store/apps/details?id=de.tu.darmstadt.uhg

3. At the first start you will have to register. The corresponding link is right beneath the Login-button. Please note that your user name will be case sensitive. If you already had an account, please create a new one specifically for this study.

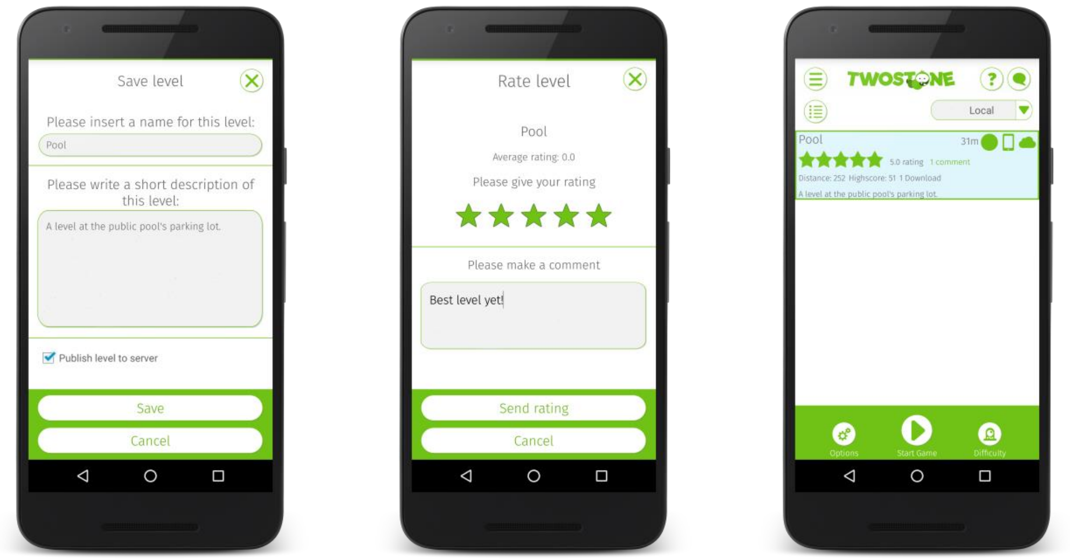


4. Once you have registered successfully, please log in to get to the main screen. There are a couple of options here. Twostone is a location-based game and for playing, you will have to create a new level at your current location. Please look for a spacious area in your vicinity. Any medium-sized lawn, park, or parking lot should do. You can also create levels on streets – but if you do so, please mind the traffic! To start the level editor, tap the "Create new level" button.

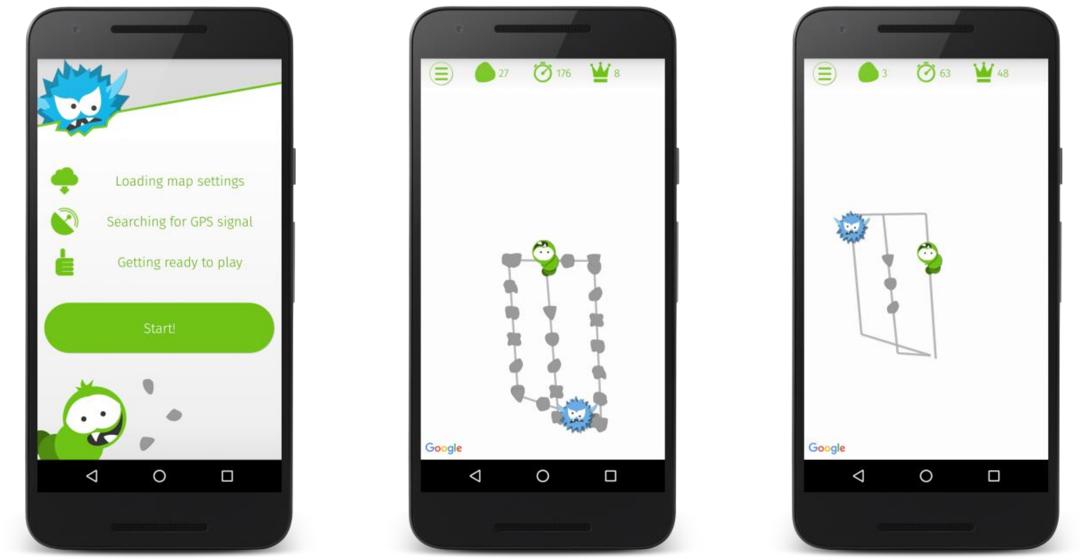
5. Creating new levels is done in a simple and straightforward way: You walk them. In the level editor you should see two buttons for adding and removing waypoints on the left side of the screen. The larger of the two, the one with the plus symbol, adds a new waypoint. Tap it once to start creating a new level. Now walk in a straight line until you need to change your direction. Place a new waypoint at this spot and then turn to the left or right in a 90 degree angle. Keep proceeding in this way until you have at least three lanes next to one another. If you finally arrive back at your original position, add a final waypoint.



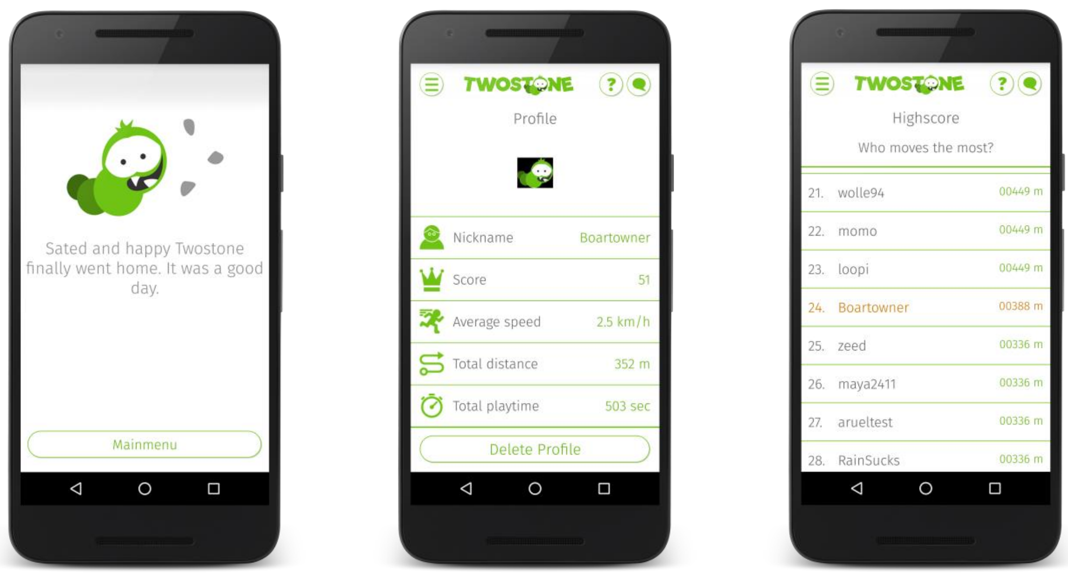
6. If you now hit the “Save level” button, Twostone will ask you to name your creation. You can also add a short level summary if you like. In any case, please make sure that the “Publish level to server”-box is checked so that other people will be able to play this level as well. Using the main menu’s “Level overview”-button, you can take a look at existent levels in your vicinity. You should give them a full 5-star rating if you like what you see. And before you are asking: You can also rate your own levels with a 5-star rating. But why would anyone want to do that? ;)



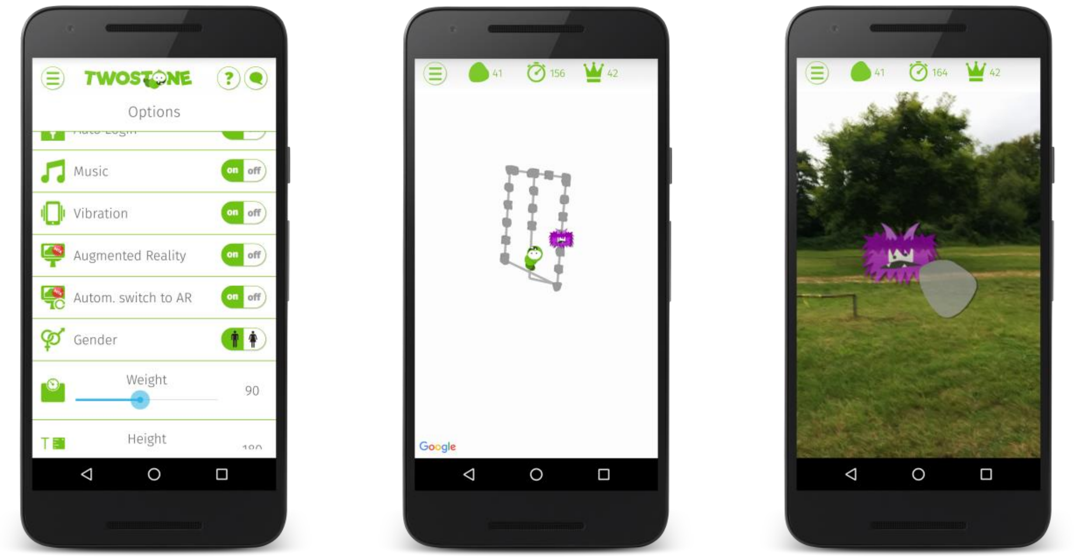
7. Use the “Quickplay”-button on the main screen to automatically start playing the level closest to you. How about trying the level you just created? You have probably already noted that Twostone resembles the game Pac-Man in many ways: The caterpillar Twostone needs to consume stones while it tries to evade a bunch of nerdowells. The main difference to this classic of video game history lies in the fact that this time, it is you who needs to do the escaping! If you ever find the game too easy or too difficult, use your smartphone’s volume buttons to dynamically adapt your opponent’s speed and intelligence.



8. After you have successfully completed your first level, why not take a look at your personal player profile or the highscore list? You can access both via the main menu.



9. There is a somewhat hidden feature that we would like to point out to you. If you go to the options screen you can enable Twostone's Augmented Reality user interface. If you do so, you will find that the game automatically switches between a bird's eye map view and a camera-based perspective, depending on how you are holding your phone. Try it!



IMPORTANT

Please create a level in close vicinity to every location that you frequently visit. Ideally, you will do this whenever you arrive at such a location the first time during the next three weeks.

If you experience any technical difficulties such as crashes, or if the game does not look anything like the one on the screenshots showed here, then please let us know.

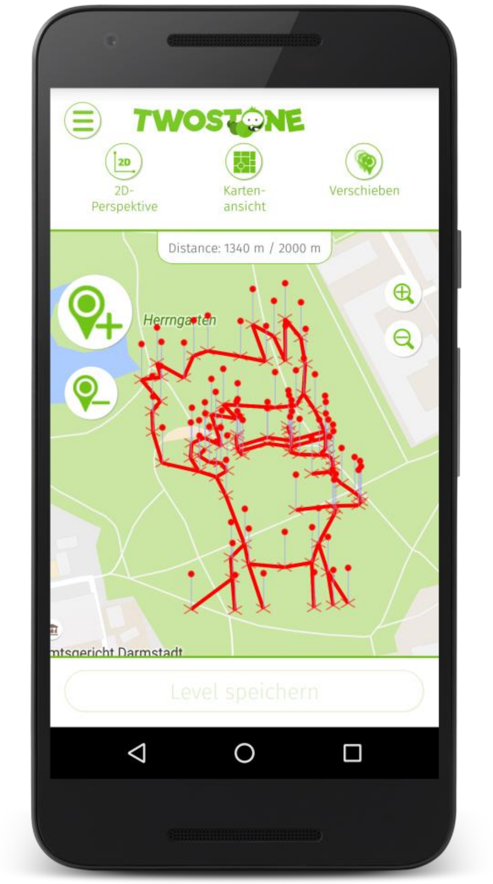
Thank you!

Manual II

This manual will explain to you how to install and initialize the “Trigger Application”. You can think of triggers as being well-timed notifications meant to remind you that there was something you wanted to do – such as playing Twostone in order to be physically active and to increase the probability of staying healthy.

In this context, we would like to thank all of you for your feedback on our game. We received lots of ideas for improvement and requests for bug fixes and we tried to implement as many of them as possible. An updated version of Twostone is now available to you from the Google Play Store. Among other things, we significantly improved the game’s level editor, which should now be a lot easier to use. Although you can still play on your old maps, you should also try to create a new one just to see the differences. Chris has taken the opportunity to design a sophisticated map named “Frystone”, which is located at the Herrngarten park and publically available. Doesn’t the map’s layout look strangely familiar? ;)

Thanks for participating and stay active!
Tim, Jens, Chris, Gerhard & Tobi



Installation and Initialization of the Trigger App

Please perform the following instructions in their given order. Please do not skip any of the steps. If you have not yet completed "Manual I", please do not proceed with this manual before you have not completely worked through the first one.

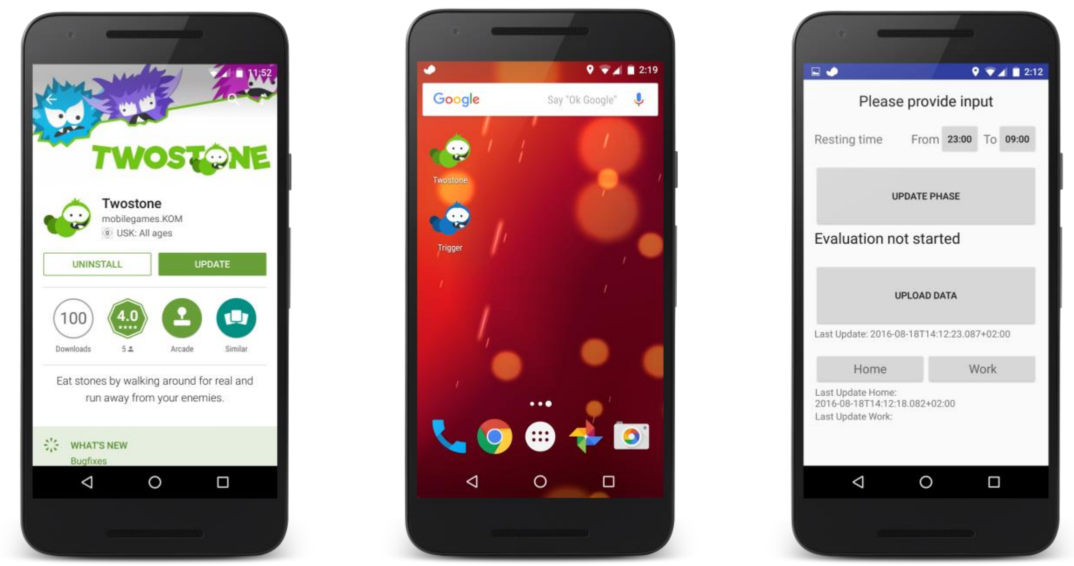
10. (Optional) Install the updated version of Twostone. To do so, go to the Google Play Store and look for "Twostone", then tap the button labeled "Update". If the respective button says "Open" instead, then the latest version of the app is already installed on your smartphone.

11. Install the "Twostone Evaluation Trigger" application (referred to in the following as the "Trigger App"). To do this, open the URL below from your smartphone, or open the Google Play Store application and look for „Twostone Evaluation".

    https://play.google.com/store/apps/details?id=de.dirty_bits.activity_trigger&hl=de

    ATTENTION: Should you already have installed the application prior to having received this manual than please contact us before you proceed!

12. If you try to install the app it will ask you for several access rights, just like other applications. Please make sure to accept all of these requests.



13. Once you have successfully installed the Trigger App, you will find it in the list of your applications; look for the blue caterpillar. Please start the app, which basically consists of a single screen. To begin with, you should define your resting time. This is the time interval during which you don't want to be disturbed because you are sleeping. On the upper right of the screen, click the left box to specify the time when you usually go to bed (such as 23:00 hours), and click the right one to specify when you usually get up (such as 09:00 hours). Please note that the Trigger App is NOT an alarm clock and that it WILL NOT wake you in time!

14. Next is the definition of your current location. We rely on the wireless network ID to figure out, where you are at. To define that the wireless network to which you are currently connected is your home network, click the button labeled "Home". Analogously, if you are at work, please click the button labeled "Work". A few remarks are required here:

    a. You can specify both locations as often as you want. You can change them at any time and if you accidently tapped the wrong button, don't worry – simply correct your choice on the next opportunity.

    b. If you do not have a wireless network at home or at work, please pick another important location instead (such as your friend's or parent's place).

    c. It is very important that you specify these locations as early as possible. Please try to keep this in mind.

15. And that's pretty much it in regard to the initialization of the Trigger App. If you close the app, you should see a small caterpillar icon on the upper left of your screen. In addition, there should now be a box with the app's icon on the start screen. This means that the Trigger App is working as intended.

16. Please also read through the notes on the next two pages and the request for feedback below.

IMPORTANT

In order to participate in the evaluation you must install the Trigger App on your smartphone until Sunday evening (21st of August) the latest. Once you have completed this manual, please tap both the "Update Phase" and the "Upload Data"-button once (tapping multiple times won't hurt either). This will inform us that a user has installed the Trigger App.

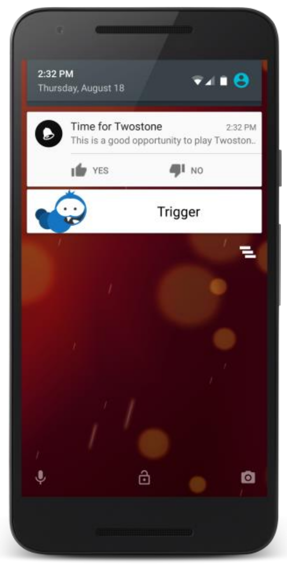
For privacy reasons, your data is not linked to your real name. This means that we cannot tell, who has installed the Trigger App; we only know the amount of people who have done so. In order to keep control of who's still in (and who's not), please send a short mail to twostonestudy@gmail.com and simply state something in the line of "I'm still participating". Thanks!

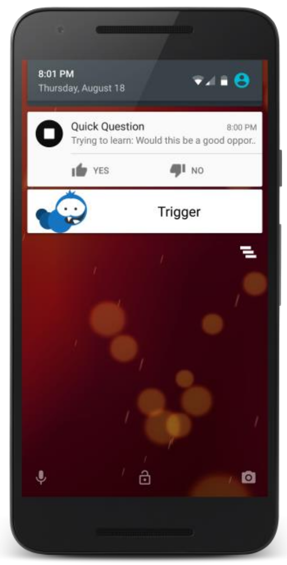
Using the Trigger App

Starting on Monday, the 22nd of August, the Trigger App will regularly ask you whether you would like to play some Twostone. This will happen by way of so-called “trigger notifications” that appear on your start screen. You can either confirm these triggers, or decline them. In the earlier case, the Twostone level closest to you will automatically be started. The evaluation ends after two weeks on the evening of Sunday, the 4th of September.

IMPORTANT! Please read through the following list carefully and please adhere to all instructions. This is essential for ensuring the study’s success.

- During the evaluation you are only allowed to play Twostone after you have been triggered and have confirmed the corresponding trigger. We do not enforce this behavior, and in theory, you could simply start Twostone by your own when you feel like it. Please refrain from doing so, however, as this would negatively affect the study results.
- Please make sure to always start the Trigger App immediately after you have started your smartphone. You start the application simply by opening its main screen. Once you see it, you can return to your start screen. The Trigger App is a so-called background service, meaning that it should be able to start itself automatically. However, this does not reliably work on all smartphone models and especially Samsung devices have been found to be a little problematic in this regard. The small caterpillar in the upper left of your screen and the box on your start screen saying “Trigger” will let you know that the application is working as it should.
- Every now and then, the application may send you a learner notification. This does not need to happen, though, as the amount of learner notifications will vary from user to user. These learner notifications are not triggers – as the name implies, the application just tries to learn about your preferences. If you want to help, you can either confirm the notification (stating that this would have been a good opportunity for playing Twostone) or decline it (indicating that you wouldn’t have played Twostone in this specific situation). However, you can always just wait for the notification to disappear by itself.





- For certain reasons the app does not know, whether a Twostone map is anywhere near you. This is why it is important that you create maps in advance at all those locations that you spend a lot of time at – please keep this in mind. If, however, the app sends you a trigger when no map is anywhere near you and you still feel like playing, please confirm the trigger and let it open Twostone for you. Then make use of the updated editor to quickly create a map and enjoy playing. :)
- Please also keep in mind that you should define your two most important places (they need to be places where you connect your phone to a wireless network). These locations can change over time, however, for instance if you go on a vacation. If "at home" is another place for you during the first week than during the second, then let the app know!
- IMPORTANT: Please tap the button labeled "Update Phase" on the morning of Monday, the 22nd. If done correctly, the "Evaluation not started"-label should disappear and be replaced by a name. Please remember this name, as we will ask you for it at a later time. Please also click the "Upload Data"-button at least once a day. Clicking it multiple times won't do any damage, either.

If you experience crashes, if you get the impression that the app is not running as it should, or if there are any other problems, please let us know as soon as possible via twostonestudy@gmail.com. Thanks for your help!

Bitte um Teilnahme

Hallo und Danke für Dein Interesse an unserer Studie. :)

Das Wissen um den Zusammenhang zwischen Sport und Gesundheit ist zwar weit verbreitet, trotzdem bewegt sich mehr als ein Drittel der europäischen Bevölkerung zu wenig. Und dabei genügen 15 Minuten Bewegung am Tag, um die eigene Lebenserwartung um bis zu drei Jahre zu erhöhen und zwar ganz ohne lästiges Schwitzen.

Seit einigen Jahren wird diskutiert, ob mobile ortsbasierte Spiele wie *Ingress* oder *Pokemon Go* dazu beitragen können, dass Menschen freiwillig das empfohlene tägliche Mindestmaß an sportlicher Aktivität erreichen. Aus diesem Grund haben wir ein eigenes ortsbasiertes Spiel entwickelt: *Twostone*.

Im Rahmen einer dreiwöchigen Evaluation von Montag, dem 15.08., bis Sonntag, dem 04.09, möchten wir nun gerne testen, welchen Effekt dieses Spiel und sein „intelligenter Reminder“ tatsächlich auf den Nutzer haben. Teilnehmer unserer Evaluation müssen über ein Android-Smartphone verfügen, idealerweise mit Android-Version 5.0 oder höher. Alle Teilnehmer sollten zudem zu Beginn und zum Ende der Evaluation jeweils einen kurzen Fragebogen ausfüllen, sowie das Spiel einmal vollständig initialisieren. Alle anderen Schritte sind komplett freiwillig. Als kleines Dankeschön erhält jeder Teilnehmer am Ende der Evaluation zwei Kinopolis-Kinofreikarten – sowie das schöne Gefühl, die Menschheit wieder ein kleines Stückchen vorangebracht zu haben. ;)

Wenn Du uns unterstützen möchtest, dann schick bitte bis zum Freitag, den 12.08., eine kurze Mail an

twostonestudy@gmail.com

Bitte schreibe in die Mail Deinen Namen und welches Smartphone Du besitzt. Wir schicken Dir dann am Samstag, den 13.08., zwei Dokumente: eine Anleitung für die Installation und die Initialisierung des Spiels, sowie den ersten von insgesamt zwei Fragebögen.

Besten Dank!
Tim, Jens, Chris, Gerhard & Tobi

Anleitung I

Danke für Deine Teilnahme an unserer Studie. Zur Teilnahme benötigst Du ein Android Smartphone. Solltest Du mehrere Geräte besitzen, dann nimm bitte dasjenige, das Du am häufigsten mit Dir führst.

In dieser Studie geht es um körperliche Bewegung. In sehr umfangreichen, langjährigen Untersuchungen haben Forscher herausgefunden, dass täglich 15 Minuten Bewegung mittlerer Intensität die Lebenserwartung um bis zu drei Jahre erhöhen. Jede weitere Minute Bewegung steigert diesen positiven Effekt auf die Gesundheit sogar noch. Doch obwohl das Wissen über den Zusammenhang zwischen Bewegung und Gesundheit weit verbreitet ist, bewegt sich ein Drittel der europäischen Bevölkerung immer noch viel zu wenig. Viele verschiedene Krankheiten sind die Folge: Bluthochdruck, Diabetes und sogar Krebs. Es gibt verschiedene und teils widersprüchliche Theorien, warum Menschen sich trotz besseren Wissens zu wenig bewegen, aber in einem Punkt sind sich alle einig: niemand tut gerne Dinge, die einfach keinen Spaß machen. Aus diesem Grund versuchen Forscher, Mittel und Wege zu finden, damit körperliche Aktivität unterhaltsamer wird und sich besser in den Alltag integrieren lässt. Ein Beispiel für einen solchen Ansatz sind die sog. „Exergames“, oft auf als „Fitnessspiele“ bezeichnet. Das sind Spiele, bei denen der Spieler sich bewegen muss, während er spielt. Das Prinzip kennen viele von Euch sicherlich noch von der *Wii* und auch für Smartphones werden immer mehr Exergames veröffentlicht, beispielsweise das derzeit sehr beliebte *Pokemon Go*.

An der TU Darmstadt haben wir ein eigenes Exergame namens *Twostone* für Android-Smartphones entwickelt und in dieser Studie möchten wir nun gerne herausfinden, welchen Effekt dieses Spiel hat. Was genau wir dabei untersuchen möchten, können wir Dir leider nicht verraten, da wir Dich dadurch möglicherweise beeinflussen würden. Verhalte Dich im Umgang mit dem Spiel einfach „ganz natürlich“, als ob wir Dich nicht beim Spielen beobachten würden.

Beobachten ist ein gutes Stichwort. Wir zeichnen verschiedene Daten über Dich auf – vom Ort, an dem Du dich befindest, über die Geschwindigkeit, mit der Du dich fortbewegst, bis hin zu dem Akkustand Deines Smartphones und wie laut es im Durchschnitt um Dich herum ist (wohlgemerkt: nur Messung von Dezibel, keine Aufzeichnung von Geräuschen oder Stimmen). Und wir tun das nicht nur, während Du spielst, sondern regelmäßig. Das klingt erst mal irritierend, aber sei Dir versichert, dass wir äußerst sensibel mit diesen Daten umgehen. Zum einen abstrahieren wir von Vorneherein. Statt „fährt gerade Zug“ kommt bei uns daher beispielsweise nur noch die Information „in einem Fahrzeug“ an. Insbesondere stellen wir aber sicher, dass jeder persönliche Bezug zwischen Dir und Deinen Daten entfernt wird. So löschen wir etwa nach der Studie alle Deine Emails, entfernen Deine Datensätze vom Server und vernichten die Namensliste der Teilnehmer. Ohne uns an dieser Stelle selbst auf die Schulter klopfen zu wollen: Google und Facebook machen das alles sicher nicht. Falls Du trotzdem Bedenken zum Thema Datenschutz haben solltest, dann setz Dich bitte mit uns in Verbindung.

Eine Sache noch: sei ein bisschen nachsichtig mit uns. Wir sind ein kleines Team und obwohl wir hart an Twostone und der Reminder-App gearbeitet haben (dazu später mehr), wird es sicherlich Fehler und Abstürze geben. Und auch Dein Akku könnte sich etwas schneller leeren, als Du es gewohnt bist. Wir würden uns freuen, wenn Du mit solchen Problemen wohlwollend und nachsichtig umgehst. Falls etwas aber einmal so gar nicht funktionieren sollte, dann wende Dich bitte an twostonestudy@gmail.com und wir versuchen, Dir schnellstmöglich zu helfen.

Viel Spaß!
Tim, Jens, Chris, Gerhard & Tobi

Ablauf der Studie

Ein paar Worte zum Ablauf der Studie. Heute, am Samstag den 13.08., hast Du zwei Dokumente von uns bekommen: diese Anleitung, sowie einen ersten Fragebogen. Bitte führe diese Anleitung komplett durch, fülle den Fragebogen aus und schicke ihn an uns zurück. Danach geht es so weiter:

- Sonntag, 14.08.: Du erhältst eine zweite Anleitung. Bitte führe auch diese Anleitung vollständig durch. Falls möglich noch am Sonntag – aber unbedingt erst nachdem Du diese erste Anleitung durchgeführt hast.

- Montag, 15.08.: Beginn der eigentlichen Studie. Die Evaluation dauert insgesamt 20 Tage, also bis einschließlich Samstag, den 03.09. Deine Aufgabe während dieser Zeit besteht darin, die in der zweiten Anleitung beschriebene Reminder-App im Hintergrund auf Deinem Smartphone laufen zu lassen und immer dann Twostone zu spielen, wenn Du Lust dazu hast.

- Sonntag, 04.09.: Wir schicken Dir einen zweiten Fragebogen. Bitte fülle diesen schnellstmöglich aus und schicke ihn an uns zurück. Als kleines Dankeschön erhältst Du von uns nach Abgabe des zweiten Fragebogens zwei Kinopolis-Kinofreikarten. Setze Dich dazu mit demjenigen in Kontakt, der Dich für die Studie geworben hat.

- Du kannst unsere Evaluationsergebnisse einsehen. Falls Du daran interessiert bist, welche Ergebnisse unsere Studie hervorgebracht hat (und was wir wissenschaftlich veröffentlichen werden), dann lass uns das wissen. Sobald wir mit der Auswertung der Daten fertig sind, schicken wir Dir eine Zusammenstellung.

Installation und Einrichten von Twostone

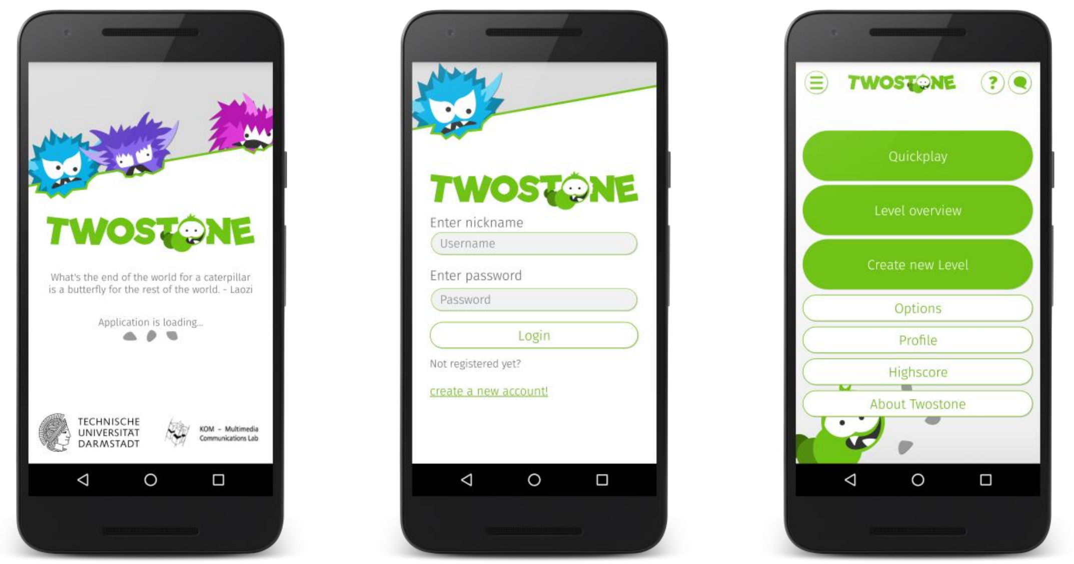
In der heutigen Anleitung kümmern wir uns um Installation und Einrichten von Twostone. Im zweiten Teil der Anleitung, den wir Dir morgen schicken werden, kommt dann noch eine Reminder-Applikation hinzu. Bitte führe diese Anleitung genau so aus, wie sie beschrieben ist. Bitte überspringe keine Anweisungen und bitte verändere nicht die Reihenfolge, in der Du die Schritte durchführst.

17. Lösche Twostone von Deinem Smartphone, falls es bereits installiert war.

18. Installiere die neuste Version von Twostone. Rufe dazu von Deinem Smartphone aus den nachfolgenden Link auf, oder suche im Google Play Store nach „Twostone“.

    https://play.google.com/store/apps/details?id=de.tu.darmstadt.uhg

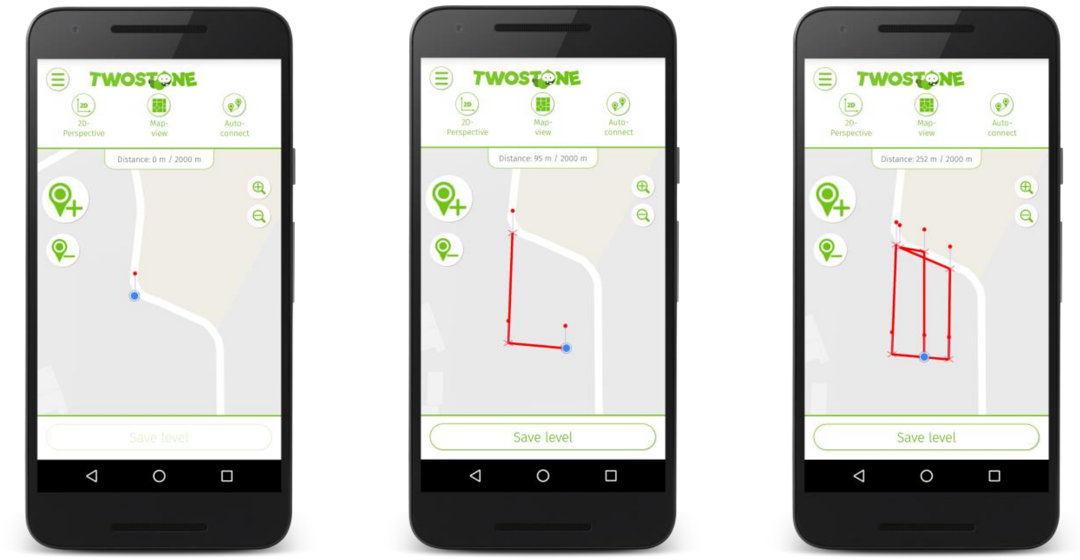
19. Beim ersten Starten von Twostone musst Du ein Benutzerkonto anlegen. Klicke dazu auf den Link unterhalb des Login-Buttons. Achtung, Dein Benutzername unterscheidet zwischen Groß- und Kleinschreibung. Falls Du bereits ein Benutzerkonto haben solltest, dann lege für diese Evaluation bitte ein weiteres Konto an.



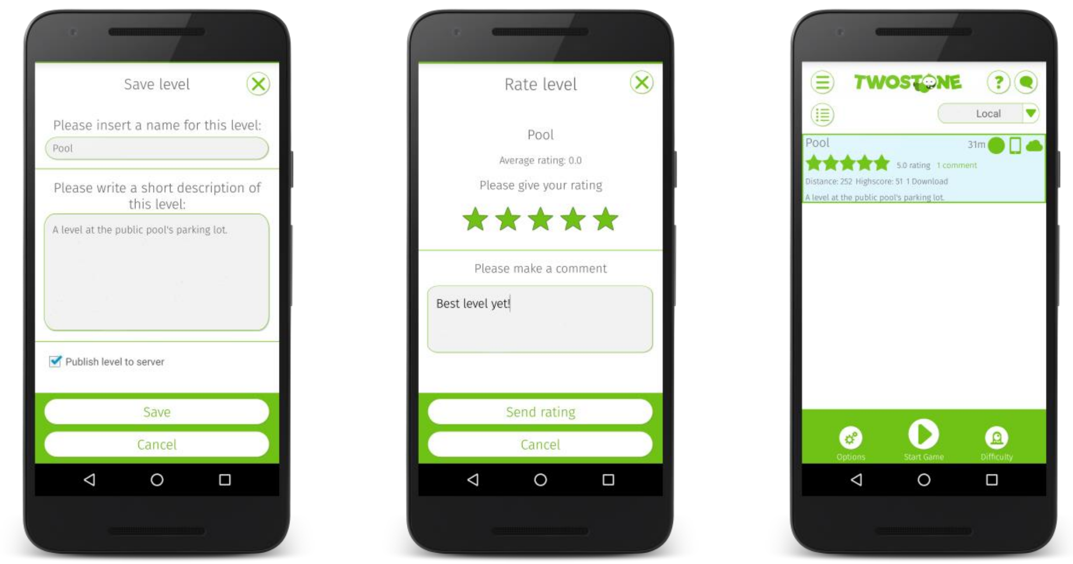
20. Nach erfolgreicher Registrierung gelangst Du auf den Hauptbildschirm. Hier stehen Dir mehrere Optionen zur Verfügung. Twostone ist ein ortsbezogenes Spiel und damit Du es spielen kannst, musst Du erst einen neuen Level anlegen. Suche dazu eine größere freie Fläche in Deiner Nähe. Geeignet sind Felder, Parks, Parkplätze, etc. Du kannst Level auch entlang von Straßenzügen anlegen, aber dann achte auf den Verkehr! Um den Leveleditor aufzurufen, drücke den „Neues Level erstellen“-Knopf.

21. Das Erstellen neuer Level erfolgt nach einem denkbar einfachen Prinzip: Du läufst sie einfach ab. Im Leveleditor siehst Du auf der linken Seite des Bildschirms zwei Knöpfe. Der größere, obere Knopf (der mit dem Plus-Zeichen) setzt einen neuen Wegpunkt. Drücke ihn einmal zu Beginn und dann laufe in gerader Linie los. Falls Du Deine Richtung veränderst, dann setze an dieser Stelle einen weiteren Wegpunkt. Versuche zunächst, nur Abzweigungen in 90 Grad Winkeln zu machen

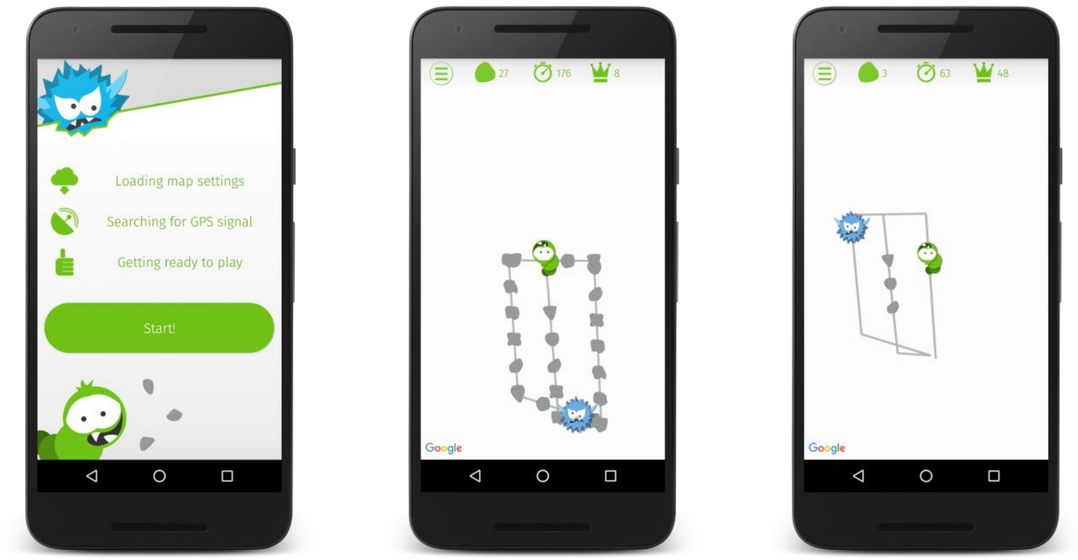
und erstelle auf diese Weise mindestens drei Spuren nebeneinander. Wenn Du schließlich an Deinen Ausgangspunkt zurückgekehrt bist, dann setze einen letzten Wegpunkt.



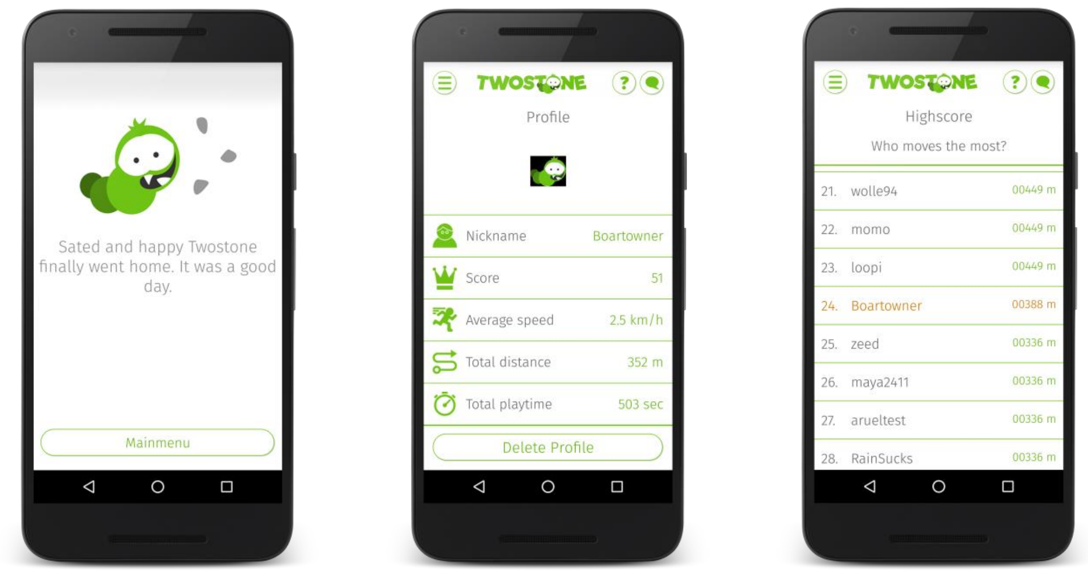
22. Wenn Du jetzt den „Speichern“-Knopf drückst, dann wirst Du aufgefordert, Deinen Level zu benennen. Du kannst auch noch eine kurze Beschreibung hinzufügen. Stelle sicher, dass die Option „Auf Server veröffentlichen“ aktiviert ist, damit sich auch andere Spieler an Deinem Level versuchen können. Über den Knopf „Levelübersicht“ im Hauptmenü kannst Du bereits erstellte Level in Deiner Nähe ansehen und bei Gefallen mit der vollen Punktzahl bewerten. Und ja, auch Deinen eigenen. Aber wer macht denn sowas? ;)



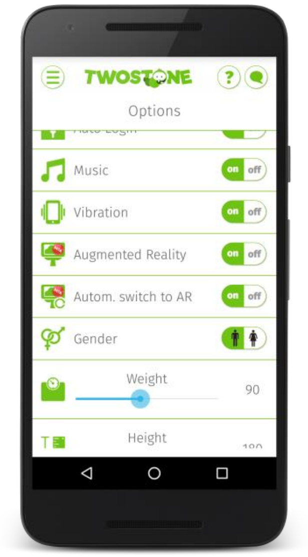
23. Über den „Schnellstart“-Knopf im Hauptmenü wird der nächstgelegene Level gestartet. Versuche Dich direkt mal an dem Level, den Du gerade erstellt hast. Wie Du siehst, ähnelt das Spielprinzip von Twostone dem von Pac-Man: Du musst als Raupe Twostone die grauen Steine fressen, während Dir bis zu fünf Unholde auf die Pelle rücken. Der Unterschied zu Pac-Man besteht darin, dass Du in Twostone tatsächlich selbst mit den Unholden um die Wette läufst. Falls Dir das Spiel mal zu leicht oder zu schwer wird, dann kannst Du Geschwindigkeit und Intelligenz Deiner Gegner jederzeit über die Lautstärkeknöpfe an der Seite Deines Smartphones dynamisch anpassen.

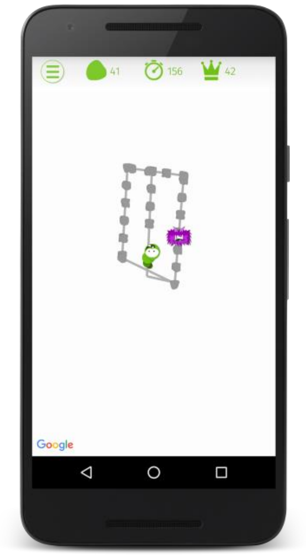


24. Wirf nach dem erfolgreichen Beenden eines Levels ruhig mal einen Blick in Dein Spielerprofil und auf die Highscore-Liste. Beides erreichst Du über das Hauptmenü.



25. Es gibt noch ein verstecktes Feature, das wir Dir nicht vorenthalten möchten. Unter dem Menüpunkt Optionen kannst Du Twostones Augmented Reality Interface aktivieren. Das Spiel wechselt dann automatisch zwischen Kartenansicht und Kamerabild, je nachdem, wie Du Dein Smartphone hältst. Probier's mal aus!





WICHTIG

Bitte erstelle in der Umgebung jedes Ortes, an dem Du dich länger aufhältst, mindestens einen Level – insbesondere in der unmittelbaren Umgebung Deiner Wohnung und Deines Arbeitsplatzes. Tu das innerhalb der nächsten drei Wochen am besten immer direkt dann, wenn Du zum ersten Mal an einem solchen für Dich zentralen Ort eintriffst.

Solltest Du technische Schwierigkeiten mit dem Spiel haben, etwa häufige Abstürze oder falls das Spiel nicht so aussieht, wie auf den hier gezeigten Screenshots, dann setz Dich bitte mit uns in Verbindung.

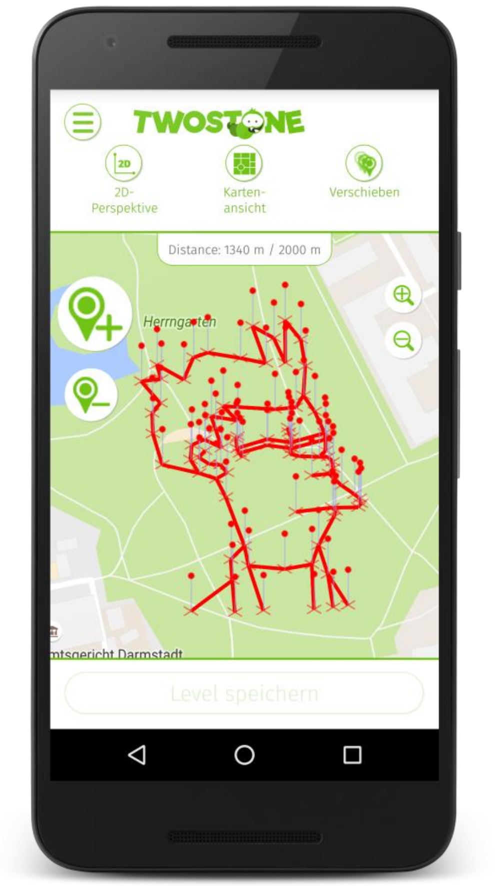
Danke!

Anleitung II

In diesem Dokument erklären wir Dir, wie Du die „Trigger Anwendung“ installierst und initialisierst. Trigger sind gewissermaßen intelligente Reminder, die Dich daran erinnern sollen, dass Du etwas Bestimmtes tun wolltest. In diesem konkreten Fall erinnern Dich die Reminder daran, dass Du Twostone spielen wolltest, um Dein tägliches Mindestmaß an Bewegung zu erreichen und um auf diese Weise die Wahrscheinlichkeit zu erhöhen, dass Du lange fit und gesund bleibst.

An dieser Stelle vielen herzlichen Dank für Eure zahlreichen Rückmeldungen zu unserem Spiel. Wir haben viele Verbesserungsvorschläge und –wünsche erhalten und uns bemüht, eine möglichst große Anzahl davon direkt umzusetzen. Im Google Play Store steht Euch jetzt ein Update für Twostone zur Verfügung, das einige Verbesserungen einführt und diverse Fehlerchen behebt. Unter anderem haben wir am Leveleditor gearbeitet, der jetzt hoffentlich einfacher zu bedienen ist und bessere Resultate erzeugt. Ihr könnt Eure alten Karten weiterhin nutzen, aber probiert ruhig auch einmal das Anlegen einer neuen Karte aus, um die Änderungen zu erleben. Chris hat sich direkt an einer etwas komplexeren Karte namens „Frystone“ probiert, die Ihr ab sofort im Herrngarten findet. Das Layout der Karte könnte dem ein oder anderen von Euch vielleicht bekannt vorkommen. ;)

Danke für Eure Teilnahme und bleibt aktiv!
Tim, Jens, Chris, Gerhard & Tobi



Installation und Einrichten der Trigger App

Bitte führe die nachfolgenden Schritte in der angegebenen Reihenfolge durch und bitte überspringe keinen der Schritte. Falls Du *„Anleitung I“* noch nicht vollständig abgearbeitet hast, dann tue das bitte unbedingt, bevor Du mit den nachfolgenden Anweisungen beginnst.

26. (Optional) Installiere das Twostone-Update. Suche dazu im Google Play Store nach „Twostone“ und klicke auf den „Update“-Knopf. Falls anstelle des „Update“-Knopfs bei Dir nur ein „Öffnen“-Knopf ist, dann hast Du bereits die neueste Version von Twostone auf Deinem Smartphone.

27. Installiere die „Twostone Evaluation Trigger“-Anwendung (im Folgenden abkürzend Trigger App genannt). Rufe dazu von Deinem Smartphone aus den nachfolgenden Link auf, oder suche im Google Play Store nach „Twostone Evaluation“.

    https://play.google.com/store/apps/details?id=de.dirty_bits.activity_trigger&hl=de

    WICHTIG: Solltest Du diese Anwendung bereits vor Erhalt dieser Anleitung auf Deinem Smartphone installiert haben, dann kontaktiere uns bitte, bevor Du fortfährst!

28. Beim Installieren der Trigger App wird Dich diese um Zugriff auf einige Funktionen des Smartphones bitten, wie Du es auch von anderen Anwendungen kennst. Bitte gewähre alle nachgefragten Zugriffe.

29. Nach der erfolgreichen Installation findest Du die Trigger App wie gewohnt in der Liste Deiner Anwendungen – sie hat eine blaue Raupe als Icon. Bitte starte die App, die im Wesentlichen nur aus einem Hauptbildschirm besteht. Als erstes sollten wir Deine Ruhezeit einstellen. Das ist die Phase, während der Du normalerweise schläfst und daher nicht gestört werden möchtest. Stelle die Uhrzeit ein, indem Du auf die beiden entsprechenden Felder oben rechts klickst. Das linke Feld steht für die Uhrzeit, zu der Du für gewöhnlich zu Bett gehst (beispielsweise 23:00 Uhr) und das rechte Feld steht für die Uhrzeit, zu der Du normalerweise aufstehst (beispielsweise 09:00 Uhr). ACHTUNG: Die Anwendung ist kein Wecker!

30. Als nächstes sollten wir den Ort spezifizieren, an dem Du dich befindest. Wir erkennen den Ort anhand des WLANs, in das Du eingeloggt bist. Um festzulegen, dass das aktuelle WLAN Dein heimisches WLAN ist, drücke bitte den Knopf „Zuhause". Falls Du dich hingegen gerade auf der Arbeit befindest, so drücke bitte den Knopf „Arbeitsort". Ein paar Anmerkungen zu diesem Punkt:

    a. Du kannst beide Orte beliebig oft festlegen und auch verändern. Es macht auch nichts, wenn Du dich mal verklickst – dann lege den Ort bei nächster Gelegenheit einfach noch mal neu fest.

    b. Falls Du zuhause oder auf der Arbeit kein WLAN haben solltest, dann nimm stattdessen diejenigen Orte mit WLAN, an denen Du dich am häufigsten aufhältst (beispielsweise bei Freunden oder Verwandten).

    c. Es ist wichtig, dass Du beide Orte so früh wie möglich festlegst. Bitte versuche, daran zu denken.

31. Und das war's auch schon in Sachen Installation und Initialisierung. Nach dem Schließen der App solltest Du in der Icon-Leiste am oberen Bildschirmrand den Schattenriss einer Raupe sehen und auf dem Startbildschirm eine Box mit dem Logo der App. Das zeigt Dir, dass die Anwendung korrekt ausgeführt wird.

32. Bitte lies Dir jetzt direkt noch die Hinweise zur Nutzung während der Evaluationsphase auf der nächsten Seite durch und bitte beachte auch die folgende Bitte um eine kurze Rückmeldung.

WICHTIG

Für die Teilnahme an unserer Studie musst Du die Trigger App unbedingt bis zum späten Sonntagabend (21.08.) installiert haben. Bitte drücke nach Abschluss der oben beschriebenen Installationsanleitung jeweils einmal den „Aktualisiere Phase"-Knopf und den „Sende Evaluationsdaten"-Knopf (es macht nichts, wenn Du die Knöpfe mehrfach drückst). Auf diese Weise werden wir darüber informiert, dass ein Nutzer die Anwendung erfolgreich installiert und initialisiert hat.

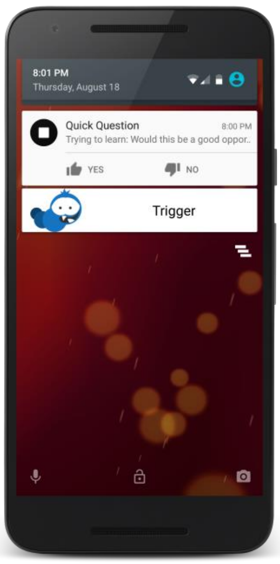
Aus Datenschutzgründen verknüpfen wir Deine Datensätze nicht mit Deinem Klarnamen. Wir wissen also nicht, welcher Nutzer die Trigger App schon installiert hat – wir kennen lediglich die Anzahl der Personen, die das getan haben. Um den Überblick darüber zu behalten, wer überhaupt noch alles dabei ist, würden wir uns daher über eine kurze Rückmeldung von Dir an twostonestudy@gmail.com sehr freuen. Ein knappes „Ich bin noch dabei" genügt uns schon. Danke!

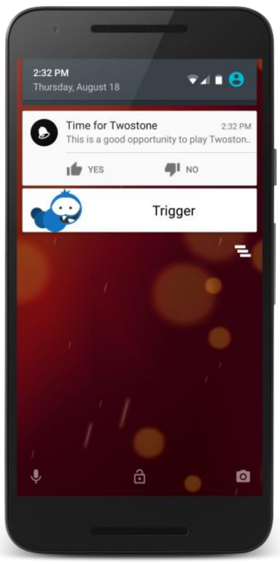
Nutzung während der Evaluationsphase

Ab Montag, dem 22.08., wird Dich die Trigger App regelmäßig dazu auffordern, Twostone zu spielen. Du erhältst dazu auf Deinem Startbildschirm eine sog. „Trigger Notification“, die Du entweder bestätigen oder ablehnen kannst. Wenn Du den Trigger bestätigst, dann wird automatisch der nächstgelegene Level gestartet. Die Evaluation endet nach zwei Wochen am Sonntagabend, dem 04.09.

WICHTIG! Bitte lies Dir die nachfolgende Liste aufmerksam durch und beachte unbedingt alle aufgeführten Punkte. Du trägst damit wesentlich zum Erfolg der Evaluation bei.

- Innerhalb der Evaluationsphase darfst Du Twostone nur spielen, wenn der Trigger Dich dazu auffordert und Du die entsprechende Nachfrage mit „Ja“ beantwortest. Im Prinzip kannst Du Twostone nach wie vor jederzeit manuell starten, aber wir bitten Dich, davon abzusehen, da Du andernfalls die Evaluationsergebnisse verfälscht.
- Bitte starte die Trigger App nach jedem Neustart Deines Smartphones sofort erneut. Tu tust dies einfach, indem Du sie ausführst. Nachdem Du den Hauptbildschirm siehst, kannst Du die Anwendung direkt wieder schließen. Die Anwendung ist ein sog. „Background Service“ und sollte sich eigentlich von selbst starten. Bei einigen Smartphone-Modellen kann das aber auch mal nicht klappen – insbesondere Samsung-Geräte machen hier ein bisschen Probleme. Über das kleine Raupenicon in der oberen Bildschirmleiste und über die Benachrichtigungsbox auf Deinem Startbildschirm kannst Du dich vergewissern, dass die Anwendung tatsächlich läuft.
- Gegebenenfalls schickt Dir die Trigger App ab und an auch mal eine Lernnachricht. Ob das überhaupt passiert, ist aber individuell verschieden und muss nicht unbedingt vorkommen. Diese Nachrichten sind keine richtigen Trigger – die Anwendung möchte nur besser verstehen, wann Du gerne Twostone spielen würdest. Wenn Du möchtest, dann kannst Du der App dadurch helfen, dass Du solche Lernnachrichten mit Ja (der Zeitpunkt der Nachricht wäre für Dich im Prinzip eine gute Gelegenheit für eine schnelle Runde Twostone) oder mit Nein (das wäre keine gute Gelegenheit) beantwortest. Du kannst diese Nachrichten aber auch einfach ignorieren – dann verschwinden sie nach kurzer Zeit wieder von selbst.





- Aus bestimmten Gründen weiß die Trigger App nicht, ob sich in Deiner Nähe ein geeignetes Spielfeld befindet. Daher ist es wichtig, dass Du an allen Orten, an denen Du dich häufig aufhältst, schon im Vorhinein Spielfelder anlegst, damit Du im Fall der Fälle dann auch direkt loslegen kannst – bitte denk daran! Falls Du mal zum Spielen aufgefordert wirst, obwohl kein Feld in der Nähe ist, Du aber trotzdem gerne spielen möchtest, dann bestätige die Rückfrage der App bitte mit „Ja“. Wenn Twostone sich geöffnet hat, dann kannst Du an Ort und Stelle direkt eine neue Karte anlegen – mit den neuesten Updates unseres Leveleditors sollte das noch schneller gehen.
- Bitte denk auch daran, im Hauptmenü der Anwendung die beiden für Dich wichtigsten Orte einzustellen (es müssen Orte sein, an denen Du in ein WLAN eingeloggt bist). Diese Orte können sich auch über die Zeit verändern, beispielsweise, wenn Du in den Urlaub fährst. Falls also für Dich beispielsweise „Zuhause“ in der zweiten Woche der Evaluation ein anderer Ort sein sollte, als während der ersten Woche, dann ist das überhaupt kein Problem – klicke den entsprechenden Knopf im Hauptmenü dann einfach noch mal und definiere so neu, was gerade Dein „Zuhause“ ist.
- WICHTIG: Bitte drücke zu Beginn der zweiten Phase (am Morgen des 22.08.) einmal den Knopf „Aktualisiere Phase“. Daraufhin erscheint unter dem Knopf ein Name – bitte merk ihn Dir, wir werden Dich später danach fragen. Bitte drücke zudem an jedem Tag der Evaluation möglichst einmal den „Sende Evaluationsdaten“-Knopf. Du kannst beide Knöpfe beliebig oft drücken, dadurch machst Du nichts kaputt.

Falls Du Abstürze erlebst, falls Du das Gefühl hast, dass die Anwendung nicht richtig funktioniert, oder falls es andere Probleme gibt, dann wende Dich bitte so schnell wie möglich an twostonestudy@gmail.com. Danke für Deine Mithilfe!

Pre-Study Questionnaire

Let's start with a few general questions. Please be honest. All answers will be anonymized and no one will be able to tell from the study results how you filled this questionnaire. Additionally, on this first page you can always refuse to answer. However, please try to not make use of this if possible.

pre01	Boy or girl?	Boy	Girl	Not telling

pre02	How old are you?	Less than 20	20-39	40-59	60 or more	Not telling

pre03	What's your highest qualification (including currently pursuing)?	None	High school	Apprentice-ship	Bachelor/ Master	Not telling

pre04	How many hours per week do you spend working or studying?	Less than 20	20-39	40-59	60 or more	Not telling

pre05	How many hours per week do you do sports or exercise?	Less than 1	1-3	4-7	More than 7	Not telling

pre06	How many hours per week do you do spend playing video games?	Less than 1	1-3	4-7	More than 7	Not telling

pre07	How many hours per week do you do spend watching TV (including Netflix, etc.)?	Less than 1	1-3	4-7	More than 7	Not telling

pre08	How many hours per week do you do spend with other hobbies (excluding TV/PC)?	Less than 1	1-3	4-7	More than 7	Not telling

On the next page we will ask for your opinion. Please read all questions carefully, but also try to answer spontaneously.

pre09	I would like to exercise more, but I simply lack the time.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pre10	Honestly: I was never into sports.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pre11	I would like to exercise more, but the conditions are not ideal.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pre12	If I do sports, I prefer the outside to indoors.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pre13	Computers and technical stuff are not my thing.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pre14	I'm always carrying my smartphone with me and I take it everywhere.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pre15	I occasionally play games on my smartphone.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pre16	My life got a lot more hectic during the last ten years.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pre17	I could explain the difference between an accelerometer and a gyroscope.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

And finally a few questions regarding your experience with video games.

pre18	Twostone	Never heard of it	Heard, but haven't played	I remember taking a look	Played it, but not anymore	Still enjoying!

pre19	Pokemon Go	Never heard of it	Heard, but haven't played	I remember taking a look	Played it, but not anymore	Still enjoying!

pre20	Ingress	Never heard of it	Heard, but haven't played	I remember taking a look	Played it, but not anymore	Still enjoying!

pre21	Zombies, Run!	Never heard of it	Heard, but haven't played	I remember taking a look	Played it, but not anymore	Still enjoying!

pre22	Wii Fit or Wii Sports	Never heard of it	Heard, but haven't played	I remember taking a look	Played it, but not anymore	Still enjoying!

pre23	Xbox Fitness, PlayStation Move Fitness or PlayStation Zumba	Never heard of it	Heard, but haven't played	I remember taking a look	Played it, but not anymore	Still enjoying!

pre24	Minecraft	Never heard of it	Heard, but haven't played	I remember taking a look	Played it, but not anymore	Still enjoying!

pre25	Grand Theft Auto V	Never heard of it	Heard, but haven't played	I remember taking a look	Played it, but not anymore	Still enjoying!

And that's it. Don't forget to safe and then send the document back to twostonestudy@gmail.com, Thanks! :)

Post-Study Questionnaire

Please try to be honest. Your answers will be aggregated and will not be reproducible.

pos01	Boy or girl?	Boy	Girl	Not telling

pos02	How old are you?	

pos03	What name did appear in your Trigger App?	

pos04	I have a lot of video gaming experience.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pos05	I'm a sportsperson.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pos06	I'm kind of a geek and love to have new technical stuff.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pos07	I'm extroverted.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pos08	I'm very busy at the moment (with work, exams, etc.).	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pos09	I try to eat healthy.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pos10	I like being outside.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pos11	I dislike sweating.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pos12	I always have my smartphone with me.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pos13	I have enough time for sport.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pos14	I like doing sport.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pos15	I like to go running.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pos16	Sport is a bit complicated for me at the moment (driving to sports range, etc.)	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pos17	I prefer team sports.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pos18	I am a competitive person.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pos19	I should probably do more sport.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pos20	I enjoy playing Twostone.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pos21	I will keep playing Twostone.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pos22	I think Twostone can help me stay active and healthy.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pos23	I think Twostone could help others stay active and healthy.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pos24	Twostone is for kids.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pos25	Twostone is for adults.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pos26	Twostone is for seniors.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pos27	I like playing my own maps.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pos28	I like playing maps that were made by others.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pos29	Twostone needs to be improved in regard to usability and bugs.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pos30	I would probably play a game like Twostone if it was of higher quality.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pos31	Playing Twostone is too complicated.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pos32	The trigger was annoying.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pos33	The trigger was more annoying during the first week.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pos34	The trigger selected good opportunities for playing.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pos35	The trigger improved over time.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pos36	I sometimes did not notice the trigger.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pos37	I sometimes ignored the trigger.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pos38	Without the trigger I would have played less Twostone.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pos39	The trigger needs to be more intelligent.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

pos40	The trigger needs to be improved in regard to usability and bugs.	Totally incorrect	Kind of incorrect	Not sure	Kind of correct	Totally correct

And that's it.

Please simply send this questionnaire back to twostonestudy@gmail.com and you're done.

Thanks for your help! :)

Vorstudienbogen

Zunächst ein paar ganz allgemeine Fragen für unsere Statistik. Bitte sei ehrlich. Alle Antworten werden von uns anonymisiert und aus den Ergebnissen der Studie wird man nicht nachvollziehen können, wie Du diesen Fragebogen beantwortet hast. Du hast zudem bei dieser ersten Gruppe von Fragen die Option, keine Angaben zu machen. Bitte versuche aber, das so wenig wie möglich zu nutzen.

pre01	Männlein oder Weiblein?	Männlein	Weiblein	Keine Angabe

pre02	Wie alt bist Du?	Unter 20	20-39	40-59	60 oder älter	Keine Angabe

pre03	Was ist Dein höchster Abschluss (auch aktuell angestrebter)?	Kein Abschluss	Schul-abschluss	Lehre	Studium	Keine Angabe

pre04	Wie viele Stunden pro Woche arbeitest Du/lernst Du?	Unter 20	20-39	40-59	60 oder mehr	Keine Angabe

pre05	Wie viele Stunden pro Woche machst Du Sport?	Unter 1	1-3	4-7	Über 7	Keine Angabe

pre06	Wie viele Stunden Pro Woche verbringst Du mit Videospielen?	Unter 1	1-3	4-7	Über 7	Keine Angabe

pre07	Wie viele Stunden pro Woche verbringst Du mit Fernsehen (auch Netflix, etc.)?	Unter 1	1-3	4-7	Über 7	Keine Angabe

pre08	Wie viele Stunden pro Woche verbringst Du mit anderen Hobbies (ohne TV/PC)?	Unter 1	1-3	4-7	Über 7	Keine Angabe

Auf der nächsten Seite kommen ein paar Fragen nach Deiner Meinung. Bitte lies Dir jede Frage genau durch, aber versuche dann möglichst spontan zu antworten.

pre09	Ich würde gerne mehr Sport treiben, aber mir fehlt einfach die Zeit dazu.	Stimmt nicht	Stimmt eher nicht	Bin mir unsicher	Stimmt ein bisschen	Stimmt absolut

pre10	Ganz ehrlich: Ich mochte Sport eigentlich noch nie so richtig.	Stimmt nicht	Stimmt eher nicht	Bin mir unsicher	Stimmt ein bisschen	Stimmt absolut

pre11	Ich würde mehr Sport treiben, aber die Rahmenbedingungen passen nicht.	Stimmt nicht	Stimmt eher nicht	Bin mir unsicher	Stimmt ein bisschen	Stimmt absolut

pre12	Wenn ich Sport mache, dann lieber draußen als drinnen.	Stimmt nicht	Stimmt eher nicht	Bin mir unsicher	Stimmt ein bisschen	Stimmt absolut

pre13	Ich hab es nicht so mit Computern und Technik.	Stimmt nicht	Stimmt eher nicht	Bin mir unsicher	Stimmt ein bisschen	Stimmt absolut

pre14	Ich habe mein Smartphone so gut wie immer und überall dabei.	Stimmt nicht	Stimmt eher nicht	Bin mir unsicher	Stimmt ein bisschen	Stimmt absolut

pre15	Ich spiele auch ganz gerne mal auf meinem Smartphone.	Stimmt nicht	Stimmt eher nicht	Bin mir unsicher	Stimmt ein bisschen	Stimmt absolut

pre16	Mein Leben ist in den letzten 10 Jahren hektischer geworden.	Stimmt nicht	Stimmt eher nicht	Bin mir unsicher	Stimmt ein bisschen	Stimmt absolut

pre17	Ich kann den Unterschied zwischen einem Accelerometer und einem Gyroskop erklären.	Stimmt nicht	Stimmt eher nicht	Bin mir unsicher	Stimmt ein bisschen	Stimmt absolut

Abschließend noch ein paar Fragen zu Deiner Erfahrung mit verschiedenen Videospielen.

pre18	Twostone	Noch nie davon gehört	Kenne ich, aber nie gespielt	Irgendwann mal angeschaut	Früher gespielt, nicht mehr	Spiele ich immer noch

pre19	Pokemon Go	Noch nie davon gehört	Kenne ich, aber nie gespielt	Irgendwann mal angeschaut	Früher gespielt, nicht mehr	Spiele ich immer noch

pre20	Ingress	Noch nie davon gehört	Kenne ich, aber nie gespielt	Irgendwann mal angeschaut	Früher gespielt, nicht mehr	Spiele ich immer noch

pre21	Zombies, Run!	Noch nie davon gehört	Kenne ich, aber nie gespielt	Irgendwann mal angeschaut	Früher gespielt, nicht mehr	Spiele ich immer noch

pre22	Wii Fit oder Wii Sports	Noch nie davon gehört	Kenne ich, aber nie gespielt	Irgendwann mal angeschaut	Früher gespielt, nicht mehr	Spiele ich immer noch

pre23	Xbox Fitness, PlayStation Move Fitness oder PlayStation Zumba	Noch nie davon gehört	Kenne ich, aber nie gespielt	Irgendwann mal angeschaut	Früher gespielt, nicht mehr	Spiele ich immer noch

pre24	Minecraft	Noch nie davon gehört	Kenne ich, aber nie gespielt	Irgendwann mal angeschaut	Früher gespielt, nicht mehr	Spiele ich immer noch

pre25	Grand Theft Auto V	Noch nie davon gehört	Kenne ich, aber nie gespielt	Irgendwann mal angeschaut	Früher gespielt, nicht mehr	Spiele ich immer noch

Das war’s schon. Speichern nicht vergessen und das Ganze einfach unkommentiert zurück an twostonestudy@gmail.com. Danke! :)

Nachstudienbogen

Bitte versuche, ehrlich zu sein. Alle Antworten werden aggregiert und sind dadurch nicht zuordbar.

pos01	Männlein oder Weiblein?	Männlein	Weiblein	Keine Angabe

pos02	Wie alt bist Du?	

pos03	Welcher Name stand bei Dir in der Trigger App?	

pos04	Ich habe viel Erfahrung mit Videospielen.	Stimmt nicht	Stimmt eher nicht	Bin mir unsicher	Stimmt ein bisschen	Stimmt absolut

pos05	Ich bin Sportler.	Stimmt nicht	Stimmt eher nicht	Bin mir unsicher	Stimmt ein bisschen	Stimmt absolut

pos06	Ich bin manchmal ein Geek und liebe neue technische Spielsachen.	Stimmt nicht	Stimmt eher nicht	Bin mir unsicher	Stimmt ein bisschen	Stimmt absolut

pos07	Ich bin extrovertiert.	Stimmt nicht	Stimmt eher nicht	Bin mir unsicher	Stimmt ein bisschen	Stimmt absolut

pos08	Zurzeit habe ich Stress (durch Arbeit, Klausuren, etc.)	Stimmt nicht	Stimmt eher nicht	Bin mir unsicher	Stimmt ein bisschen	Stimmt absolut

pos09	Ich bin bemüht, mich gesund zu ernähren.	Stimmt nicht	Stimmt eher nicht	Bin mir unsicher	Stimmt ein bisschen	Stimmt absolut

		Stimmt nicht	Stimmt eher nicht	Bin mir unsicher	Stimmt ein bisschen	Stimmt absolut
pos10	Ich bin gerne draußen.					
pos11	Ich schwitze nicht gerne.					
pos12	Ich habe mein Smartphone immer dabei.					
pos13	Ich habe im Alltag ausreichend Zeit für Sport.					
pos14	Ich mache gerne Sport.					
pos15	Ich gehe gerne joggen.					
pos16	Sport ist für mich derzeit aufwändig (Fahrt zum Training, usw.)					
pos17	Ich mag Mannschafts-sportarten lieber.					
pos18	Ich messe mich gerne mit anderen.					
pos19	Ich sollte vermutlich insgesamt mehr Sport treiben.					

pos20	Ich spiele gerne Twostone.	Stimmt nicht	Stimmt eher nicht	Bin mir unsicher	Stimmt ein bisschen	Stimmt absolut

pos21	Ich werde Twostone auch weiterhin spielen.	Stimmt nicht	Stimmt eher nicht	Bin mir unsicher	Stimmt ein bisschen	Stimmt absolut

pos22	Ich glaube, dass Twostone mir dabei helfen kann, aktiv und gesund zu bleiben.	Stimmt nicht	Stimmt eher nicht	Bin mir unsicher	Stimmt ein bisschen	Stimmt absolut

pos23	Ich glaube, dass Twostone anderen helfen könnte, aktiv und gesund zu bleiben.	Stimmt nicht	Stimmt eher nicht	Bin mir unsicher	Stimmt ein bisschen	Stimmt absolut

pos24	Twostone ist ein Spiel für Kinder.	Stimmt nicht	Stimmt eher nicht	Bin mir unsicher	Stimmt ein bisschen	Stimmt absolut

pos25	Twostone ist ein Spiel für Erwachsene.	Stimmt nicht	Stimmt eher nicht	Bin mir unsicher	Stimmt ein bisschen	Stimmt absolut

pos26	Twostone ist ein Spiel für Senioren.	Stimmt nicht	Stimmt eher nicht	Bin mir unsicher	Stimmt ein bisschen	Stimmt absolut

pos27	Ich spiele gerne auf selbsterstellten Karten.	Stimmt nicht	Stimmt eher nicht	Bin mir unsicher	Stimmt ein bisschen	Stimmt absolut

pos28	I spiele gerne auf Karten, die von anderen Spielern erstellt wurden.	Stimmt nicht	Stimmt eher nicht	Bin mir unsicher	Stimmt ein bisschen	Stimmt absolut

pos29	Twostone muss noch stabiler werden (in Bezug auf Nutzer-freundlichkeit & Bugs).	Stimmt nicht	Stimmt eher nicht	Bin mir unsicher	Stimmt ein bisschen	Stimmt absolut

pos30	Ich würde vielleicht ein Spiel wie Twostone spielen, aber es müsste von höher Qualität sein.	Stimmt nicht	Stimmt eher nicht	Bin mir unsicher	Stimmt ein bisschen	Stimmt absolut

pos31	Es ist zu aufwändig, Twostone zu spielen.	Stimmt nicht	Stimmt eher nicht	Bin mir unsicher	Stimmt ein bisschen	Stimmt absolut

pos32	Der Trigger hat genervt.	Stimmt nicht	Stimmt eher nicht	Bin mir unsicher	Stimmt ein bisschen	Stimmt absolut

pos33	Der Trigger hat in der ersten Woche mehr genervt, als später.	Stimmt nicht	Stimmt eher nicht	Bin mir unsicher	Stimmt ein bisschen	Stimmt absolut

pos34	Der Trigger hat gute Gelegenheiten für das Spielen herausgesucht.	Stimmt nicht	Stimmt eher nicht	Bin mir unsicher	Stimmt ein bisschen	Stimmt absolut

pos35	Der Trigger wurde mit der Zeit besser.	Stimmt nicht	Stimmt eher nicht	Bin mir unsicher	Stimmt ein bisschen	Stimmt absolut

pos36	Ich habe den Trigger manchmal nicht bemerkt.	Stimmt nicht	Stimmt eher nicht	Bin mir unsicher	Stimmt ein bisschen	Stimmt absolut

pos37	Ich habe den Trigger manchmal ignoriert.	Stimmt nicht	Stimmt eher nicht	Bin mir unsicher	Stimmt ein bisschen	Stimmt absolut

pos38	Ohne den Trigger hätte ich weniger Twostone gespielt.	Stimmt nicht	Stimmt eher nicht	Bin mir unsicher	Stimmt ein bisschen	Stimmt absolut

pos39	Der Trigger muss noch intelligenter werden.	Stimmt nicht	Stimmt eher nicht	Bin mir unsicher	Stimmt ein bisschen	Stimmt absolut

		Stimmt nicht	Stimmt eher nicht	Bin mir unsicher	Stimmt ein bisschen	Stimmt absolut
pos40	Der Trigger muss noch stabiler werden (in Bezug auf Nutzer-freundlichkeit & Bugs).					

Und das war’s.

Bitte schicke den ausgefüllten Fragebogen zurück an twostonestudy@gmail.com.

Danke für Deine Unterstützung! :)

Appendix D – Author’s Publications

MAIN PUBLICATIONS

[DGH+16] Tim Dutz, Augusto Garcia-Agundez, Sandro Hardy, Stefan Göbel, Ralf Steinmetz. Towards effective interventive health applications – on the problem of user triggering. In: Proceedings of the HCI International ‘16, Springer, Berlin, Germany, 2016.

[DHK+14] Tim Dutz, Sandro Hardy, Martin Knöll, Stefan Göbel, Ralf Steinmetz. User interfaces of mobile exergames. In: Proceedings of the HCI International ‘14, Springer, Berlin, Germany, 2014.

[KDH+14] Martin Knöll, Tim Dutz, Sandro Hardy, Stefan Göbel. Urban exergames – how architects and serious gaming researchers collaborate on digital games that make you move. In: Virtual and Augmented Reality in Healthcare, Vol. 1, Springer, Berlin, Germany, 2014.

[DKH+13] Tim Dutz, Martin Knöll, Sandro Hardy, Stefan Göbel. Calm gaming – how mobile devices could change the face of serious gaming. In: i-com, Vol. 12, No. 2, 2013.

OTHER PUBLICATIONS

[GDG16] Augusto Garcia-Agundez, Tim Dutz, Stefan Göbel. Adapting smartphone-based photoplethysmography to suboptimal conditions. To be published, 2016.

[GSD+16] Augusto Garcia-Agundez, Saroj Sharma, Tim Dutz, Stefan Göbel. Ein smartphonebasiertes Framework für Patientenfernüberwachung. In: Proceedings of Mensch und Computer ’16, Oldenbourg Verlag, Munich, Germany, 2016.

[BD16] Andreas Braun, Tim Dutz. Low-cost indoor localization using cameras – evaluating AmbiTrack and its applications in Ambient Assisted Living. In: Journal of Ambient Intelligence and Smart Environments, Vol. 8, No. 3, IOS Press, 2016.

[HFD+15] Sandro Hardy, Florian Feldwieser, Tim Dutz, Stefan Göbel, Ralf Steinmetz, Elisabeth Steinhagen-Thiessen. ALFRED Back Trainer – conceptualization of a serious game-based training system for low back pain rehabilitation exercises. In: Proceedings of the First Joint International Conference for Serious Games, Springer, Berlin, Germany, 2015.

[HDW+14-a] Sandro Hardy, Tim Dutz, Josef Wiemeyer, Stefan Göbel, Ralf Steinmetz. Adaptation und Personalisierung in RehaGames. In: Neurologie & Rehabilitation, Vol. 4, 2014.

[HDW+14-b] Sandro Hardy, Tim Dutz, Josef Wiemeyer, Stefan Göbel, Ralf Steinmetz. Framework for personalized and adaptive game-based training programs in health sport. In: Multimedia Tools and Applications, Vol. 74, No. 14, 2014.

[HKD+14] Sandro Hardy, Angelika Kern, Tim Dutz, Christoph Weber, Stefan Göbel, Ralf Steinmetz. What makes games challenging? Considerations on how to determine the ‘challenge’ posed by an exergame for balance training. In: Proceedings of the 2014 ACM International Workshop on Serious Games, ACM, 2014.

[KKN+14] Martin Knöll, Johannes Konert, Katrin Neuheuser, Sandro Hardy, Tim Dutz, Michael Gutjahr, Annette Rudolph-Cleff, Joachim Vogt, Stefan Göbel. Interdisciplinary course on urban health games. In: Proceedings of the 1st Workshop on Smart City Learning, European Conference on Technology Enhanced Learning (EC-TEL), 2014.

[GMW+14] Stefan Göbel, Florian Mehm, Viktor Wendel, Johannes Konert, Sandro Hardy, Christian Reuter, Michael Gutjahr, Tim Dutz. Erstellung, Steuerung und Evaluation von Serious Games. In: Informatik Spektrum, Vol. 37, No. 6, 2014.

[MDW14] Martin Majewski, Tim Dutz, Reiner Wichert. An optical guiding system for gesture based interactions in smart environments. In: Proceedings of the HCI International '14, Springer, Berlin, Germany, 2014.

[DMM+14] Tim Dutz, Martin Majewski, Stefanie Müller, Andreas Braun, Johannes Konert, Felix Kamieth, Henrik Rieß, Antonija Mrsic Carl, Denise Bender, Verena Brückner, Peter Klein, Reiner Wichert, Stefan Göbel, Ralf Steinmetz. inDAgo – ein Mobilitätsunterstützungssystem für Senioren auf dem Weg in die Praxis. Tagungsband des 8. AAL-Kongresses des VDE, 2014.

[KDH+13] Martin Knöll, Tim Dutz, Sandro Hardy, Stefan Göbel. Active design – how the built environment matters to mobile games for health. In: Proceedings of the FROG '13, Vienna, Austria, 2013.

[BD13] Andreas Braun, Tim Dutz. AmbiTrack – indoor localization and tracking of multiple users in smart environments with a camera-based approach. In: Evaluating AAL Systems Through Competitive Benchmarking, Springer, Berlin, Germany, 2013.

[SGD+13] Carsten Stocklöw, Andrej Grguric, Tim Dutz, Tjark Vandommele, Arjan Kuijper. Resource management for multimodal and multilingual adaptation of user interfaces in Ambient Assisted Living environments. In: Proceedings of the HCI International '13, Springer, Berlin, Germany, 2013.

[MDK+13] Alexander Marinc, Tim Dutz, Felix Kamieth, Maxim Djakow, Pia Weiss. Creating rule sets for smart environments through behavior recording. In: Proceedings of the HCI International '13, Springer, Berlin, Germany, 2013.

[BDK13] Andreas Braun, Tim Dutz, Felix Kamieth. Capacitive sensor-based hand gesture recognition in ambient intelligence scenarios. In: Proceedings of the 6th International Conference on Pervasive Technologies Related to Assistive Environments, No. 5, ACM, 2013.

[BDA+13] Andreas Braun, Tim Dutz, Michael Alekseew, Philipp Schillinger, Alexander Marinc. Marker-free indoor localization and tracking of multiple users in smart environments using a camera-based approach. In: Proceedings of the HCI International '13, Springer, Berlin, Germany, 2013.

[KMR+13] Peter Klein, Stefanie Müller, Henrik Rieß, Felix Kamieth, Tim Dutz, Steffi Hußlein. Methoden und Möglichkeiten zur Akzeptanzsteigerung von Mobilitäts-Assistenzsystemen am Beispiel des Forschungsprojekts inDAgo. In: Lebensqualität im Wandel von Demografie und Technik - 6. Deutscher AAL-Kongress mit Ausstellung, VDE-Verlag, 2013.

[KDW+13] Felix Kamieth, Tim Dutz, Pia Weiss, Stefanie Müller, Christian Reuter, Reiner Wichert, Peter Klein, Stefan Göbel. Navigationsassistenz für ältere Menschen im öffentlichen Nahverkehr. In: Lebensqualität im Wandel von Demografie und Technik - 6. Deutscher AAL-Kongress mit Ausstellung, VDE Verlag, 2013.

[KMB+12] Kerstin Klauß, Stefanie Müller, Andreas Braun, Tim Dutz, Felix Kamieth, Peter Klein. Synergieeffekte aus der Kombination verschiedener AAL Lösungen. Proccedings of Mensch und Computer ‘12, Oldenbourg Verlag, Munich, Germany, 2012.

[LMD+12] Myriam Lipprandt, Alexander Marinc, Tim Dutz, Guido Moritz, Marco Eichelberg, Reiner Wichert, Andreas Hein. Evaluation of AAL middleware platforms. In: Proceedings of the 4th AAL Forum Eindhoven, The Netherlands, 2012.

Appendix E – Supervised Students' Theses

MASTER THESES

[KOM-M-0575]	Johny George Malayil. Smartphone-based emotional response analysis. Master thesis, TU Darmstadt, Germany, 2016.
[KOM-M-0573]	Johannes Heucher. Dynamic content creation for pervasive games. Master thesis, TU Darmstadt, Germany, 2016.
[KOM-M-0572]	Jens-Uwe Sperling. Implementation and evaluation of varied indicator sets for pervasive user triggers. Master thesis, TU Darmstadt, Germany, 2016.
[KOM-M-0544]	Oliver Welther. Development of a mobile modular context awareness framework. Master thesis, TU Darmstadt, Germany, 2016.
[KOM-M-0543]	Alexander Schmitt. Interventive applications – conceptualization of an intelligent user trigger for older adults. Master thesis, TU Darmstadt, Germany, 2016.
[KOM-M-0535]	Tiange Hu. User triggering with smartphone sensors. Master thesis, TU Darmstadt, Germany, 2015.
[KOM-M-0503]	Augusto Garcia. Smartphone-based photoplethysmography under suboptimal conditions – adapting to real life scenarios. Master thesis, TU Darmstadt, Germany, 2015.
[KOM-M-0498]	Patrick Hock. Augmented reality in pervasive urban gaming. Master thesis, TU Darmstadt, Germany, 2014.
[KOM-M-0491]	Philipp Dürr. Heart rate sensitive gaming. Master thesis, TU Darmstadt, Germany, 2014.

BACHELOR THESES

[KOM-B-0549]	Paul Schweiger. An analysis of the motivational mechanics of mobile exergames. Bachelor thesis, TU Darmstadt, Germany, 2016. (Joined supervision with Dipl.-Psych. Michael Gutjahr)
[KOM-B-0541]	Gabriela Marques. Design and implementation of an iOS-based framework for user activity recognition and cross-platform communication. Bachelor thesis, TU Darmstadt, Germany, 2016.
[KOM-B-0540]	Viktoria Swiatkowski. Comparing analysis of user triggering concepts in technical environments. Bachelor thesis, TU Darmstadt, Germany, 2016.
[KOM-B-0522]	Glen Wang. Eye tracking and gaze interaction on smartphones. Bachelor thesis, TU Darmstadt, Germany, 2015. (Joined supervision with Laila Shoukry, M.Sc.)
[KOM-B-0521]	Philipp von Bauer. Adapting game parameters by using physiological user experience measurement. Bachelor thesis, TU Darmstadt, Germany, 2015. (Joined supervision with Dipl.-Psych. Michael Gutjahr.)
[KOM-B-0518]	Nikolay Dimitrov. User ability adaptation in mobile exergames. Bachelor thesis, TU Darmstadt, Germany, 2015.

[KOM-B-0517] Kevin Mais. Context-aware triggering for mobile exergames. Bachelor thesis, TU Darmstadt, Germany, 2015.

[KOM-B-0516] Aline Praetorius. Conception and implementation of a visualization tool for psychophysiological information. Bachelor thesis, TU Darmstadt, Germany, 2015. (Joined supervision with Dipl.-Psych. Michael Gutjahr.)

[KOM-B-0512] Stefan Wegener. Design and development of a smartphone camera-based sport performance analysis tool for application in bouldering training sessions. Bachelor thesis, TU Darmstadt, Germany, 2015.

[KOM-B-0503] Polona Caserman. Development of an adaptive learning game for autistic children. Bachelor thesis, TU Darmstadt, Germany, 2015. (Joined supervision with Laila Shoukry, M.Sc.)

[KOM-B-0492] Tobias Welther. Holding hand recognition for smartphones. Bachelor thesis, TU Darmstadt, Germany, 2014.

[KOM-B-0491] Oliver Welther. 3D-gesture recognition for smartphones. Bachelor thesis, TU Darmstadt, Germany, 2014.

[KOM-B-0489] Chris Michel. Ad hoc creation of game areas for location-based exergames. Bachelor thesis, TU Darmstadt, Germany, 2014.

[KOM-B-0488] Gerhard Säckel. Augmented reality interfaces for location-based exergames. Bachelor thesis, TU Darmstadt, Germany, 2014.

STUDIENARBEITEN

[KOM-B-0526] Martin Möller. Design and implementation of user interaction mechanisms and information visualization tools for a multiplayer strategy game. Studienarbeit, TU Darmstadt, Germany, 2015.

[KOM-B-0520] Lukas Fey. Design and implementation of a multiplayer engine for a cross-platform strategy game. Studienarbeit, TU Darmstadt, Germany, 2015.

[KOM-B-0519] Thomas Tregel. Design of a multiplayer framework for mobile exergames. Studienarbeit, TU Darmstadt, Germany, 2015.

Appendix F – Curriculum Vitae

PERSONAL INFORMATION

Full Name	Tim Alexander Dutz
Nationality	German
Date of Birth	July 20th, 1978
Place of Birth	Darmstadt, Germany
Marital Status	Married, one child

ACADEMIC CAREER

2013 – today	Research Associate / Ph.D. Student Department of Electrical Engineering Technische Universität Darmstadt, Germany R&D projects: ALFRED (EU), Smastra (German BMBF), UHG (TU Darmstadt) Management: supervisor for a total of 14 student assistants (in 3 teams) Honors: participant of the exclusive *Software Campus* program
2010 – 2013	Research Associate Fraunhofer IGD Darmstadt, Germany R&D projects: universAAL (EU), inDAgo / optimAAL / RAALI (all Ger. BMBF)
2002 – 2010	Student (acquired degree: Dipl.-Inf. Univ.) Department of Computer Science Technische Universität München, Germany
1998 – 2002	Student (no degree) Department of History and Philosophy / Department of Economics Technische Universität Darmstadt, Germany
1997 – 1998	Military Service
1988 – 1997	Student (acquired degree: university entrance qualification) Ludwig-Georgs-Gymnasium LGG Darmstadt, Germany

TEACHING ACTIVITIES

2013 – today	Thesis supervisor for a total of 26 students (see appendix E for the full list)
2013 – today	'Serious Games Seminar' supervisor for a total of 22 students
2013 – today	'Serious Games Lab Course' supervisor for a total of 41 students (in 14 groups)
2013 – today	Contribution to lecture series 'Serious Games' for a total of 4 lectures
2013 – 2015	Contribution to lecture series 'Urban Health Games' for a total of 3 lectures
2010 – 2013	Contribution to lecture series 'Ambient Intelligence' for a total of 4 lectures

Appendix G – Erklärung laut §9 der Promotionsordnung

Ich versichere hiermit, dass ich die vorliegende Dissertation allein und nur unter Verwendung der angegebenen Literatur verfasst habe.

Die Arbeit hat bisher noch nicht zu Prüfungszwecken gedient.

Darmstadt, 14. September 2016