Moving forward

Supporting physical activity behavior change through intelligent technology

Julia S. Mollee
Promotor:

prof. dr. Jan Treur
Vrije Universiteit Amsterdam
Behavoural Informatics

Co-promotores:

dr. Michel C. A. Klein
Vrije Universiteit Amsterdam
Behavoural Informatics

dr. Saskia J. te Velde
VU medical center
Epidemiology & Biostatistics

Reading committee:

prof. dr. ir. Aart T. van Halteren
Philips Research /
Vrije Universiteit Amsterdam

dr. Jaap R. C. Ham
Eindhoven University of Technology
Human-Technology Interaction

prof. dr. Susan Michie
University College London
Centre for Behaviour Change

prof. dr. Mark A. Neerincx
Delft University of Technology
Human-Centered Computing

dr. Annette C. M. ten Teije
Vrije Universiteit Amsterdam
Knowledge Representation & Reasoning

This research is financed by the partnership program ‘Healthy Lifestyle Solutions’ (HLS), of technology foundation STW, the National Initiative Brain & Cognition (Nationaal Initiatief Hersenen & Cognitie, NIHC) and Philips Research (grant no. 12014).

SIKS Dissertation Series No. 2018–10
The research reported in this thesis has been carried out under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems.

Cover design by Radúz Mollee.


© 2018 Julia S. Mollee. All rights reserved.
MOVING FORWARD

Supporting physical activity behavior change through intelligent technology
promotor: prof.dr. J. Treur
copromotoren: dr. M.C.A. Klein
dr. S.J. te Velde
Part One

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Motivation</td>
<td>15</td>
</tr>
<tr>
<td>1.2</td>
<td>Context</td>
<td>15</td>
</tr>
<tr>
<td>1.2.1</td>
<td>Active2Gether project</td>
<td>16</td>
</tr>
<tr>
<td>1.2.2</td>
<td>Physical (in)activity</td>
<td>17</td>
</tr>
<tr>
<td>1.2.3</td>
<td>Behavior change</td>
<td>18</td>
</tr>
<tr>
<td>1.2.4</td>
<td>Computer technology and behavior change</td>
<td>21</td>
</tr>
<tr>
<td>1.3</td>
<td>Research objective and research questions</td>
<td>22</td>
</tr>
<tr>
<td>1.3.1</td>
<td>Research question 1</td>
<td>22</td>
</tr>
<tr>
<td>1.3.2</td>
<td>Research question 2</td>
<td>23</td>
</tr>
<tr>
<td>1.3.3</td>
<td>Research question 3</td>
<td>23</td>
</tr>
<tr>
<td>1.3.4</td>
<td>Research question 4</td>
<td>23</td>
</tr>
<tr>
<td>1.4</td>
<td>Methods</td>
<td>24</td>
</tr>
<tr>
<td>1.4.1</td>
<td>Systematic reviews</td>
<td>24</td>
</tr>
<tr>
<td>1.4.2</td>
<td>Qualitative interviews</td>
<td>24</td>
</tr>
<tr>
<td>1.4.3</td>
<td>Computational modeling</td>
<td>25</td>
</tr>
<tr>
<td>1.4.4</td>
<td>User studies</td>
<td>25</td>
</tr>
<tr>
<td>1.4.5</td>
<td>Data analysis</td>
<td>26</td>
</tr>
<tr>
<td>1.4.6</td>
<td>Development and implementation</td>
<td>27</td>
</tr>
<tr>
<td>1.5</td>
<td>Contributions</td>
<td>27</td>
</tr>
<tr>
<td>1.6</td>
<td>Thesis organization</td>
<td>28</td>
</tr>
<tr>
<td>1.6.1</td>
<td>Thesis type</td>
<td>28</td>
</tr>
<tr>
<td>1.6.2</td>
<td>Thesis outline</td>
<td>28</td>
</tr>
<tr>
<td>1.6.3</td>
<td>Chapters</td>
<td>29</td>
</tr>
</tbody>
</table>
# Part Two

## 2 Review of behavior change techniques in apps

### 2.1 Background

### 2.2 Methods

#### 2.2.1 Inclusion criteria

#### 2.2.2 Search strategy

#### 2.2.3 Screening procedure

#### 2.2.4 The taxonomy

#### 2.2.5 Scoring

#### 2.2.6 Extracted data

#### 2.2.7 Analyses

### 2.3 Results

### 2.4 Discussion

### 2.5 Conclusions

## 3 Review of technological features in apps

### 3.1 Introduction

### 3.2 Methods

#### 3.2.1 Identification

#### 3.2.2 Screening

#### 3.2.3 Scoring

#### 3.2.4 Analyses

### 3.3 Results

#### 3.3.1 Overview of included apps

#### 3.3.2 Number of features per app

#### 3.3.3 Number of apps per feature

#### 3.3.4 Differences between types of apps

#### 3.3.5 Correlation between number of features and usability rating

### 3.4 Discussion

### 3.5 Conclusion

## 4 User preferences for apps that promote physical activity

### 4.1 Background

### 4.2 Methods

#### 4.2.1 Design

#### 4.2.2 Recruitment

#### 4.2.3 Procedures

#### 4.2.4 Data management and analysis
12.4 Results 219
12.4.1 Social comparison preferences 219
12.4.2 Exploratory data analysis 219
12.4.3 Hypothesis H1: Preferred upward, presented upward social comparison 219
12.4.4 Hypothesis H2: Preferred upward, presented downward social comparison 219
12.4.5 Hypothesis H3: Preferred downward, presented downward social comparison 220
12.4.6 Hypothesis H4: Preferred downward, presented upward social comparison 221
12.4.7 Hypothesis H5: Preferred vs. non-preferred social comparison 221
12.5 Discussion 223
12.6 Conclusion 224

V Part Five

13 Description of the development and content of Active2Gether 231
13.1 Introduction 233
13.2 Methods 235
13.2.1 Intervention development 235
13.2.2 Target population 236
13.3 Development of the Active2Gether intervention 237
13.3.1 Step 1: Defining the intervention's primary and secondary objectives 237
13.3.2 Step 2: Defining the theoretical framework 237
13.3.3 Step 3: Developing the Active2Gether intervention 238
13.3.4 Step 4: Pilot testing 248
13.3.5 Step 5: Testing and evaluating the intervention 248
13.4 Summary of the Active2Gether intervention 248
13.5 Discussion 249

14 Description of the technical design of Active2Gether 263
14.1 Introduction 265
14.2 System description 266
14.2.1 Data collection 266
14.2.2 Architecture 267
14.2.3 Layout and visual design 269
14.2.4 Model-based reasoning 269
14.2.5 Selection and tailoring of coaching messages 274
14.2.6 Social comparison 275
14.3 Discussion 277
14.3.1 Data collection 277
14.3.2 Architecture 278
14.3.3 Layout and visual design 278
14.3.4 Model-based reasoning 279
14.3.5 Selection and tailoring of coaching messages 279
14.4 Conclusions and future work

15 Exploring use and effects of Active2Gether

15.1 Background

15.2 Methods

15.3 Results

15.4 Discussion

15.5 Conclusion

16 Evaluation of user appreciation and adherence of Active2Gether

16.1 Introduction

16.2 Active2Gether system

16.3 Methods

16.4 Results
### Part Six

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>Discussion &amp; Conclusion</td>
<td>339</td>
</tr>
<tr>
<td>17.1</td>
<td>Summary of research contributions</td>
<td>339</td>
</tr>
<tr>
<td>17.2</td>
<td>Discussion of research questions and results</td>
<td>340</td>
</tr>
<tr>
<td>17.2.1</td>
<td>Research question 1</td>
<td>341</td>
</tr>
<tr>
<td>17.2.2</td>
<td>Research question 2</td>
<td>343</td>
</tr>
<tr>
<td>17.2.3</td>
<td>Research question 3</td>
<td>345</td>
</tr>
<tr>
<td>17.2.4</td>
<td>Research question 4</td>
<td>348</td>
</tr>
<tr>
<td>17.3</td>
<td>Ethical aspects</td>
<td>350</td>
</tr>
<tr>
<td>17.3.1</td>
<td>Non-maleficence and beneficence</td>
<td>350</td>
</tr>
<tr>
<td>17.3.2</td>
<td>User-centricity</td>
<td>351</td>
</tr>
<tr>
<td>17.3.3</td>
<td>Multiple user groups</td>
<td>351</td>
</tr>
<tr>
<td>17.3.4</td>
<td>Privacy</td>
<td>351</td>
</tr>
<tr>
<td>17.3.5</td>
<td>Data protection</td>
<td>352</td>
</tr>
<tr>
<td>17.3.6</td>
<td>Security</td>
<td>352</td>
</tr>
<tr>
<td>17.3.7</td>
<td>Autonomy</td>
<td>352</td>
</tr>
<tr>
<td>17.3.8</td>
<td>Transparency</td>
<td>353</td>
</tr>
<tr>
<td>17.3.9</td>
<td>Equality, dignity and inclusiveness</td>
<td>353</td>
</tr>
<tr>
<td>17.4</td>
<td>Future work</td>
<td>354</td>
</tr>
<tr>
<td>17.5</td>
<td>Conclusion</td>
<td>356</td>
</tr>
</tbody>
</table>

### Part Seven

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Summary</td>
<td>363</td>
</tr>
<tr>
<td></td>
<td>Samenvatting</td>
<td>365</td>
</tr>
<tr>
<td></td>
<td>Shrnutí</td>
<td>367</td>
</tr>
<tr>
<td></td>
<td>Acknowledgments</td>
<td>369</td>
</tr>
<tr>
<td></td>
<td>List of publications</td>
<td>371</td>
</tr>
<tr>
<td></td>
<td>SIKS dissertation series</td>
<td>375</td>
</tr>
</tbody>
</table>
Introduction

1.1 Motivation
1.2 Context
1.3 Research objective and research questions
1.4 Methods
1.5 Contributions
1.6 Thesis organization
1. Introduction

Physical inactivity is an increasingly serious health problem. On a personal level, it is associated with a higher risk of many non-communicable diseases (such as cardiovascular disease, cancer and diabetes) and it increases the likelihood of developing depression or anxiety. On a societal level, the increased prevalence of physical and mental illness through physical inactivity implies a growing burden on the health care system. The large number and the popularity of health & fitness apps available in the app stores indicate that many people are interested in using such apps to change their lifestyle for the better. The rise of mobile technology and wearables offers new opportunities: to make these apps more informed, more tailored, more adaptive and overall smarter or more ‘intelligent’. The research presented in this thesis investigates different aspects of using technology to stimulate behavior change for physical activity, and it contributes to the development of an intelligent physical activity promotion app.

The remainder of this chapter is organized as follows. Section 1.1 explains the overall motivation behind the research described in this thesis. In Section 1.2, the background to which this research was done is described. Section 1.3 formulates the research objective and research questions. In Section 1.4, the methods applied throughout the chapters of this thesis are outlined. The scientific contributions of the work presented in this thesis are described in Section 1.5. Finally, Section 1.6 explains how the remainder of this thesis is organized.

1.1 Motivation

Physical activity is an important prerequisite for global health. Despite the well-known benefits for both physical and mental health (Eime et al., 2013; I.-M. Lee et al., 2012; Paluska and Schwenk, 2000; Reiner et al., 2013; World Health Organization, 2010), approximately 50% of the adult population in western countries are less physically active than recommended by health authorities (World Health Organization, 2014b). This has serious consequences: insufficient engagement in moderate to vigorous physical activity has been associated with increased risks of cardiovascular diseases, cancer, diabetes and mental illness (I.-M. Lee et al., 2012). Research has shown that physical activity levels decrease with age, in particular when transitioning from adolescence into adulthood (Bell and C. Lee, 2005; Kwan et al., 2012). Therefore, effective and engaging interventions are needed to increase and maintain physical activity levels, with a special focus on young adults.

It is believed that modern (mobile) technology provides an opportunity to support people to become or remain physically active (Knight et al., 2015; Payne et al., 2015). After all, smartphones and smartphone applications (apps) are well intertwined in modern society
Introduction

Chapter 1. Introduction

(Center, 2016; TelecomNieuwsNet, 2016), always accessible to the user, and they can lower
the barrier for people to address their health problems (Griffiths et al., 2006). Also, initial
evidence indicates that such mobile interventions are effective (Stephens and Allen, 2013).
However, although physical activity can be increased significantly through simple solutions
as self-monitoring and the use of pedometers (Bravata et al., 2007; Lubans et al., 2009), it
is challenged whether these changes are durable over the long term (Gierisch et al., 2015).
Therefore, physical activity interventions based on technological features require a more
solid scientific basis in order to yield long-term effects.

It has been established that social processes play an important role in achieving and
maintaining a healthy lifestyle (Zimmerman and Connor, 1989). These social influences rely
on several motivational mechanisms, such as priming, social norms, observational learning,
social facilitation, social support and social comparison (Bandura, 1998; Buunk et al., 2013;
Cheng et al., 2014; Festinger, 1954; McNeill et al., 2006). When developing effective
physical activity interventions, the potential of such mechanisms should not be overlooked,
but exploited in well-advised and effective ways.

The main research objective of this thesis is to explore the field sketched above, which
addresses physical activity promotion from an evidence-based perspective, thereby applying
techniques from modern (mobile) technology and artificial intelligence as well as scientific
knowledge from human-directed disciplines within psychology and social sciences, while
taking the importance of social processes into account. Because of its interdisciplinary
nature, the subject of this thesis will be addressed from different perspectives and by
incorporating various aspects. This way, a broad picture of endeavors to approach this
subject will be unfold.

1.2 Context

The research described in this thesis is motivated by both contemporary societal issues, state
of the art scientific knowledge about them, and recent advances in technology, which allows
a new perspective on ways to solve those issues. In order to position the work in this thesis,
it is important to sketch the background to which this research was done. Therefore, this
section is devoted to delineating certain aspects of its context.

First, since this work is part of a bigger research project (called Active2Gether), Section
1.2.1 describes the goals of that project in more detail. In Section 1.2.2, the prevalence
of physical inactivity in society that partly motivated this research is explained. Then, Section
1.2.3 elaborates on elements that play a role in behavior change. Finally, Section 1.2.4
sketches the developments in the use of computer technology in behavior change initiatives.

1.2.1 Active2Gether project

Active2Gether is a research project financed by the partnership program Healthy Lifestyle
Solutions (HLS) of technology foundation STW, the National Initiative Brain & Cognition
(Nationaal Initiatief Hersenen & Cognitie, NIHC) and Philips Research (STW, 2012). It
aims at developing an innovative solution that supports people to be more physically active.
More specifically, the project focuses on young adults, as physical activity levels are known
to decline rapidly in this age group (Bell and C. Lee, 2005; Kwan et al., 2012). Two
core components of the Active2Gether project are the use of modern mobile technology to
unobtrusively monitor the users’ behavior, scientific knowledge about such behaviors and
the employment of social processes to support the users in achieving or maintaining healthy activity levels.

The Active2Gether project is a research collaboration between the Behavioural Informatics group in the Artificial Intelligence section at the Vrije Universiteit in Amsterdam and the Department of Epidemiology and Biostatistics of the EMGO+ Institute for Health and Care Research at the VU Medical Center. The expertise of the two disciplines involved is reflected in the application of principles from artificial intelligence and the focus on evidence-based behavior change techniques.

Altogether, the ingredients of the Active2Gether project were combined in an innovative behavior change intervention. We developed a smartphone app aimed at young adults that coaches them to increase or maintain their physical activity levels in daily life, by incorporating social influences and processes and by using intelligent reasoning mechanisms (including context awareness, personalized feedback, model-based prediction and tailoring). This led to new insights regarding the use of artificial intelligence technology for automated coaching and practical lessons learnt regarding their development and implementation, as well as an extensive and rich dataset that allows us to gain insight in different aspects of physical activity behavior.

1.2.2 Physical (in)activity

Despite the well-known health benefits of physical activity (I.-M. Lee et al., 2012; World Health Organization, 2010), 23 percent of the adult population worldwide did not meet the recommended guidelines in 2010 (World Health Organization, 2014b). Moreover, engagement in moderate to vigorous physical activity decreases with age, in particular when transitioning from adolescence into (young) adulthood (Bell and C. Lee, 2005; Kwan et al., 2012). Systematic reviews reported that levels of physical activity in Europe vary across countries, ranging from 15.6 percent in Israel to 84.8 percent in Slovakia who met the guidelines (Marques et al., 2015). In the Netherlands, approximately one third of the adult population does not meet the Dutch guidelines for healthy physical activity (Wendel-Vos, 2014). Moreover, compared to other European countries, Dutch people lead a relatively sedentary lifestyle, with over 60 percent spending at least 5.5 hours sitting on a usual day and 25 percent at least 8.5 hours (TNS Opinion & Social, 2014).

One contributing factor to the increasingly sedentary lifestyles is that people are more inclined to make use of passive modes of transportation. Active traveling modes such as biking and walking can contribute to a healthy level of physical activity (Sahlqvist et al., 2012). Another explanation for sedentary lifestyles is related to the work environment, as many people work in offices and carry out their responsibilities sitting at a desk. Research suggests that having desk jobs increases health risks up to 50 percent. Integrating brief active interruptions in work routines could help to increase physical activity and lower health risks (Levine, 2014).

Insufficient physical activity is one of the leading risk factors for premature mortality and non-communicable diseases as cardiovascular diseases, cancer and diabetes (I.-M. Lee et al., 2012; World Health Organization, 2014b). In addition, physical inactivity is associated with osteoporosis, high blood pressure, high blood cholesterol and obesity (World Health Organization, 2014b). The WHO estimates that a lack of physical activity leads to 3.2 million deaths per year globally (World Health Organization, 2009, 2014b). Related, overweight and obesity cause more deaths today than underweight (World Health Organization, 2004).
On a personal level, the health consequences of physical inactivity may lead to a decrease in quality of life and a decrease in life expectancy. Also, related health care costs may pose a financial burden on patients. For organizations, physical inactivity can have an adverse effect, since it can diminish employee productivity by causing both absenteeism (when employees do not come to work) and presenteeism (when employees come to work, but are not fully functioning). This causes a direct economic loss for employers. On a global level, research has shown that the economic burden of physical inactivity amounted to $67.8 billion in 2013 (Ding et al., 2016). This includes both direct costs for health care expenditure and indirect costs through productivity losses, and because of the conservative methodology of the researchers, it is likely to even be an underestimation.

### Behavior change

Since over two-thirds of the total number of global deaths is caused by noncommunicable diseases (World Health Organization, 2014a), healthy lifestyles could contribute significantly to preventing these premature deaths. Therefore, an important effort to control noncommunicable diseases is to focus on modifiable behavioral factors (such as tobacco/alcohol use, excess salt intake and physical inactivity) to decrease the risk of these diseases (World Health Organization, 2014a). This requires a change in behavior, which has shown to be difficult (Bouton, 2014). This section highlights some of the personal, social and environmental factors that play a role in behavior change for physical activity.

#### Personal factors

Personal characteristics that influence behavior change for healthy lifestyles have been identified and documented in many different theories of behavior change. Some factors appear in multiple theories, whereas some others are more specific to certain theories. This section introduces some of the most prominent relevant personal factors.

Self-efficacy is a construct that is incorporated (in some form) in many theories within the disciplines addressing behavior change, such as the social-cognitive theory, the theory of planned behavior, the health action process approach and the health belief model (Ajzen, 1991; Bandura, 2004; Rosenstock, 1974; Schwarzer, 2002). Self-efficacy represents an individual’s confidence in their own ability to perform a task or to achieve certain goals. It is believed to influence how high people set their goals, but it can also moderate the process of turning intentions into actions. The level of self-efficacy is influenced by prior performance on the same or a related task: a positive evaluation of one’s own behavior helps to build trust in the efficacy, whereas a feeling of failure undermines it. Additionally, feelings of self-efficacy can be influenced by social persuasion or vicarious learning. The health action approach distinguishes different types of self-efficacy for different stages in the behavior change process (Schwarzer, 2002). In the theory of planned behavior, self-efficacy is represented by the concept of ‘perceived behavioral control’ (Ajzen, 1991).

One’s intentions are another prevalent personal factor in behavior change. They indicate the extent to which an individual is ready and willing to perform a given behavior. Intentions play a key role in, for example, the health action process approach, the social cognitive theory and the theory of planned behavior (Ajzen, 1991; Bandura, 2004; Schwarzer, 2002), as they are believed to be an immediate antecedent of behavior. One’s intentions can be influenced by many other factors, such as the self-efficacy, perceived barriers and facilitators to perform the given behavior, and social or personal norms.
Another factor that is believed (in some theories) to influence the intentions is the concept of *outcome expectations*. Outcome expectations represent the expected benefits of performing the behavior, whether it is on a personal, physical or social level. In the theory of planned behavior, this notion is represented by the ‘attitude’ towards the behavior (Ajzen, 1991), and the health belief model speaks of ‘perceived benefits’ (Rosenstock, 1974). In addition to affecting the intentions, some of the theories also document another effect of the outcome expectations, such as to the behavior directly (social cognitive theory), to the perceived behavioral control or to the subjective norm (theory of planned behavior).

According to some theories, *perceived barriers* (or *impediments*) and *facilitators* (or *resources*) play a role in the formation of intentions as well (Bandura, 2004; Rosenstock, 1974; Schwarzer, 2002). Although these factors are quite often external phenomena (for example, impediments such as foul weather, lack of time, inept opening hours of some facility), it is the individual’s *perception* of how insurmountable they are that determines their influence on the behavior change process. The perception of barriers is again influenced by the self-efficacy: the more confident people are, the more likely they will assess their inhibiting conditions as surmountable. Facilitators work in the exact opposite direction, with a stronger perception of facilitators leading to easier execution of the behavior.

An important step in achieving behavior change is the translation of intentions into actual behavior, and thereby overcoming any perceived barriers. This requires *self-regulation skills* (Bandura, 1991; Baumeister and Heatherton, 1996). In the health action approach, a distinction is made between action planning and coping planning skills (Schwarzer, 2002). The former is related to initiating behavior, whereas the second represents one’s ability to overcome obstacles. In the self-regulation theory, four different components of self-regulation are distinguished: *standards* of some desirable behavior, *monitoring* by comparing the self to the standard, a source of self-regulatory strength (i.e., *willpower*), and *motivation* to meet the standards (Baumeister and Vohs, 2007).

**Social factors**

Personal factors are not the only determinants of behavior, as attitudes and behaviors are also influenced by interactions with other people. Therefore, social processes play an important role in achieving and maintaining a healthy lifestyle, and should not be overlooked when investigating behavior change (Zimmerman and Connor, 1989). Many mechanisms that underlie such social factors have been identified (e.g., Bandura, 1998; Cheng et al., 2014; McNeill et al., 2006), and it is beyond the scope of this chapter to discuss them all. Therefore, this section highlights a number of social factors that are relevant to behavior change.

*Social norms* are the expectations for a certain behavior that are formed by interactions with other individuals. These norms can influence one’s behavior through both social and self-evaluation (Bandura, 2004). Complying with normative behavior leads to social acceptance and praise, which reinforces the behavior. In contrast, deviating from the norms leads to social disapproval. Similarly, meeting social standards for behavior gives a positive internal reaction, whereas violating the norms leads to negative self-evaluation, thereby reinforcing or undermining certain behavior.

*Social facilitation* is the tendency for certain tasks or behaviors to be easier in the presence of others (Harkins, 1987). The effect occurs when being observed while performing a task (especially when the task is simple) (Bond and Titus, 1983), or when carrying out a task together. Therefore, it may be easier to adopt and internalize new behavior when doing
it together than when trying to achieve behavior change alone.

Adopting new behavior is also easier when people feel supported by their social environment. Such social support exists in different forms (Langford et al., 1997). For instance, instrumental support encompasses very practical support to help others to achieve their goals, for example by providing financial aids or material goods. Informational support can also be practical, but relates to offering useful information, advice and guidance, rather than tangible help. A third form is emotional support, which stands for expressions of empathy and encouragement.

Another mechanism that plays a part in behavior change is social comparison (Buunk et al., 2013; Festinger, 1954). Social comparison is the act of comparing one’s own behavior (or behavioral outcomes) to another individual. It exists in two variants: downward social comparison and upward social comparison (Festinger, 1954), depending on whether the target (with whom one compares oneself) performs worse (i.e., downward) or better (i.e., upward) than the individual. Both variants can be effective and encouraging, for instance by boosting one’s self-view or by motivating improvement, but also counter-effective and discouraging, for instance by advocating inferior standards or by threatening the self-view (Corcoran et al., 2011). The strength of the effect of social comparison depends on several factors, including one’s closeness and similarity to the other individual, or one’s closeness to some relevant performance metric (for example, number one position) (Garcia et al., 2013). When designed carefully and applied correctly, social comparison can be an effective tool for empowering people to achieve behavior change.

In addition to the specific mechanisms discussed above, we can also look at the effect of the social environment from a social network perspective. It is known that emotions, attitudes and behaviors can spread through social networks via network ties, by means of a process called social contagion (Christakis and Fowler, 2013). Even though it also works for less desirable phenomena, such as obesity (Christakis and Fowler, 2007) and violence (Fagan et al., 2007), social contagion can also play a role in beneficial health behaviors. When trying to achieve behavior change, focusing on changing the behavior of specific actors in a social network could lead to a more widespread change in the network, for example because these actors have a large number of connections (Valente and Pumpluan, 2006), or because they hold key positions in the network, thereby strongly influencing the diffusion to other parts of the community (Borgatti, 2006).

**Environmental factors**

Although some behavior change theories do lightly touch upon external factors that influence behavior change (through perceived barriers and facilitators or social influences), human behavior might be more influenced by environmental factors than these theories suggest. The ANGELO framework proposed by Swinburn et al. (1999) distinguishes four types of environmental factors: physical, economical, political and sociocultural (Swinburn et al., 1999). Examples of social influences are described in the previous section, although sociocultural factors – as the name suggests – also comprise more abstract cultural influences, such as parenting styles or religious viewpoints.

The physical environment refers to the accessibility of resources and opportunities to perform a given behavior. In the domain of physical activity, it includes, for example, the availability of sports clubs in the neighborhood, the presence of safe and well maintained bike lines, and a building layout of the workplace that encourages walking, for example to
1.2 Context

The physical environment can play a role in nudging people to make healthy lifestyle choices, for example when an entrance naturally leads to a staircase, and the elevators can be found around the corner. The health belief model incorporates such (and other) stimuli as ‘cues to action’ (Rosenstock, 1974).

The economical environment concerns the financial costs associated with performing a desirable behavior. When considering physical activity, one could think of the costs for memberships of sports clubs or fitness centers, for sports equipment, and for example the costs of purchasing and maintaining a bike (as compared to the costs for traveling by car or public transport). In addition, people’s incomes also form part of the economical environment, as well as the overall economic situation of a country or municipality.

The fourth type of environmental influence, the political environment, refers to laws, regulations and policies that enable or facilitate the desirable behavior. In case of physical activity promotion, this could include an employer’s incentive for the employee to bike to work (for example through a program that allows employees to advantageously purchase a bike), or (a discount for) a fitness membership included in the collective labor agreement. Policies on national level could stimulate healthy behavior in a similar manner, or indirectly through making such corporate programs and policies attractive to employers.

In addition to the four types of environmental elements, the ANGEL0 framework also distinguishes different levels of environmental size, namely the microenvironmental settings (e.g., homes, schools, workplaces) and the macroenvironmental sectors (e.g., media, transport systems, sports/leisure industry) (Swinburn et al., 1999). Other researchers have identified three levels: the micro level, which represents the actual surroundings of the behavior (such as homes, workplaces); the meso level, which comprises the characteristics of the neighborhood, community and organizations (such as availability of bike lines); and the macro level, which refers to the societal system that can influence behavior through policies, cultural norms, and et cetera (Booth et al., 2001; Brug et al., 2007).

1.2.4 Computer technology and behavior change

As introduced in Section 1.1, the advances in (mobile) technology have lead to many new opportunities to support behavior change. Not only has the transition from traditional face-to-face coaching sessions and print-based interventions to Internet-based interventions increased the reach of interventions to the masses, it has also opened up possibilities to provide coaching at any time and any place. The advantages of smartphones, smartphone apps and wearables as mobile coaching systems for physical activity (or any other health behavior) are in line with that trend, as they are well intertwined in modern society, always accessible to the user, and because they can lower the barrier for people to address their health problems (Griffiths et al., 2006; Pantelopoulos and Bourbakis, 2010).

Using computer technology in behavior change interventions has many benefits. First of all, smartphones and wearable devices allow for continuous monitoring of users through their built-in sensors. This leads to a theoretically complete picture of the individual’s behavior, as well as the possibility to respond to certain events immediately when they take place. In the domain of physical activity, the most relevant sensors in such devices are accelerometers or pedometers, heart rate sensors and location measurements (such as GPS, or location detection through cell towers of Wi-Fi signals). In addition, online data can be used to enrich the context information about the user. For example, current and forecast weather predictions in the user’s area can be retrieved online, as well as traffic conditions...
and travel routes. Also, the user’s location information can be enriched with GIS data to understand the type of location (e.g., sports club, shopping area, public transport hub).

Second, the computing power of mobile devices (and/or remote systems) enables more intelligent interpretation of the sensor data. For example, interpretation of data from one or a combination of sensors could be used for the detection of activities (such as walking, biking or driving a car). Both Android and iOS nowadays offer built-in mechanisms to detect important locations and travel modes (Apple Inc., 2017; Google LLC, 2017), and some physical activity promotion apps are based on this principle as well (Human.co, 2016).

In addition to sensing and interpreting the user’s behavior, administering a behavior change intervention through the smartphone allows for tailoring of the coaching. This could entail simple personalization features, such as greeting the users with their own name or providing a type of coaching based on a user categorization. However, advances in artificial intelligence provide opportunities for far more complex tailoring. Rather than providing a one-size-fits-all approach, where all (categories of) users receive the same type of feedback and advice, sophisticated reasoning mechanisms can be used to decide which support action to perform for a specific user. The reasoning mechanism could be based on, for example, agent-based models, bayesian inference or rule-based systems. Such underlying models can be developed top-down, from theories and available literature, or in a bottom-up manner, by applying machine learning techniques to collected data.

Continuous information about the user can be used for continuous adaptation of the coaching to the users as well. For example, if (global or personal) parameters play a role in a reasoning mechanism as described above, the available data can be used to tune the parameter values over time, in order to match the users’ behavior as closely as possible.

Finally, the smartphone opens up the possibility to use online social networks as a new source of information and to influence on behavior change. Through online friendship connections, the user’s social environment can be investigated, and it can be used to steer social processes that play a role in behavior change (see also Section 1.2.3).

1.3 Research objective and research questions

The aim of the research presented in this thesis is to investigate and develop methods that can be used to counteract physical inactivity, by combining evidence-based intervention design with principles from artificial intelligence. More specifically, the overall research question addressed in this thesis is as follows:

How can mobile technology and artificial intelligence techniques be applied in the design of a behavior change system that aims to increase physical activity levels in young adults?

This overall research question is expanded into four subquestions. These questions are stated below, accompanied by a brief explanation.

1.3.1 Research question 1

What are requirements for mobile behavior change interventions for physical activity based on the state of the art of such interventions and user preferences of the target population?

This research question is made up of two parts: in order to contribute to the current offer of behavior change interventions for physical activity, it is important to have an
understanding of both the state of the art of such interventions and the intended users’ preferences. With respect to the state-of-the-art physical activity promotion apps, it is interesting to know to what extent existing mobile interventions apply behavior change techniques. Also, the extent to which technological features are applied provides relevant information about the current offer of such interventions. In both cases, exploring this leads to insights about gaps in the state of the art and therefore opportunities to develop an intervention that goes beyond existing interventions. On the other hand, knowledge on user preferences is important to be able to respond to their needs and wishes for such an application, and to thereby improve user engagement and adherence.

### 1.3.2 Research question 2

*What role can dynamic computational models play in the development of an intelligent mobile intervention for physical activity?*

This research question focuses on one area of artificial intelligence, i.e. dynamic computational modeling, and how it can be applied in the domain of behavior change systems. A first step is to actually develop such a dynamic computational model of influences on physical activity behavior, based on theoretical and empirical evidence found in literature. Especially when aiming to deploy such a model in a real-life behavior change intervention, it is important to investigate the validity of the model’s simulation outcomes. However, more practical considerations also come into play, for example whether the simulation outcomes also meet other requirements than validity only.

### 1.3.3 Research question 3

*How can an individual’s social network be used to influence his/her physical activity behavior?*

As explained in Section 1.2.3, social processes (e.g., social contagion, social comparison) play an important role in behavior change. This research question focuses on the analysis of such processes and the development of methods to employ social influences to support behavior change. Analysis of social processes, for example by comparing interventions with or without a social component or by explaining behavioral patterns in data with models of social influence, leads to better understanding of the dynamics involved. A thorough understanding of these dynamics is essential when trying to influence social processes for the benefit of behavior change.

### 1.3.4 Research question 4

*To what extent can the answers to the questions above be used to design, implement, exploit and evaluate a personalized mobile intervention for physical activity promotion?*

As one of the objectives of the Active2Gether project (see Section 1.2.1) is to develop and evaluate a behavior change intervention for physical activity, exclusively theoretical endeavors do not suffice to achieve this goal. This raises a number of different questions and challenges, specifically with respect to practical considerations of designing, implementing and exploiting such an innovative behavior change system. Also, evaluating both the effectiveness and user appreciation reveals important lessons learnt and directions for future research and improvement.
1.4 Methods

The broad scope and the interdisciplinarity of the work presented in this thesis also implies a wide range of different methods applied. Among these are methods that are traditionally typical for social sciences and methods that are more related to computer science. The sections below give an overview of the methods applied throughout this thesis.

1.4.1 Systematic reviews

Systematic reviews are a very suitable means to investigate the state of the art of mobile interventions for physical activity (see research question 1). One option would be to analyze the literature on this topic, but that is still only sparsely available. Therefore, analyzing the actual offer of physical activity apps provides a more insightful overview of the characteristics of currently available apps.

When conducting the review of the selected apps, the information in the app description in the app stores provides a good starting point for the analyses. However, additionally downloading and exploring the apps leads to more complete and more reliable findings. In the reviews in this thesis, the apps are scored on two different aspects: the application of evidence-based behavior change techniques (such as self-monitoring, providing feedback, etc.) and the use of technological features (such as using built-in or external sensors, visualizing aggregations of data, etc.). In the former case (Chapter 2), the scoring is based on an existing taxonomy of behavior change techniques (Abraham and Michie, 2008), slightly adapted to fit the scope of mobile rather than traditional interventions. In the other case (Chapter 3), a custom framework of technological features is devised.

These systematic app reviews serve different purposes. In the first place, the reviews aim to sketch the landscape of currently available apps for physical activity promotion, whether the focus is on the number of evidence-based behavior change techniques implemented or the extent to which technological features are incorporated. At the same time, the findings reveal opportunities for improvement for the next generation of physical activity apps. Lastly, the scores obtained from the review combined with meta information about the apps (for example, its price) allow for identification of connections between certain aspects, which can provide additional insight in the characteristics of these apps.

1.4.2 Qualitative interviews

To investigate the user preferences for physical activity promotion apps (as required to answer research question 1), qualitative interviews provide rich information, that would arguably be hard to obtain from quantitative research methods. While a continuum of interview types are available (i.e., from structured to unstructured), focus group interviews are especially suitable. Since the interviews take place in a group setting, interactions between participants may help to stimulate creativity and to bring new ideas to the surface.

In this thesis, focus group interviews are conducted to assess user preferences (i.e., their requirements and wishes) for physical activity coaching apps. The results are described in Chapter 4. In this case, the interview sessions are preceded by a three-week period in which the participants used an existing physical activity promotion app, in order to give them some relevant experience and as a stimulus to jump start the discussions. In addition, a discussion guide with open-ended questions and a number of prompt statements are created to streamline the discussions and to provoke the participants to share their views. In order
to prevent information getting lost due to participants not willing to share their opinion on a certain topic in the group, they are asked to for written comments at the end of the discussion.

Upon collection of the interview data, the discussions are transcribed verbatim. From these reports, relevant fragments (based on the discussion guide) are selected and various codes and subcodes are created. After further review and rearranging the codes, general themes are identified and organized in a tree diagram. This process enables the discovery of overall patterns in the sizeable set of individual opinions and statements, and thereby facilitates forming a coherent picture of the target users’ preferences. This is an important base when developing a physical activity app, so engagement is not threatened by unfulfilled expectations.

1.4.3 Computational modeling

One of the techniques from artificial intelligence that is studied in this thesis, is dynamic computational modeling. It refers to the use of computer systems to simulate behavior of complex systems, in order to study, predict and better understand them. Research question 2 focuses on different aspects of using computational models in the development of mobile physical activity promotion interventions.

In the creation of computational models, four iterative stages can be distinguished. First, the conceptualization stage deals with identifying relevant concepts and relationships between concepts from literature. In the formalization stage, these concepts and relationships are defined in more detail. For the concepts, this amounts to deciding on a operationalization of the variable (e.g., qualitative or quantitative, range of possible values). For the relationships, the formalization step consists of describing the influences between the concepts over time mathematically. It is understandable that the conceptualization and formalization processes go hand in hand.

The third and fourth stages, simulation and evaluation, are also closely connected. Through simulations of the formalized model, simulation traces are obtained. These simulations can be initialized based on available data or on carefully devised scenarios. Evaluation can be done in different ways. For example, reviewing the simulation traces on face validity provides a first indication of the model’s correctness. In addition, verification of expected patterns in the simulations through so-called ‘property checking’ is a more structured approach. Also, the model can be validated by comparing its simulation outcomes with empirical data.

In this thesis, a computational model of influences on physical activity behavior, together with a preliminary verification, is presented in Chapter 5. Chapter 6 provides an in-depth evaluation of (an updated version of) that same model. A further analysis, taking other evaluation measures than goodness of fit to empirical data into account, is described in Chapter 7. At the same time, Chapter 7 is also a proof-of-concept for a new application of parameter tuning techniques, namely the evaluation of model behavior.

1.4.4 User studies

As the work in this thesis builds towards a practical application, it is important to set up real-life test scenarios to review intermediate design choices, to validate assumptions or models, and eventually to evaluate the final product. Consequently, experiments and user
studies play a substantial role in answering research question 2, research question 3 and research question 4.

In each of these experiments, data is collected from a number of participants (ranging from ten to more than a hundred). Depending on the research question at hand, several types of data are collected. Data is collected unobtrusively where possible, i.e. without the participants’ interference. This increases user friendliness and allows for more continuous data collection, since participants don’t have to actively provide the data. In the scope of this thesis, unobtrusive data collection is achieved, for example, by collecting location information through an app on the smartphone and activity data through an activity tracker (e.g., Fitbit One).

However, some notions are difficult to measure directly. Especially for such types of data, questionnaires are appropriate means to collect information. This is, for example, the case with psychological states, such as one’s feelings of self-efficacy or beliefs with respect to outcome expectations. For the assessment of such constructs, validated questionnaires are used where possible. The same holds for subjective information, for example regarding the participants’ opinions. A third category of data to be collected through questionnaires is information that is technically difficult to measure. In the scope of this thesis, this concerns for example transportation modes (e.g., walking, biking, car, bus, train) and sports participation (e.g., swimming). In order to increase user friendliness, the (short) questionnaires to collect this information are presented conveniently on the smartphone and where possible based on triggers.

During the development of this thesis, several smaller intermediate or pilot experiments were conducted. In Chapter 7, user data is used to analyze the behavior of a computational model of influences on physical activity behavior. Chapter 9 describes a validation of a computational model of social contagion based on collected empirical data. In Chapter 12, data is collected to test a design principle related to social comparison. In addition to the intermediate experiments, a final experiment was conducted in which the Active2Gether intervention was tested. Chapter 15 describes the analyses of the effectiveness of the intervention and an evaluation of the user experiences is presented in Chapter 16. The data collected in this final experiment is also used in Chapter 6 to validate a computational model of influences on physical activity behavior.

1.4.5 Data analysis

In the experiments described in the previous section, many different types of data were collected to answer various research questions. As a consequence, a wide variety of data analysis approaches are required. The most suitable type of data analysis depends heavily on the research question at hand, as well as on the available data. This section illustrates a number of different data analysis approaches applied throughout this thesis.

One important objective pursued with data analysis is model validation. Validation of computational models, as briefly introduced in Section 1.4.3, can be approached in many different ways. This thesis contains two validations of a computational model of social contagion (in Chapter 9 and Chapter 11), and a validation of a computational model of influences on physical activity behavior (in Chapter 6). These chapters present a wide range of tests for the models’ validity (depending on the available data and the focus of the validation), for each of which a suitable statistical approach was selected. Among those are a Mann-Kendall correlation test between predicted and empirical data, a Mann-Whitney test
to compare error sizes between two models, and a Spearman rank correlation test between predicted and empirical changes.

In addition to using data analysis for model validation, this thesis also contains applications of data analysis for other purposes. For each task, appropriate statistical tests are conducted, among which regression models (in Chapter 10), Mann-Kendall tests for significance of trendlines (in Chapter 9 and Chapter 12), a log-rank test on Kaplan-Meier survival curves (in Chapter 16), Anova or Krukal-Wallis tests and Tukey or Mann-Whitney post-hoc tests to examine differences between conditions (in Chapter 16).

1.4.6 Development and implementation

In addition to the scientific methods described in the previous sections, a considerable part of the work behind this thesis consisted of the development and implementation of the Active2Gether system. The challenges that come into play when actually developing such a system are covered by research question 4. The design and development of the intervention is described in Chapter 13, and Chapter 14 describes the technical aspects of the development of the Active2Gether system in detail.

From a technical perspective, the Active2Gether system includes four main components. First, a MySQL database is used to store relevant user data, such as activity data, important locations and their characteristics, assessments of psychological constructs, etc. Second, a set of Python scripts, together called the ‘reasoning engine’, read the database contents to decide on the system’s next actions (e.g., sending the user a supportive coaching message). Third, a Java-based program regularly checks the availability of relevant messages or questions, and takes care of sending them to the user. Fourth, the interaction with the users is done through an Android app: it shows a dashboard with recent activity data, it collects location data, and it presents the questions and messages to the user.

The development of the Active2Gether system required careful formalization of conceptual ideas and integration of its various components, and above all thorough testing and debugging.

1.5 Contributions

The scientific contributions of the research presented in this thesis will be elaborated on in the Discussion & Conclusion (Chapter 17). This section briefly outlines the overall foreseen contributions that warrant the added value of this thesis.

As explained in Section 1.1, the overall aim of the work described in this thesis is to contribute to the field of study that pursues healthy lifestyle promotion by combining knowledge and methods from (health) psychology and artificial intelligence. The integration of these two disciplines implies an interdisciplinary approach, in which each of the perspectives is reflected accurately.

Although the combination of (health) psychology and artificial intelligence is not unique, its application to the actual development of a behavior change intervention for physical activity is relatively unprecedented. To this purpose, a variety of techniques that aim to promote physical activity by using ‘intelligent’ approaches are (developed and) investigated. This produces many insights in the added value of such techniques and their advantages and disadvantages. Some of these insights contribute to the understanding of human behavior and dynamics underlying behavior change, whereas others provide practical directions
for the development of physical activity interventions. One can imagine that some of the investigated techniques should be relatively easily transferable to other aspects of healthy lifestyle (such as sleeping, nutrition, smoking habits and alcohol intake), thereby extending the scientific value to beyond physical activity only.

The interdisciplinary nature of this thesis, as well as the wide variety of techniques investigated on their ability to contribute to behavior change, goes hand in hand with a wide range of methods to approach the overall research objective. As Section 1.4 exemplifies, these include methods that are traditionally typical for social sciences and methods that are more related to computer science.

Finally, another important contribution is the collection of several rich datasets regarding physical activity behavior and a diversity of related factors. These datasets form the basis for answering specific research questions in some of the chapters of this thesis, but the variety of information stored in these datasets allows for a wide range of further analyses. Therefore, various interesting research questions that were beyond the scope or time limit of this work will still be able to be investigated based on this data.

1.6 Thesis organization

This section delineates the structure of this thesis. First, Section 1.6.1 explains the type of this thesis. Then, Section 1.6.2 describes the thesis’ different parts. Finally, each of the chapters (and the publications they are based on) are briefly outlined in Section 1.6.3.

1.6.1 Thesis type

This thesis is a cumulative thesis, which means that it is composed of a collection of independent articles. As some elements had to be explained in multiple papers, this implies that certain notions or ideas are repeated in more than one chapter and therefore some parts may show some overlap. However, this also implies that each of the chapters can be read as a standalone narrative.

1.6.2 Thesis outline

This thesis consists of six parts in total, including the Introduction (Part I) and the Discussion & Conclusion (Part VI). The core of this thesis is organized into four parts (Part II – Part V), each consisting of multiple chapters. Each of these four parts addresses one of the research questions introduced in Section 1.3. Below, the remaining five parts and the overall contents of their chapters are outlined.

Part II: Investigating the state of the art and user preferences

Part II provides an answer to research question 1, as introduced in Section 1.3. It first presents an overview of the state of the art of behavior change interventions that promote a physically active lifestyle. This is done through two systematic reviews and content analyses: one focusing on the use of behavior change techniques (in Chapter 2) and one focusing on the application of technological features (in Chapter 3). In addition, the results of exploring user preferences regarding such interventions through focus group discussions are presented in Chapter 4.
Part III: Modeling influences on physical activity

Part III contributes to answering research question 2. Chapter 5 presents a dynamic computational model of influences on physical activity behavior. Chapter 6 provides a preliminary validation of a refined version of this model, based on empirical physical activity data. In Chapter 7, it is investigated how parameter tuning techniques can be used to get more insight in the behavior of such a computational model. The latter is especially useful when employing a computational model in a real-life application, as in the Active2Gether intervention.

Part IV: Using the social environment to influence behavior

Part IV deals with research question 3. It contains four chapters that –each from a different perspective– focus on the role of social influences on physical activity behavior. Chapter 8 investigates how interventions in the structure of social networks can lead to improved behavior, and how the best locus of such interventions can be identified. In Chapter 9, a computational model that describes contagion of behaviors, attitudes and emotions through a social network is validated using real-life data. Chapter 10 demonstrates the effect of online communities in a corporate physical activity promotion program, and Chapter 11 investigates the role of social contagion in explaining changes in physical activity in the same dataset. Finally, Chapter 12 investigates the relation of social comparison implementations with the users’ preferences.

Part V: Combining all components of the Active2Gether intervention

Part V covers the topics raised in research question 4. It combines the components that were investigated in each of the previous parts. First, Chapter 13 describes the stepwise approach of the overall design of the Active2Gether intervention. Then, Chapter 14 provides more detail on the technical development of the Active2Gether intervention and its underlying system. In Chapter 15, the results of a user study on the effectiveness of the Active2Gether intervention are presented. Finally, Chapter 16, discusses the results of a user evaluation study of the Active2Gether intervention, in which the experiences of users with the system are investigated.

Part VI: Discussion and Conclusion

The final chapter of this thesis (Chapter 17) provides a discussion of the results obtained in each of the parts, reflects on ethical issues, and looks ahead at future work related to the currently presented ideas and endeavors.

1.6.3 Chapters

Most of the chapters of this thesis appeared as peer-reviewed published articles, one is still under review and one will be submitted for publication. This section provides an overview of the articles that the chapters were based on, together with a brief explanation of my individual contributions to each piece of work.

My personal contributions included conceiving the review setup, collecting the app data, conducting the review, providing intellectual input to the review and manuscript, and approving the final version of the article.


My contributions included conceiving the review setup, co-supervising the student having the lead in the review, conducting the review, performing the analyses, writing most of the manuscript and incorporating feedback in the final version of the article.


My contributions to this work included providing intellectual input, co-supervising the students having the lead in the focus group interviews, assisting with the focus group discussions, providing feedback on the manuscript, and approving the final version of the article.


The authors can be regarded to have made equal contributions to the work and are therefore in alphabetical order. My personal contributions included exploring relevant literature, co-designing and co-implementing the computational model, running a substantial part of the simulations, writing a substantial part of the manuscript and incorporating feedback in the final version of the article.


My contributions included co-designing the experimental setup, conducting a substantial part of the data collection, performing the analyses, writing most of the manuscript and incorporating feedback in the final version of the article.

**Chapter 7** has been published as: **Julia S. Mollee**, Eric F. M. Araújo, and Michel C. A. Klein (2017). “Exploring Parameter Tuning for Analysis and Optimization of a

My contributions included co-designing the experimental setup, conducting a substantial part of the data collection, performing the analyses, writing most of the manuscript and incorporating feedback in the final version of the article.


The authors can be regarded to have made equal contributions to the work and are therefore in alphabetical order. My personal contributions included co-designing the experimental setup, implementing part of the software, performing part of the analyses, writing a substantial part of the manuscript, providing feedback on the manuscript and incorporating feedback in the final version of the article.


My contributions included co-designing the experimental setup, co-supervising the student in charge of conducting the experiment, providing feedback on the analyses and the manuscript, and approving the final version of the article.


My contributions included co-designing the experimental setup, providing feedback on the analyses and the manuscript, and approving the final version of the article.


My contributions included co-designing the experimental setup, co-performing the analyses, writing most of the manuscript, incorporating feedback on parts of the manuscript, providing feedback on parts of the manuscript, and approving the final version of the article.

Chapter 12 has been published as: Julia S. Mollee and Michel C. A. Klein (2016). “The effectiveness of upward and downward social comparison of physical activity in an

My contributions included co-designing the experimental setup, co-supervising the student in charge of conducting the experiment, exploring relevant literature, performing the analyses, writing most of the manuscript, and incorporating feedback in the final version of the manuscript.

Chapter 13 is based on the article that is conditionally accepted for publication as: Anouk Middelweerd, Saskia J. te Velde, Julia S. Mollee, Michel C.A. Klein, and Johannes Brug (2018). “Development of Active2Gether: An app-based intervention combining evidence-based behavior change techniques with a model-based reasoning system to promote physical activity among young adults”. In: Journal of Medical Internet Research.

My contributions included co-designing the intervention, co-designing the experimental setup and co-conducting the intervention study, providing feedback on the manuscript and approving the final version of the article.

Chapter 14 has been published as: Michel C. A. Klein, Adnan Manzoor, and Julia S. Mollee (2017). “Active2Gether: A Personalized m-Health Intervention to Encourage Physical Activity”. In: Sensors 17.6, pages 1436–1451.

The authors can be regarded to have made equal contributions to the work and are therefore in alphabetical order. My contributions included co-designing and implementing a substantial part of the proposed system, writing a substantial part of the manuscript, providing feedback on the manuscript, incorporating feedback in the final version of the manuscript.

Chapter 15 is based on the manuscript of the article that will be submitted as: Anouk Middelweerd, Julia S. Mollee, Michel C.A. Klein, Adnan Manzoor, Johannes Brug, and Saskia J. te Velde (2018). Exploring use and effects of an app-based intervention to promote physical activity: Active2Gether.

My contributions included co-designing the intervention, implementing a substantial part of the intervention, co-designing the experimental setup and co-conducting the intervention study, assisting in the data preprocessing, providing feedback on the manuscript and approving the final version of the article.


My contributions included co-designing the experimental setup, co-conducting the data collection study, performing the analyses, writing most of the manuscript, and incorporating feedback in the final version of the article.

Additional publications

In addition, the following four articles have been written in the context of the creation of this thesis, but are not included.


*The authors can be regarded to have made equal contributions to the work, and are therefore in alphabetical order.
References


REFERENCES

Fagan, Jeffrey, Deanna L. Wilkinson, and Garth Davies (2007). “Social contagion of violence”. In:


Part Two

2 Review of behavior change techniques in apps ........................................ 41
  2.1 Background
  2.2 Methods
  2.3 Results
  2.4 Discussion
  2.5 Conclusions

3 Review of technological features in apps  55
  3.1 Introduction
  3.2 Methods
  3.3 Results
  3.4 Discussion
  3.5 Conclusion

4 User preferences for apps that promote physical activity  75
  4.1 Background
  4.2 Methods
  4.3 Results
  4.4 Discussion
  4.5 Conclusions
Abstract
Physical inactivity contributes to approximately 3.2 million deaths annually and is the fourth leading risk factor for premature death. Over the recent years, the number of smartphone applications (apps) for health and fitness has grown rapidly, and they might form part of the solution to physical inactivity. In May 2013, the iTunes and Google Play stores contained 23,490 and 17,756 health and fitness apps, respectively. However, the quality of these apps, in terms of applying established health behavior change techniques, remains unclear.

This study investigates use of behavior change techniques in apps (developed for Android or iOS) that promote physical activity among adults through tailored feedback. Sixty-four apps were downloaded, reviewed and rated based on the taxonomy of behavior change techniques used in the interventions. Mean and ranges were calculated for the number of observed behavior change techniques. Using nonparametric tests, we compared the number of techniques observed in free and paid apps and in iTunes and Google Play.

On average, the reviewed apps included 5 behavior change techniques (range 2–8). Techniques such as self-monitoring, providing feedback on performance and goal setting were used most frequently, whereas other techniques such as motivational interviewing, stress management, relapse prevention, self-talk, role models and prompted barrier identification were not.

The present study demonstrated that apps promoting physical activity applied an average of 5 out of 23 possible behavior change techniques. No differences for paid and free apps or between app stores were found. The most frequently used behavior change techniques in apps were similar to those most frequently used in other types of physical activity promotion interventions.
This chapter appeared as:

2.1 Background

Physical inactivity contributes to approximately 3.2 million deaths annually and is the fourth leading risk factor for premature death (World Health Organization, 2009, 2014). Despite the fact that many people do not comply with physical activity recommendations (World Health Organization, 2010, 2014), smartphone applications (apps) that promote physical activity are popular: of the 875,683 active apps available in iTunes and the 696,527 active apps in Google Play, 23,490 and 17,756 were categorized as Health and Fitness (148Apps.biz, 2013; AppBrain, 2013). Therefore, it is worthwhile to study the potential of apps that aim to promote physical activity, especially because 56% of the US adults owns a smartphone (Center, 2013). Health behavior change interventions are more likely to be effective if they are firmly rooted in health behavior change theory (Foster et al., 2013; Noar and Mehrotra, 2011; Webb et al., 2010). Webb et al. have noted the importance of behavior change theories in Internet-based interventions (Webb et al., 2010). Additionally, earlier studies have suggested that individually tailored feedback (i.e., feedback based on the user’s own characteristics (Kreuter et al., 1999)) and advice is more likely to be effective than generic information about physical activity (Foster et al., 2013; Lustria et al., 2013; Van den Berg et al., 2007).

Many advantages of using the Internet as a delivery mode apply to smartphone apps too: constantly accessible, adjustable to the needs of the user (Griffiths et al., 2006), able to provide (computer-) tailored feedback, large reach and interactive features. Because people carry smartphones and can access data anywhere and anytime, physical activity behavior change promotion apps offer the opportunity to provide tailored feedback and advice at the appropriate time and place. Therefore, apps offer new opportunities to deliver individually tailored interventions, including real-time assessment and feedback that are more likely to be effective.

Apps are relatively new tools in physical activity interventions and only very little research has been published to date on the content and the effectiveness of physical activity apps. It remains unclear to what extent apps differ in their relevant content and if these differences mediate effectiveness. Previous research suggests that the use of behavior change techniques to address behavioral determinants conceptualized in behavior change theory, is linked to effectiveness (Abraham and Michie, 2008). Therefore, it can be proposed that the presence of behavior change techniques in general and some specific behavior change techniques in particular is an indicator of potential effectiveness. Abraham and Michie developed a taxonomy to identify behavior change techniques in a range of health promotion interventions (Abraham and Michie, 2008). The taxonomy can be used to identify techniques or combinations of techniques that enhance effectiveness. The most frequently applied behavior change techniques in traditional interventions are goal setting (Abraham and Michie, 2008), prompting intention formation (Abraham and Michie, 2008), providing feedback on performance (Abraham and Michie, 2008), self-monitoring (Abraham and Michie, 2008) and reviewing behavioral goals (Conn et al., 2011; Michie, Abraham, et al., 2009). A large body of work has been published using the taxonomy in health promotion interventions (Conn et al., 2011; Michie, Abraham, et al., 2009; Michie, Jochelson, et al., 2009; Webb et al., 2010), but so far, no study has been conducted with the aim to review application of behavior change techniques in apps.

Therefore, the present study aims to review apps developed for iOS and Android that promote physical activity among adults through individually tailored feedback and
advice. Recent reviews have concluded that health promoting apps lack the use of behavior change theories in promoting behavior changes such as smoking cessation, weight-loss, and increased physical activity (Abroms et al., 2011; Breton et al., 2011; Cowan et al., 2013; West et al., 2012). Only one earlier study focused on the use of behavior change theories in apps that target physical activity (Cowan et al., 2013). However, the authors limited their search to iTunes and excluded apps that targeted other health behaviors in addition to physical activity (e.g., apps that combined physical activity and diet information). Another limitation of their review was that it included apps that only provided information or solely used GPS tracking to promote physical activity. In addition, the authors used a first generation iPad to download and review the apps and consequently had to exclude apps that were not compatible with this tablet. To improve upon the existing body of research on this topic, the current study reviews the use of behavior change techniques in physical activity apps available in both app stores (i.e., iTunes and Google Play) restricted to apps that utilize tailored feedback. Because previous studies reported a significant association between price and the inclusion of behavior change theories (Cowan et al., 2013; West et al., 2012), free and paid apps will be compared. Since we derived apps from two different online sources that differ in their acceptance policy, we additionally assessed whether the number of behavior change techniques differed between apps available in iTunes and Google Play.

2.2 Methods

2.2.1 Inclusion criteria

This review included apps that were available through iTunes and Google Play. Apps were included if they (i) were in English, (ii) promoted physical activity, (iii) followed the official recommendation of 150 minutes of moderate to vigorous physical activity per week (World Health Organization, 2010), (iv) were primarily aimed at healthy adults, and (v) provided individually tailored feedback. Thus, apps that specifically focused on children, adolescents, older adults, pregnant women, unhealthy adults or individuals with disabilities were excluded because of the differences in physical activity guidelines for these groups (World Health Organization, 2010). Apps that provided feedback by showing logged statistics without feedback or without information about progress toward a personal user-set goal were also excluded.

2.2.2 Search strategy

The study sample was identified through systematic searches in iTunes and Google Play. Apps from iTunes were identified between August and September 2012, and apps in Google Play were identified between November 2012 and January 2013. Because the two reviewers (AM and JM) screened the apps on different days, there was a slight variation in the number of apps offered in the app stores. During the search and screening period, iTunes updated its search strategies (on August 24, 2012), which reduced the number of apps retrieved with a specific search term. In case one of the reviewers retrieved fewer apps than the other due to this update, the results from the earlier search were included.

Search terms were based on Boolean logic and included AND combinations for physical activity, healthy lifestyle, exercise, fitness, coach, assistant, motivation, and support.
2.2 Methods

2.2.3 Screening procedure

Because the screening procedure for iTunes differed to some extent from Google Play, the screening procedures are reported separately. If an app had a free version and a paid version, the free version was downloaded first. If the paid version had relevant extra features (tailored feedback or additional features not available for the free version), it was also downloaded and evaluated. This method was applied for both screening procedures. If the same version of an app was available in iTunes and in Google Play, the iTunes version was downloaded and assessed for eligibility. For both iTunes and Google Play, the identification and eligibility phases of screening were performed by two researchers (AM and JM, or AM and StV), and differences between the two reviewers were resolved by discussion and/or involving the third reviewer.

First, the screening procedure was conducted for apps available in iTunes. Figure 2.1 provides a schematic overview of the decision sequence.

In the identification phase, search terms were entered in iTunes. In the screening phase, the app description and screenshots were reviewed based on the inclusion criteria. If the app appeared to be eligible, it was downloaded to an iPhone 4S smartphone and assessed for eligibility. In the eligibility phase, a reviewer explored each app by using all of its available functions.

Google –including Google Play– has a somewhat different search algorithm than iTunes. For example, it extends the search by recognizing synonyms and personal preferences, resulting in twice as many hits compared to iTunes. Therefore, the review steps were adapted for Google Play. Google Play’s search algorithms also prioritize search results, meaning that the first results listed are the most relevant and the closest to the search terms. Therefore, the adjusted screening method specified that for search terms revealing over 1,000 apps, the title, description, and screenshots of the first 100 apps were first screened carefully. If at least five out of the first 100 apps met the inclusion criteria, the next 100 apps were also screened. If one app was selected in the second group of 100 apps, the screening procedure was continued with the next 100 apps, and so on, until no apps were selected in a group of 100 screened apps. All remaining apps (AM = 1,801, JM = 1,331) were additionally screened for possible eligibility based on their title. If the title indicated possible eligibility, the app was screened for inclusion. This screening procedure was applied for eight search terms that revealed over 1,000 apps: “physical activity”, “healthy lifestyle AND fitness”, “fitness AND exercise”, “fitness AND coach”, “fitness AND motivation”, “fitness AND support”, “exercise AND support”, and “physical activity AND support”.

Figure 2.1 provides a schematic overview of the decision sequence for the decision sequence for Google Play apps as well. In the identification phase, search terms were entered in Google Play. In the screening phase, the app description and screenshots were reviewed based on the inclusion criteria. Apps that appeared to be eligible were downloaded to an HTC Rhyme smartphone and were fully explored by using all functions available in the app.

Apps commercially available do not provide detailed (intervention) descriptions or published study protocols, therefore an alternative approach was chosen to detect behavior change techniques in apps. Firstly, all available functions (e.g. information, chat, monitoring options, reminders and graphs) were explored by using the app for about half an hour. Secondly, the apps were running in the background for a couple days so the authors were able to read the reminders and push messages.
Figure 2.1: Flow chart: schematic overview of the selection process for apps eligible for full review. This flow chart provides a schematic overview of the selection process of eligible apps available in iTunes and Google Play (GP). The initials of the main reviewers are reported as JM and AM. a Apps on the list of one researcher were untraceable for the other researcher. b Apps to which the adjusted screening method had been applied and only the titles were screened. c Apps that were not available in English or Dutch. d The main focus of the apps was not physical activity (PA) promotion. e Apps that focused on diet and weight loss. f The main focus of the apps was not physical activity (PA) promotion or weight loss. g Apps that targeted people with injuries or disabilities. h Apps that targeted children or older adults. i Apps did not follow the guidelines for physical activity. j Apps that did not provide tailored feedback. k Apps that were detected in the first screening step and were not available for download. l After downloading the app, it did not work. m An extra monitor or device was needed to receive tailored feedback. n Before using the app, a credit card was needed to deduct money as a penalty if the user did not achieve self-defined goals. o The same app was available under a different name but with the same features. p The app had a free and a paid version, but the paid version did not have additional features.
The apps were rated based on the taxonomy of behavior change techniques used in interventions (Abraham and Michie, 2008). This taxonomy was developed to identify potentially effective behavior change techniques used in interventions (Abraham and Michie, 2008) and was previously used to identify behavior change techniques in interventions that aimed to increase physical activity (Abraham and Michie, 2008; Golley et al., 2011; Michie, Abraham, et al., 2009; Webb et al., 2010). The taxonomy distinguished 26 behavior change techniques. Three of these techniques had low inter-rater reliability and were thus not included in the present review (Abraham and Michie, 2008), resulting in an adapted version of the taxonomy with 23 items.

Each app was scored by two reviewers (AM, JM) on all 23 items of the adapted taxonomy. Each app received a score of 0–23 representing the number of behavior change techniques identified. The results were entered into an electronic database (Microsoft Access 2003). In preparation for scoring each app, the reviewers studied a coding manual and discussed each item of the taxonomy carefully. For example, self-monitoring was defined as all features helping in keeping record of the behavior (e.g., GPS-tracking, diary, accelerometer). Specific goal setting was defined if a feature helps with detailed planning, the goal had to be clearly defined. Plan social support was seen as all features offering social support (e.g., possibility to link with social networking sites, chat possibilities).

The apps were scored independently, and a percentage of agreement was calculated to assess inter-rater reliability between reviewers. The percentage of exact agreement was 44%, and 91% of the scores were within a difference of 1 point. Nine percent of the apps had a disagreement of >1 point (but with a maximum of 3 points). Subsequently, differences in interpretation were resolved by discussion.

The name of the app, the name of the app producer, the date it was downloaded, the name of the app store, and the price were collected for each app in addition to the app’s score based on the number of behavior change techniques it used.

Means and ranges were calculated for the sum behavior change technique scores and the price of apps. Significant differences in the use of behavior change techniques (between iTunes and Google Play and between free and paid apps) and in price (between iTunes and Google Play) were assessed with Mann Whitney U tests (significance level of $p < .05$). To compare iTunes and Google Play, apps available in both stores were excluded, otherwise the same app would be included twice in the same analyses (once in the iTunes group, once in the Google Play group).

Due to the time differences mentioned earlier, reviewer AM detected 1,913 apps in iTunes and 5,540 apps in Google Play and reviewer JM detected 1,968 apps in iTunes and 5,217
apps in Google Play. The current review included 41 apps available in iTunes and 23 apps available in Google Play, of which 30 and 21, respectively, were free. The mean price of the paid apps was €2.06 (range €0.79–8.99) for iTunes and €1.88 (range €0.76–2.99) for Google Play. Seven apps were available in both iTunes and Google Play for free.

The average number of behavior change techniques included in the eligible apps was 5 (range 2–8). Table 2.1 shows the sum score for behavior change techniques for each app. One app had a score of 8 out of 23.

### Table 2.1: The number of behavior change techniques (BCTs) in apps.

<table>
<thead>
<tr>
<th>App</th>
<th>App store</th>
<th>Price (€)</th>
<th>Score BCT</th>
</tr>
</thead>
<tbody>
<tr>
<td>RunKeeper - GPS Track Run Walk</td>
<td>Google Play</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Big Welsh Walking Challenge</td>
<td>iTunes</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>GymPush</td>
<td>iTunes</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Hubbub Health</td>
<td>iTunes</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>My Pocket Coach (a life, wellness &amp; success coach)</td>
<td>iTunes</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Sixpack - Personal Trainer</td>
<td>iTunes</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Teemo: the fitness adventure game!</td>
<td>iTunes</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>fitChallenge</td>
<td>iTunes</td>
<td>0.89</td>
<td>6</td>
</tr>
<tr>
<td>FitCoach - powered by Lucozade Sport</td>
<td>iTunes</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Fitness War</td>
<td>iTunes</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Running Club</td>
<td>iTunes</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Sworkit Pro</td>
<td>Google Play</td>
<td>0.76</td>
<td>6</td>
</tr>
<tr>
<td>Take a Walk Lite</td>
<td>iTunes</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Track &amp; Field REALTIMERUN (GPS)</td>
<td>iTunes</td>
<td>0.89</td>
<td>6</td>
</tr>
<tr>
<td>Withings- Lose Weight, Exercise, Sleep Better, Monitor Your Heart</td>
<td>iTunes/Google Play</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>1UpFit</td>
<td>iTunes</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>All-in Fitness: 1000 Exercises, Workouts &amp; Calorie Counter</td>
<td>iTunes</td>
<td>8.99</td>
<td>5</td>
</tr>
<tr>
<td>Be Fit, Stay Fit Challenge</td>
<td>Google Play</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Endomondo Sports Tracker</td>
<td>Google Play</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Everywhere Run! - GPS Run Walk</td>
<td>Google Play</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Fit Friendzy</td>
<td>iTunes</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>FitCommit - Fitness Tracker and Timer</td>
<td>iTunes</td>
<td>1.59</td>
<td>5</td>
</tr>
<tr>
<td>Fitocracy Fitness Game,Tracker</td>
<td>iTunes/Google Play</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Healthy Heroes</td>
<td>iTunes</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Improver</td>
<td>iTunes</td>
<td>0.79</td>
<td>5</td>
</tr>
<tr>
<td>Macaw</td>
<td>iTunes/Google Play</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Make your move</td>
<td>iTunes</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Nexercise = fun weight loss</td>
<td>iTunes/Google Play</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Nike+ Running</td>
<td>Google Play</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Noom CardioTrainer</td>
<td>Google Play</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>ShelbyFit</td>
<td>iTunes</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>SoFit</td>
<td>Google Play</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Strava Cycling</td>
<td>Google Play</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Tribesports</td>
<td>Google Play</td>
<td>0</td>
<td>5</td>
</tr>
</tbody>
</table>
2.4 Discussion

Providing feedback (n = 64), self-monitoring (n = 62), and goal setting (n = 40) were used most frequently, whereas motivational interviewing, stress management, relapse prevention, self-talk, role modeling, and prompted barrier identification were not used in any of the screened apps (see Figure 2.2).

Free and paid apps did not differ with respect to the use of behavior change techniques (p = .18). No differences in price were found between apps available in iTunes and Google Play (p = .14). Similarly, apps available in iTunes and Google play did not differ with respect to the number of behavior change techniques used (p = .39).

The current review aimed to evaluate the use of behavior change techniques in apps available through iTunes and Google Play that target physical activity and use tailored feedback, based on an established taxonomy of such techniques (Abraham and Michie, 2008; Norman et al., 2007). The 64 apps included in the review used on average 5 different behavior change techniques, and none of the apps used more than 8 or less than 2. Providing feedback and self-monitoring were the most frequently used technique. At least two behavior change techniques were identified in each of the apps included in the review, which suggests that app developers attempt to use behavior change theory to some extent. However, the results
Chapter 2. Review of behavior change techniques in apps

Figure 2.2: Frequencies of the 23 behavior change techniques used in apps. Behavior change techniques are scored using the taxonomy created by Abraham and Michie (2008), ranked by the most frequently applied techniques.

also indicate that the inclusion of established behavior change techniques is far from optimal in most apps.

Studies in which behavior change theories in apps were operationalized have concluded that apps generally lack the use of theoretical constructs (Breton et al., 2011; Cowan et al., 2013; West et al., 2012). For example, West et al. concluded that 1.86% of the apps in Health & Fitness included all of the factors of the Precede Proceed Model (West et al., 2012). Similarly, Cowan et al. found that key constructs of behavior change theories were seldom used in apps that target physical activity (Cowan et al., 2013). Lastly, Breton et al. found a lack of adherence to evidence-based practice in apps targeting weight loss (average 3 practices, range 0–12) (Breton et al., 2011). The findings of the present review are somewhat more favorable than earlier findings from the reviews described above. The more frequent use of behavior change techniques in the apps reviewed in the current study may be a consequence of the inclusion criteria. We only included apps that provided individually tailored feedback and excluded generic information apps, which may have resulted in the exclusion of apps that were not based on theoretical constructs. In addition, technological
development in recent years may have resulted in the ability to develop more advanced app features, including the use of a wider range of behavior change techniques. Another finding that deviates from previous studies is that free and paid apps did not differ in the number of behavior change techniques used, whereas previous reviews found that price was positively associated with use of theoretical constructs (Cowan et al., 2013; West et al., 2012). The differences in findings may be explained by the number of paid apps included, which was much higher in our review compared to previous reviews (Cowan et al., 2013; West et al., 2012).

Previous reviews that applied Abraham and Michie’s taxonomy (Abraham and Michie, 2008) to assess the number of behavior change techniques used in non-app interventions identified on average 6–8 behavior change techniques (Abraham and Michie, 2008; Golley et al., 2011; Michie, Abraham, et al., 2009). Frequently used behavior change techniques are: self-monitoring, feedback on performance and goal setting (Golley et al., 2011; Michie, Abraham, et al., 2009; Webb et al., 2010). Interventions including self-monitoring in combination with providing feedback, specific goal setting, prompt intention formation or prompt review behavioral goals showed larger effect sizes (Conn et al., 2011; Michie, Abraham, et al., 2009). Furthermore, studies reported inconclusive conclusions regarding the number of behavior change techniques that are associated with larger effects: a systematic review on web-based interventions reported that interventions that included larger numbers of behavior change techniques are more likely to be effective (Webb et al., 2010), whereas another meta-analysis suggests that the number of included behavior change techniques is not associated with a larger effect (Michie, Abraham, et al., 2009).

Although we found that the average number of behavior change techniques used in apps was lower than previously reported for other types of physical activity promotion, the most frequently used types of behavior change techniques used were similar (Golley et al., 2011; Michie, Abraham, et al., 2009; Webb et al., 2010). It remains unclear if lack of theory-driven behavior change techniques in apps is due to technical difficulties or due to other factors. However, the findings of the current review, combined with our knowledge about what specific behavior change techniques have been effective in other types of behavior change interventions, suggest that apps may be an effective way to promote physical activity.

Unfortunately, little is currently known about the effect of apps on physical activity. The current review provides information about the content of apps, but future research should study how behavior change techniques can be translated into apps. Additionally, future research should examine the effectiveness of apps and which behavior change techniques or combinations of techniques are more effective.

This review indicates that apps have the potential to provide tailored feedback and to integrate behavior change techniques. Smartphones with Internet access and apps turn a cell phone into a portable personal computer. This technology offers the opportunity for ecological momentary assessment (EMA) and makes it feasible to provide timely messages based on the user’s location (Lin, 2013; Norman et al., 2007). The application of smartphones and apps in health behavior interventions are growing rapidly, however little has been published about the interventions using the new technology to provide real-time feedback (Riley et al., 2011).

A collaboration between app developers, health professionals, and behavior change experts could increase the use of behavior change techniques in apps and may open a new scale of possibilities in health promotion.
Strengths and limitations

Scoring the content of apps is susceptible to rater bias. The level of inter-rater reliability in this review was lower than that of previous content analyses of apps (Cowan et al., 2013; West et al., 2012). This study's relatively low inter-rater reliability may be because Abraham and Michie’s taxonomy (Abraham and Michie, 2008) was originally designed to score other behavior change interventions than smartphone app-based interventions. Applying the taxonomy to apps forced the researchers to translate the strategies into app functionalities. Following this logic, the researchers had to score each app based on what they observed. Although the researchers reviewed the apps carefully, behavior change strategies in apps may have been overlooked or interpreted differently, and some behavior change techniques may be more obvious than others. Thus, some of the behavior change techniques may be hidden in the app features and may therefore not been detected, especially follow-up prompts.

This study evaluated the use of behavior change techniques in apps that target physical activity but provides no information about the effectiveness of these apps. Further research is needed to evaluate the effectiveness of apps that promote physical activity. The strengths of the present review include the extensive search strategy, the inclusion of both iTunes and Google Play, and the independent rating of the apps by two reviewers. Moreover, rating of the apps was not limited to apps that were free but also included retail apps. Finally, rating was done after downloading and using all of the app’s functions rather than solely using screen shots.

2.5 Conclusions

The present study demonstrates that apps promoting physical activity applied an average of five behavior change techniques. There was no difference in the number of identified behavior change techniques between free and paid apps. The most frequently used behavior change techniques in apps were goal setting, self-monitoring and feedback on performance, which was similar to the ones most frequently used in other types of physical activity promotion interventions. The findings of the present study showed that apps can substantially be improved regarding the number of applied techniques.

Competing interests

The authors declared that they have no competing interests.

Authors’ contributions

AM: Conducted the review, performed the analyses, drafted the manuscript and incorporated all feedback. JM: Conducted the review, provided intellectual input to the review and manuscript and approved the final version. NydW: Provided intellectual input, provided feedback on the manuscript and approved the final version of the manuscript. JB: Provided intellectual input to the design and execution of the review and to the manuscript, provided feedback and approved the final version of the manuscript. StV: Designed the review and proved intellectual input the execution of the review, screened part of the applications, provided feedback and approved the final version of the manuscript. All authors read and approved the final manuscript.
References


Breton, Emily R., Bernard F. Fuemmeler, and Lorien C. Abroms (2011). “Weight loss—there is an app for that! But does it adhere to evidence-informed practices?” In: Translational behavioral medicine 1.4, pages 523–529.


Foster, Charles, Justin Richards, Margaret Thorogood, and Melvyn Hillsdon (2013). “Remote and web 2.0 interventions for promoting physical activity”. In: The Cochrane Library.


3. Review of technological features in apps that promote physical activity

Abstract

Despite the well-known health benefits of physical activity, a large proportion of the population does not meet the guidelines. Hence, effective and widely accessible interventions to increase levels of physical activity are needed. Over the recent years, the number of health and fitness apps has grown rapidly, and they might form part of the solution to the widespread physical inactivity. However, it remains unclear to which extent they make use of the possibilities of mobile technology and form real e-coaching systems.

This study aims to investigate the current landscape of smartphone apps that promote physical activity for healthy adults. Therefore, we present a framework to rate the extent to which such apps incorporate technological features, and we show that the physical activity promotion apps included in the review implemented an average of approximately eight techniques and functions. The features that were implemented most often were user input, textual/numerical overviews of the user’s behavior and progress, sharing achievements or workouts in social networks, and general advice on physical activity. The features that were present least often were adaptation, integration with external sources, and encouragement through gamification, some form of punishment or the possibility to contact an expert. Overall, the results indicate that apps can be improved substantially in terms of their utilization of the possibilities that current mobile technology offers.
This chapter appeared as:

3.1 Introduction

Despite the well-known health benefits of physical activity, 23 percent of the adult population worldwide does not meet the recommended guidelines (World Health Organization, 2014). Systematic reviews concluded that levels of physical activity in Europe vary across countries, ranging from 15.6 percent in Israel to 84.8 percent in Slovakia who meet the guidelines (Marques et al., 2015). In the Netherlands, approximately one third of the adult population does not meet the Dutch guidelines for healthy physical activity (Wendel-Vos, 2014). Moreover, compared to other European countries, Dutch people lead a relatively sedentary lifestyle, with 25 percent spending at least 8.5 hours sitting on a usual day and over 60 percent at least 5.5 hours (TNS Opinion & Social, 2014). Insufficient physical activity is one of the leading risk factors for premature mortality and avoidable health-related issues as cardiovascular diseases, cancer and diabetes (Lee et al., 2012; World Health Organization, 2014). Thus, effective and widely accessible interventions to increase levels of physical activity are needed.

Smartphones and smartphone applications (apps) could be useful as mobile coaching systems that aim to increase levels of physical activity, as they are well intertwined in modern society, always accessible to the user, and because they can lower the barrier for people to address their health problems (Griffiths et al., 2006). Despite the fact that many adults do not meet the activity guidelines, apps that focus on health and fitness promotion are popular. To illustrate, the numbers of health and fitness apps are still growing and the iTunes App Store contained 71,895 health and fitness apps in 2016 (Steel Media Ltd., 2016), including both free and paid apps. Moreover, also traditional interventions have been influenced by ICT developments and make use of mobile phones and the internet. Web-based and mobile (app) based interventions (i.e., eHealth and mHealth interventions) provide opportunities for delivering personalized materials to promote physical activity on a population level (Krebs et al., 2010; Vandelanotte et al., 2016). Several reviews and meta-analyses of eHealth interventions targeting physical activity found small effects on levels of physical activity in favor of the intervention groups (Davies et al., 2012; Krebs et al., 2010; Webb et al., 2010). mHealth interventions that were included in systematic reviews and meta-analyses mainly consisted of interventions delivered via SMS or a personal device assistant (PDA) and showed promising results (Fanning et al., 2012; Head et al., 2013; Muntaner et al., 2016; Vandelanotte et al., 2016). However, to date no systematic reviews on the effectiveness of app-based interventions to promote physical activity are available.

Smartphones offer a wide range of technological possibilities, as part of or in addition to techniques used in eHealth and mHealth, such as telecommunication, sensoring/monitoring, and any-time any-place support. Even though no systematic reviews on the effectiveness of mobile interventions to promote physical activity have been published yet, there are several content analyses available focusing on the inclusion of behavior change theories and behavior change techniques. Those reviews showed that the apps were generally lacking foundation in behavior change theories and the use of behavior change techniques that are associated with effectiveness (Conroy et al., 2014; Cowan et al., 2013; Direito et al., 2014; Middelweerd, Mollee, et al., 2014; West et al., 2012). Behavior change techniques that were often included in apps were self-monitoring, providing feedback on performance and goal-setting (Middelweerd, Mollee, et al., 2014). However, sensoring and monitoring can be done in various ways and it remains unclear to what extent current physical activity apps make use of the technological possibilities to help the user to be physically active and thus
actually deliver the promises of mobile coaching systems. For example, features as self-monitoring can be based on different types of inputs, e.g., user input (i.e., diary) or sensor data obtained from the phone or from external sensors, such as a Fitbit or a GPS-watch.

Although some technological features can be mapped to behavior change techniques (that again can be associated with effectiveness), it is currently unknown what features implicate higher effectiveness of physical activity apps. It is an interesting first step to investigate the prevalence of those features in the current supply of physical activity apps. Therefore, the aim of the present paper is to inventory the landscape of the state-of-the-art smartphone apps that promote physical activity, in order to (1) gain insight in technological possibilities and (2) identify missed opportunities. More specifically, a framework of technological features is proposed, and a set of apps is selected systematically for the content analysis to discern how often those features are implemented. In addition, it is investigated whether the price of an app and the type of app store in which they are available (Google Play Store vs. iTunes App Store) are correlated with the number of features that are implemented. Also, we explored whether the apps’ number of features are correlated with the reviewers’ ratings of their usability.

The remainder of this paper is organized as follows. Section 3.2 describes the methods of identifying, screening and scoring the eligible apps, including the framework used to score the apps. The results are presented in Section 3.3, and they are reflected upon in the discussion in Section 3.4. Finally, Section 3.5 provides a conclusion.

### 3.2 Methods

This section describes the process of searching, screening and selecting the apps to be included in the systematic review, as well as the scoring procedure and how the scores were analyzed.

#### 3.2.1 Identification

For this review, the Google Play Store and the iTunes App Store were searched for relevant apps. In the first quarter of 2015, Android and iOS (the mobile operating systems served by these two app stores) accounted for 96.7% of the market share. The remaining 3.3% is covered by Windows Phone, Blackberry OS and other mobile operating systems (IDC, 2015). For reasons of efficiency, only apps from the app stores of the two market leaders were reviewed in this study.

The Google Play Store and the iTunes App Store were searched between April and May 2015. The search terms used to search the app stores were based on an exploration of the 20 most popular apps in the ‘Health & Fitness’ category of both app stores. The descriptions of those apps were screened and the most prevalent terms were listed. The resulting list of key words was used to construct a set of combined search terms: coach fitness, coach exercise, coach fit, coach workout, coach training, fitness exercise, fitness fit, fitness workout, fitness training, exercise fit, exercise workout, exercise training and physical activity. These search terms were used to identify relevant physical activity apps in the two app stores, up to a maximum of 100 apps per search term. This lead to 100 screened apps per search term and app store, except for physical activity and coach fit, that yielded only 48 and 69 results in the iTunes App Store. Thus, a total of 2,517 apps was identified.
3.2 Methods

3.2.2 Screening

The total number of 2,517 identified apps were screened for inclusion in the app review. The screening procedure consisted of evaluating the app description and screenshots in the app stores, in order to determine whether the app met the predefined inclusion criteria. Some apps that were included based on this screening, were still excluded in a later stage after downloading and further exploring the app.

The general inclusion criteria stated that (i) the app is in either English or Dutch, (ii) the app promotes physical activity, (iii) the app is aimed at a healthy population, rather than some specific target group, (iv) the app is focused on adult users, i.e. suitable for users 18 to 65 years of age, (v) the app is not specifically focused at male or female users, and (vi) the app offers more than static information only.

This leads to the following list of exclusion criteria for the apps that were identified through the initial search:

1. General:
   a. Language: The app is in a language other than English or Dutch.
   b. Gender: The app is aimed at male or female users specifically.
   c. Age/Health: The app is not aimed at adults, but at children, adolescents or elderly people specifically, or the app is not aimed at a healthy population, but a specific target group, such as people with obesity or other physical problems or illnesses.

2. Aim:
   a. Dieting: The app is aimed at weight loss, for example through information about dieting, nutrition, calorie counting, without (substantial) physical activity component.
   b. Brain: The app is aimed at brain training to improve cognitive capacities.
   c. Tactics: The app is aimed at teaching tactics (for sports, games or exams).
   d. Games: The app is a game that does not require or promote physical activity.
   e. Mind: The app is aimed at stimulating the mind, through for example meditation and mindfulness.
   f. Specific: The app is aimed at very specific physical activity, such as training one particular muscle group.

3. Methods:
   a. Testing: The app only offers a test of physical fitness or endurance, without further support or advice to become more physically active.
   b. Timer: The app only offers a timer.
   c. Information: The app only offers static information, such as opening times of local sports clubs.
   d. Book/Magazine: The app is a digital version of a book or magazine about physical activity or health.

4. Other: Any other reason why an app was excluded, that does not fit in the reasons listed above. For example, the app only offers a heart rate measurement tool.

After the first screening of the 2,517 identified apps, 227 apps remained to be reviewed. Of those 227 apps, 113 were found in the iTunes App Store, 89 in the Google Play Store
and 25 in both app stores. In the next step, another 58 apps were excluded, for example because they were seemingly removed from the app stores, because the app required external hardware or a paid subscription, or because they did not meet the inclusion criteria for the review after all. For the remaining 169 apps, targeted search revealed in which app store(s) they were actually available, irrespective of which app store they were originally identified in. This led to a total of 38 apps in the iTunes App Store, 39 apps in the Google Play Store and 92 in both app stores.

Figure 3.1 provides a schematic overview of the decision sequence from the identification to the inclusion of the apps. Please note that as soon as one exclusion criterion was identified in the app, the app was rejected and that criterion was registered. This implies that there could have been more reasons why the apps were not eligible than represented in this overview. Apps that were identified multiple times through different search terms are registered under ‘doubles’.

![Figure 3.1: Schematic overview of the selection process for apps eligible for full review.](image-url)
3.2 Methods

3.2.3 Scoring

This section describes how the selected apps were scored. In Section 3.2.3, the framework used for scoring is introduced and explained. In Section 3.2.3, the procedure of scoring the apps is described, including an analysis of the agreement between the different raters.

Scoring Framework

The selected apps were scored using a framework of smartphone features (techniques and functionalities) that can be used to monitor or encourage physical activity in an e-coaching system. The framework was designed based on relevant literature, in order to ensure good coverage of the important features. First, it was partly based on a systematic review of scientific publications on smartphone applications that aim to increase physical activity levels (Bort-Roig et al., 2014). In this review, the authors investigated 26 articles reporting about the viability of smartphones to measure and/or influence physical activity, but did not investigate the described apps themselves. Second, the framework was partly based on literature on desired features in smartphone applications that promote physical activity (Middelweerd, van der Laan, et al., 2015; Rabin and Bock, 2011).

The resulting framework consists of 50 items, which are organized into five categories about app features (measuring & monitoring, information & analysis, support & feedback, adaptation and social), and two categories for additional information (usability and other). The categories were established through discussions between the authors, and allow for analysis of the apps on different levels of abstraction. The two additional categories are subordinate in the current work, as they concern other aspects than app features, but serve as a basis for further analyses on the collected data, such as investigating correlations between the perceived usability of apps and their implemented features.

1. Measuring & Monitoring:
   This category contains items about how the app receives its input. This could be, for example, through user input, through built-in smartphone sensors, or through external (hardware) sensors or other sources.

2. Information & Analysis:
   This category contains items about how the collected data is analyzed, summarized and represented.

3. Support & Feedback:
   In this category, the items cover what kind of support or feedback the app offers the users. For example, does it provide auditory or visual (real-time) feedback, or feedback based on the user’s context, etc.

4. Adaptation:
   This category contains items about whether (and to what extent) the app adapts to the user, e.g. does the app automatically adjust the goals to the user’s behavior?

5. Social:
   This category investigates the social aspect of the app. For example, is there a community within the app, or is there a possibility to connect to external social networking platforms? Additionally, the category contains items about the functionalities within these online communities, e.g. is it possible to send messages, to compete, or to see a leaderboard?

6. Usability:
In this category, the apps are scored on clarity (how easy is it to find information) and attractiveness (does the app look appealing), on a scale from 1 to 5.

7. **Other**

   This category contains items that do not fit in the other categories. The items cover whether there is a website available where the users can view their data in (more) detail, and whether the app offers in-app purchases that enable more functionalities.

See the Appendix for the complete scoring framework.

**Scoring Procedure**

The scoring procedure consisted of downloading each of the 227 selected apps and exploring the different functionalities offered by the app. Actually downloading the app yields more reliable results than merely screening the description and screenshots in the app store (Conroy et al., 2014; Yang et al., 2015). The hands-on experience with each app took approximately 15 minutes. If some of the functionalities were dependent on actual use (i.e., responding to registered accelerometer or location data), the app was kept running in the background for a couple of hours to days, to see if other functionalities would be revealed.

The framework described in Section 3.2.3 was implemented in Microsoft Excel. Each item from the framework was assigned a 1 or 0: a 1 if the answer to the question was ‘yes’ and a 0 if the answer was ‘no’ or if it was not clear whether the app included the described feature. If an item consisted of several subitems, it was awarded a 1 if at least one of the subitems was also awarded a 1, and a 0 otherwise. This allows for analysis of the apps’ features on different levels of abstraction. In addition to the scores on the framework, the name of the app, the name of the app store and the price of the app were registered during the scoring process.

Four reviewers (RK, JM, AM and RFH) contributed to scoring the 227 selected apps. Each app was scored by two reviewers. Of the 227 apps, RK and JM reviewed a set of 125 apps, RK and RFH reviewed 65 apps, and the remaining 37 apps were reviewed by RK and AM. In order to ensure consensus on the interpretation of the framework, it was discussed extensively before the start of the app review. In addition, decisions on interpretation that arose during the scoring process were documented and continuously shared among the four reviewers. After the review of the apps, possible discrepancies between the scores of the two reviewers for each app were resolved by discussion. If the disagreement was not resolved easily (e.g., if a reviewer overlooked a functionality), a score of ‘0.5’ was registered, in order to reflect the disagreement or ambiguity.

The inter-rater reliability was assessed with Cohen’s kappa, by calculating the agreement for each app separately and taking the average. This resulted in $\kappa = 0.69$, which indicates a substantial agreement. However, since some of the apps were scored with many 0’s, the probability of chance agreement is relatively high. This results in a relatively low value for $\kappa$, even though the percentage of agreement between the reviewers is high, namely 90% of all item scores.

### 3.2.4 Analyses

Using the scores obtained as described in Section 3.2.3, the results can be analyzed from different perspectives. First, the scores allow insight in the extent to which such features are incorporated in physical activity promotion apps, by looking at the scores per reviewed app.
Second, the scores can be used to investigate how often certain features are implemented, by looking at the sum score per item in the framework.

In addition, we used the results to investigate whether there is a difference in the number of features applied in free or paid apps, or between apps from the two different app stores. This could reveal whether paid apps are generally more sophisticated (in terms of implemented features) than free apps, or vice versa. Similarly, if apps in one of the app stores are generally equipped with more features, this could indicate a difference in the selection/admission mechanism of the specific app store. The significance of these potential differences was assessed using a Mann Whitney U test with a significance level of $\alpha = 0.05$. Also, we used Spearman’s correlations to investigate whether the number of features implemented in apps is correlated with the reviewers’ ratings for the clarity and attractiveness of the apps, as captured by the *usability* category of the framework.

### 3.3 Results

In this section, we elaborate on the results of reviewing the 169 apps that remained after careful screening and selection.

#### 3.3.1 Overview of included apps

Of the 169 apps, 39 apps were found only in the Google Play Store, 38 apps only in the iTunes App Store and 92 apps in both app stores. Of the apps selected from the Google Play Store, 34 were free and 5 were paid, with an average price of €2.48 (range €0.76–3.39). In the iTunes App Store, 35 of the apps found were free and 3 were paid, with an average price of €1.66 (range €0.99–2.99). Of the 92 apps found in both app stores, 83 apps were free and 9 were paid, with an average price of approximately €2.48 (range €0.99–2.99).

#### 3.3.2 Number of features per app

First, we investigated how many features are generally included in the apps. The items in category 6 (usability) and category 7 (other) were not considered in this analysis, since they do not represent technological features. Also, if an item was divided into several subitems, only these subitems were considered in counting the number of features, in order to avoid double counting of features. The framework contains 37 of such subitems.

The average number of features included in the eligible apps was 8.18 (range 0.5–19.5). Table 3.1 shows the 18 apps with the highest numbers of features. The app with the highest score was Endomondo, with 19.5 out of the possible 37 features (53%). Endomondo is available for free in the Google Play and the iTunes app store.

Table 3.2 shows the 16 apps with the lowest numbers of features. The app with the lowest score was 7 Minute Workout by mphan, with 0.5 out of the possible 37 features (1.3%). 7 Minute Workout is available for free in the Google Play Store.

#### 3.3.3 Number of apps per feature

Second, we investigated which features and categories of features were implemented most often. Figure 3.2 shows how often the five categories of features were implemented in
Table 3.1: Overview of apps with highest number of features.

<table>
<thead>
<tr>
<th>#</th>
<th>App</th>
<th>App store</th>
<th>Price (€)</th>
<th>Number of features</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Endomondo</td>
<td>Google Play / iTunes</td>
<td>Free</td>
<td>19.5</td>
</tr>
<tr>
<td>2</td>
<td>Speedo Fit</td>
<td>iTunes</td>
<td>Free</td>
<td>18.5</td>
</tr>
<tr>
<td>3</td>
<td>Nike+ Fuel</td>
<td>iTunes</td>
<td>Free</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Pedometer++</td>
<td>iTunes</td>
<td>Free</td>
<td>18</td>
</tr>
<tr>
<td>4</td>
<td>Runkeeper</td>
<td>Google Play / iTunes</td>
<td>Free</td>
<td>17.5</td>
</tr>
<tr>
<td>5</td>
<td>Fitbit</td>
<td>Google Play / iTunes</td>
<td>Free</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Map My Fitness</td>
<td>Google Play / iTunes</td>
<td>Free</td>
<td>17</td>
</tr>
<tr>
<td>6</td>
<td>Running Plan by Gipis Coach</td>
<td>iTunes</td>
<td>Free</td>
<td>16.5</td>
</tr>
<tr>
<td>7</td>
<td>Health Mate</td>
<td>Google Play / iTunes</td>
<td>Free</td>
<td>16</td>
</tr>
<tr>
<td>8</td>
<td>FitStar Personal Trainer</td>
<td>iTunes</td>
<td>Free</td>
<td>15.5</td>
</tr>
<tr>
<td></td>
<td>Strava GPS Running and Cycling</td>
<td>Google Play / iTunes</td>
<td>Free</td>
<td>15.5</td>
</tr>
<tr>
<td>9</td>
<td>CARROT Fit</td>
<td>iTunes</td>
<td>€2.99</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>NexTrack: Making Exercise Fun</td>
<td>Google Play</td>
<td>Free</td>
<td>15</td>
</tr>
<tr>
<td>10</td>
<td>Mywellness</td>
<td>Google Play / iTunes</td>
<td>Free</td>
<td>14.5</td>
</tr>
<tr>
<td></td>
<td>Runtastic GPS running***</td>
<td>Google Play / iTunes</td>
<td>Free</td>
<td>14.5</td>
</tr>
<tr>
<td></td>
<td>Shapelink Fitness Journal</td>
<td>Google Play / iTunes</td>
<td>Free</td>
<td>14.5</td>
</tr>
<tr>
<td></td>
<td>Sports Tracker</td>
<td>Google Play / iTunes</td>
<td>Free</td>
<td>14.5</td>
</tr>
</tbody>
</table>

The 169 apps. Clearly, the categories Measuring & Monitoring, Information & Analysis and Support & Feedback were well represented, with 141 to 161 apps (approximately 83% to 95%) that include at least one of the features from that category. Social features were implemented in 121 of the 169 apps (72%). However, adaptation was part of only 7 out of the 169 apps (4%).

Figure 3.2: Frequencies of the five categories of features implemented in the apps.
Table 3.2: Overview of apps with lowest number of features.

<table>
<thead>
<tr>
<th>#</th>
<th>App</th>
<th>App store</th>
<th>Price (€)</th>
<th>Number of features</th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td>30 Day Fitness Challenges (Happy-Dev38)</td>
<td>Google Play</td>
<td>Free</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Boss Fit Solo</td>
<td>Google Play / iTunes</td>
<td>€2.24</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Extreme Bootcamp! With Lacey Stone</td>
<td>Google Play / iTunes</td>
<td>Free</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Global Cycle Coach</td>
<td>Google Play / iTunes</td>
<td>Free</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Scientific 7 Minute Workout</td>
<td>Google Play / iTunes</td>
<td>Free</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>The Ultimate Workout 1</td>
<td>iTunes</td>
<td>€1.99</td>
<td>2</td>
</tr>
<tr>
<td>36</td>
<td>7 Minute Bootcamp Workout</td>
<td>iTunes</td>
<td>Free</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>7-Minute Workout</td>
<td>Google Play / iTunes</td>
<td>Free</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>Cardio Fitness Exercises</td>
<td>Google Play / iTunes</td>
<td>Free</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>ENERGETICS Training</td>
<td>Google Play / iTunes</td>
<td>Free</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>Fitness Trip</td>
<td>Google Play</td>
<td>Free</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>HIIT - intervalworkout</td>
<td>Google Play / iTunes</td>
<td>Free</td>
<td>1.5</td>
</tr>
<tr>
<td>37</td>
<td>5 Minute Morning Workout routines</td>
<td>Google Play / iTunes</td>
<td>Free</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>10 Daily Exercises</td>
<td>Google Play / iTunes</td>
<td>Free</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Spin It</td>
<td>iTunes</td>
<td>€0.99</td>
<td>1</td>
</tr>
<tr>
<td>38</td>
<td>7 Minute Workout (mphan)</td>
<td>Google Play</td>
<td>Free</td>
<td>0.5</td>
</tr>
</tbody>
</table>

The next five figures show in more detail how often the features in these five categories were applied in the set of 169 apps, ordered by the most frequently implemented features. Figure 3.3 shows that many apps made use of user input, and a reasonable number incorporated data from a built-in sensor, but external sources were used very rarely. In Figure 3.4, we see that the user’s data was usually summarized textually or numerically, and less often in a visual format. Figure 3.5 shows that real-time feedback was usually offered as audio, whereas feedback on the overall progress of the user’s behavior is usually presented in textual format. The type of feedback that was used most often was generic advice or tips about physical activity, whereas gamification, punishment and context-aware feedback were very rare among the apps. It also shows that users more often got to set their own goal, rather than the app deciding on a personal goal for them. In Figure 3.6, we see that very few apps incorporated some form of adaptation to the user. Figure 3.7 indicates that sharing workouts or achievements among users was quite common; other social features (e.g., ranking or competition among users) were much less prevalent.
3.3.4 Differences between types of apps

The average number of features implemented in free apps was 8.26 out of 37, and paid apps contained 7.63 features on average. Using a two-tailed Mann-Whitney U test, we determined that this difference was not significant ($p = 0.64$). The free apps received an average rating of 3.32 out of 5.0 for usability, whereas the paid apps were marked with an average rating of 3.23 out of 5.0. This difference was again not significant ($p = 0.66$).
The average number of features implemented in apps that were only available in the Google Play Store was 6.95, and apps that were only found in the iTunes App Store contained 8.83 features on average. This difference was not significant ($p = 0.07$). The apps from the Google Play Store were marked with an average rating of 3.23 out of 5.0, whereas the apps from the iTunes App Store received an average rating of 3.32 out of 5.0. The difference was again not significant ($p = 0.56$).

### 3.3.5 Correlation between number of features and usability rating

The range of the number of features implemented in the eligible apps was between 0.5 and 19.5. The reviewers’ ratings for the two usability items (clarity and attractiveness) both spanned the entire range from 1.0 to 5.0, with an average of 3.5 and 3.1 respectively. Using Spearman’s rank correlation, we determined that there was no correlation between the number of features implemented in an app and its rating for clarity ($r_s = 0.085$, $p = 0.138$). However, there was a moderate positive correlation between the number of features and rating for attractiveness, which was statistically significant ($r_s = 0.477$, $p < 0.001$).

### 3.4 Discussion

The current review was designed to investigate the use of technological features in apps available through the Google Play and iTunes app stores that aim to promote a physically active lifestyle. To do so, a framework of techniques and functionalities that can be used to monitor or encourage physical activity was constructed. This framework organized the features in five different categories, to be able to evaluate the apps on a higher level. Two additional categories were added to the framework to collect additional information about the apps.

The 169 apps included in the review implemented approximately 8 features out of the 37
Chapter 3. Review of technological features in apps

in the framework on average. The highest number of features found in an app was 19.5 and
the lowest number was 0.5. Disagreements on the presence of a feature between reviewers
were reflected by awarding a score of ‘0.5’. The features that were implemented most often
were user input (to log activities or to form a personal profile), a textual/numerical overview
of the user’s behavior and progress, sharing achievements or workouts in internal or external
social networks, and general advice on physical activity. The features that were identified
least often were adaptation, integration with external sources, and encouragement through
gamification, some form of punishment or the possibility to contact an expert through the
app. There were no differences found between apps from the Google Play or the iTunes
app store, nor between free and paid apps. We determined that there was no correlation
between the number of features implemented in an app and the reviewers’ rating for its
clarity. Apparently, the ease to find information in an app is not compromised by the number
of implemented features. However, there was a moderate positive correlation between the
number of features and the rating for attractiveness. This implies that relatively simplistic
apps (in terms of numbers of implemented features) are generally less visually appealing
than apps with a larger number of implemented features. Although the ratings are based
on the subjective evaluation of only two reviewers, these results could indicate that more
sophisticated apps (that are equipped with more features) are developed with more care for
their visual design as well.

The results demonstrate that some features or categories of features are applied quite
often, but other functionalities are almost never implemented. Examples of features that
were rarely present in the reviewed apps are integration with information from external
sources, such as the user’s calendar or the local weather forecast, and adaptation to the
user’s behavior. All of these rarely implemented functionalities could enhance the feel of
the app being an intelligent virtual personal coach, since they imply a better understanding
of the user’s personal context and progress. This is in line with the wishes and expectations
of users regarding physical activity apps (Middelweerd, van der Laan, et al., 2015), and
therefore suggests an important area of possible improvement of physical activity apps.
Although we did not investigate why certain features were or were not implemented, a
plausible hypothesis is that it is related to the technical and conceptual complexity of the
implementation. Future work could provide more insight into this question.

The interpretation of the apps’ score in terms of implementation of technological features
depends on the research question under consideration. For example, one could investigate
whether the presence of specific features is related to positive user experiences (e.g., as
provided via the ratings in the app store), or whether some features are correlated with the
effectiveness of the app. Up to now, the effectiveness of such features in physical activity
apps is unknown. Therefore, it is not (yet) possible to argue which apps are more effective
than others based on the implemented features. A plausible hypothesis is that elements from
each category are necessary to create a versatile and complete app. Also, some subitems
seem superior to others: for example, automatic registration of physical activity through
(built-in) sensors is usually more user-friendly than manual input of activities. However,
such evaluations also depend heavily on user preferences, robustness of (the implementation
of) the technology and the objectives of an app, and are therefore difficult to claim on a
global level. Nonetheless, the proposed framework provides a valuable tool in such more
specifically motivated evaluations of apps.

One of the limitations of this study – and this type of research in general – is that it
provides a snapshot of the landscape at a certain time point. This means that the results that are valid now could be different after some time. In addition, because the number of health and fitness apps has grown to unmanageable numbers, a search strategy is necessary to find a selection of apps to review. This inevitably implies that not all apps can be covered, and certain eligible apps can be missed. Another limitation is that the review depends on the visibility of the features: if certain features are used in the background, they can be missed by the reviewers. Also, the results are susceptible to the reviewers’ interpretation. This is reflected in the ‘half’ features in the scores and the imperfect inter-rater reliability. In future work, interviews with app developers could reveal in more detail which features they did or did not implement, and for what reason.

Among the strengths of the present review are the large number of apps covered in the screening (n = 2,517), the relatively large number of apps analyzed (n = 227) and included (n = 169) in the review, which was performed by multiple independent reviewers, and the inclusion of both free and paid apps from the two largest app stores. Moreover, the rating of the apps was done based on downloading and using all functions of the app, rather than considering the app description and screenshots only.

Another contribution of this work is the proposed framework. Although this framework’s relevance is also subject to advances in modern technology, it provides a basis for the prevalence of technological features in physical activity apps. In addition, the hierarchical organization of the items in the framework allows for analysis of the apps on different levels of abstraction. Depending on the research question under consideration, one could investigate only the main items or only the subitems, or for example focus on coverage across all categories rather than a simple count of the number of implemented features. These possibilities of different perspectives pave the way for further analysis of the current range of physical activity apps.

The findings of this review are in line with other content analyses of physical activity apps. Although they generally focused on the application of behavior change techniques rather than technological features, these reviews also established that apps are generally lacking in such generic techniques for behavior change (Conroy et al., 2014; Cowan et al., 2013; Direito et al., 2014; Middelweerd, Mollee, et al., 2014; West et al., 2012). Similarly to the results of this study, these reviews showed that there is considerable room for improvement of the content of physical activity apps. Overall, it seems that smartphones provide a wide range of possibilities for more intelligent physical activity promotion interventions, but the developers of such apps are not yet taking full advantage of them.
workouts in social networks, and general advice on physical activity. The features that were identified least often were adaptation, integration with external sources, and encouragement through gamification, some form of punishment or the possibility to contact an expert. Furthermore, we determined that there was no correlation between the number of features implemented in an app and the reviewers’ rating for its clarity, but there was a moderate positive correlation with the rating for the app’s attractiveness. Overall, the results indicate that physical activity apps can be enhanced substantially in terms of their utilization of the possibilities that current mobile technology offers.

Acknowledgments

The authors would like to thank Rukshar Wagid Husain for her contribution to reviewing the apps and her input in the discussions about the framework.
References


Chapter 3. Review of technological features in apps


Appendix – Framework of Functionalities

1. Measuring & Monitoring
   a. Does the app ask for input from the user?
      i. Does it ask for input to set up a personal profile?
      ii. Does it ask for input to log activities?
   b. Does the app use built-in smartphone sensors to collect data?
      i. Does it use a motion sensor, e.g. a (combination of a) gyroscope and/or accelerometer?
      ii. Does it use a location sensor, e.g. based on GPS?
      iii. Does it use any other built-in sensor?
   c. Does the app support the use of external sensors?
      i. Does it support the use of external sources on the Internet or other apps?
         1. Does it support integration with weather forecasts?
         2. Does it support integration with calendars?
         3. Does it make use of location-specific information?
         4. Does it support the use of information from other sources on the Internet?
         5. Does it support the use of information from other (physical activity) apps?
      ii. Does it support the use of external hardware?

2. Information & Analysis
   a. Does the app show data of the user’s physical activity?
      i. Does it show data in graphical format (graphs, diagrams, etc.)?
      ii. Does it show the data in textual/numerical format (percentages, etc.)?
   b. Does the app provide a summary of the user’s progress?
      i. Does it do so in visual format (graphs, calendars, etc.)?
      ii. Does it do so in textual/numerical format (list of logs, statistics, etc.)?

3. Support & Feedback
   a. Does the app offer tips and advice on physical activity?
   b. Does the app offer real-time feedback?
      i. Does it do so in auditory format?
      ii. Does it do so in textual format?
   c. Does the app enable the user to set goals?
      i. Does it allow the user to set their own goal?
      ii. Does it set personal goals for the user?
   d. Does the app offer feedback on the user’s progress?
      i. Does it do so in visual format (color coding, smileys, etc.)?
      ii. Does it do so in textual format (messages, etc.)?
   e. Does the app give the user feedback based on their physical/social context?
   f. Does the app give some sort of punishment for not meeting goals/commitments?
   g. Does the app support physical activity in game format?
      i. Does it do so in terms of achievements, badges, etc.?
      ii. Does it do so in terms of a virtual game environment?
h. Does the app enable the user to contact an expert?

4. Adaptation
   a. Does the app adapt the user’s goals automatically, based on the user’s behavior?
   b. Does the app implement any other form of adaptation?

5. Social
   a. Does the app make use of an integrated social network?
   b. Does the app support a connection to an external social network (e.g., Facebook, Twitter)?
   c. Does the app allow the user to send messages to other users?
   d. Does the app allow the user to share information, experiences or achievements?
   e. Does the app enable competition between users?
   f. Does the app include a ranking between users?
   g. Does the app offer any other social functionalities?

6. Usability
   a. Is the app clear? (Scale 1-5.)
   b. Is the app appealing? (Scale 1-5.)

7. Other
   a. Is there a web-based version of the app, where the user can view (more) detail information?
   b. Does the app offer in-app purchases?
Abstract

Physical activity has many well-known health benefits. The transition from adolescence to early adulthood is a critical period in which there is a decline in physical activity. College and university students make up a large segment of this age group. Smartphones may be valuable aids to promote and support physical activity among this subpopulation. Therefore, the purpose of this qualitative study was to explore the preferences of Dutch young adults for a physical activity smartphone application (PA app).

Thirty Dutch students (aged 18–25 years) used a PA app for three weeks and subsequently attended a focus group discussion (k = 5). To streamline the discussion, a discussion guide was developed covering seven main topics, including general app usage, usage and appreciation of the PA app, appreciation of and preferences for its features and the sharing of PA accomplishments through social media. The discussions were audio and video recorded, transcribed and analyzed according to conventional content analysis.

The participants, aged 21±2 years, were primarily female (67%). Several themes emerged: app usage, technical aspects, PA assessment, coaching aspects and sharing through social media. Participants most often used social networking apps, communication apps and content apps. They preferred a simple and structured layout without unnecessary features. Ideally, the PA app should enable users to tailor it to their personal preferences by including the ability to hide features. Participants preferred a companion website for detailed information about their accomplishments and progress, and they liked tracking their workout using GPS. They preferred PA apps that coached and motivated them and provided tailored feedback toward personally set goals. They appreciated PA apps that enabled competition with friends by ranking or earning rewards, but only if the reward system was transparent. They were not willing to share their regular PA accomplishments through social media unless they were exceptionally positive.

In conclusion, this study showed that the participants prefer PA apps that coach and motivate them, that provide tailored feedback toward personally set goals and that allow competition with friends.
This chapter appeared as:

4.1 Background

The positive effects of regular physical activity (PA) are well known, yet many people do not comply with PA guidelines (World Health Organization, 2010, 2014). Sixty-four percent of Dutch young adults (aged 18–34 years) meet the guidelines for being physically active at a moderate intensity for at least 30 minutes per day and at least 5 days per week (Hildebrandt et al., 2013). The transition from adolescence to adulthood and the period of early adulthood itself is a critical period during the life course where the decline of PA appears to accelerate (Kwan et al., 2012; Larouche et al., 2012). Previous studies indicate that the rate at which PA decreases varies between men and women, and men who transition into a university are more likely to adopt a less physically-active lifestyle (Kwan et al., 2012). In the Netherlands, many students who enter university move away from home, start to live on their own or in student housing communities and combine their study obligations with part-time jobs and social commitments. This may result in a reduction in free time that was previously available for PAs (Allender et al., 2008).

Smartphones and smartphone applications (apps) are popular, especially among highly educated young adults (Centraal Bureau voor de Statistiek, 2013) and offer new possibilities for promoting PA. The rapidly growing number of fitness apps that are commercially available indicate their popularity (Steel Media Ltd., 2015). However, recent studies and a systematic review show that most of them are minimally based on established behavior change theories and techniques (Breton et al., 2011; Cowan et al., 2013; Middelweerd et al., 2014; West et al., 2012). The review by Middelweerd et al. (2014) further demonstrates that when established behavior change techniques are included, self-monitoring (e.g., GPS, diary, or accelerometer), goal-setting features and social support by connecting with social networking sites (e.g., Facebook or Twitter) were applied most frequently.

A small number of studies examine the usability and effectiveness of PA apps to increase PA in healthy (young) adults (Casey et al., 2014; L. G. Glynn, Hayes, Casey, F. Glynn, Alvarez-Iglesias, Newell, ÓLaighin, Heaney, O’Donnell, et al., 2014; Kirwan et al., 2012; Morrison et al., 2014; Spook et al., 2013). L. G. Glynn, Hayes, Casey, F. Glynn, Alvarez-Iglesias, Newell, ÓLaighin, Heaney, O’Donnell, et al. report significant increases (1029 steps) in daily step activity in the intervention group using an app that offered feedback graphs and continuous feedback (L. G. Glynn, Hayes, Casey, F. Glynn, Alvarez-Iglesias, Newell, ÓLaighin, Heaney, O’Donnell, et al., 2014). Kirwan et al. conclude that a smartphone app in addition to website-delivered intervention could enhance engagement and increase levels of PA (Kirwan et al., 2012). Thus far, PA app interventions are commonly used as supplemental tools, complementing the primary goal of keeping track of personal goals (Kirwan et al., 2012; Morrison et al., 2014), making ecological momentary assessments (Spook et al., 2013) or providing feedback for current behavior (L. G. Glynn, Hayes, Casey, F. Glynn, Alvarez-Iglesias, Newell, ÓLaighin, Heaney, and Murphy, 2013). Yet, little is known about the preferences of young adults for specific behavior change techniques applied in a PA app that stands on its own.

Social networking sites such as Facebook and Twitter are popular among Dutch young adults: 98% use Facebook and/or Twitter (Centraal Bureau voor de Statistiek, 2013). Many PA apps offer the possibility of sharing one’s activities through social media. However, little is known about whether Dutch students like the possibility of sharing PA-app-based tracking of their activities through social media and if they actually share their results.

Developing a theory-based, effective and engaging PA app that is also based on user
preferences and opinions is a complex process, as are all thoroughly-developed theory- and evidence-based interventions (Bartholomew et al., 1998). A recent review identifies key features that facilitate PA engagement: real-time feedback, social networking, expert consultations and goal setting. In addition, disruptive prompts, text messaging and competition-based strategies reportedly limit engagement in PA (Bort-Roig et al., 2014). However, little is known about the usage, appreciation and preferences of students (aged 18–25 years) for various features in such apps. Understanding their needs, expectations and preferences is the first step in designing more effective PA apps.

This study aimed to gain insight into the role, usability and appreciation of an existing PA app that allows sharing of activities through social media, called Nexercise (Nexercise Inc., 2014). Three research questions were addressed:

1. How do Dutch bachelor’s and master’s students (aged 18–25 years) use and appreciate the various features of an existing PA app?
2. What are the preferences of Dutch bachelor’s and master’s students regarding a new PA app?
3. How do Dutch bachelor’s and master’s students use and appreciate the option to share accomplishments through social media?

4.2 Methods

4.2.1 Design

A qualitative design was used to explore respondents’ preferences, attitudes and experiences regarding PA apps; for this reason, focus group discussions were the chosen format (Hsieh and Shannon, 2005). To ensure meaningful focus group discussions, participants must have had some experience with a PA app. They were asked to download the Nexercise app (version 2.2.3; www.nexercise.com) (Nexercise Inc., 2014) and then use it during the three weeks preceding the discussions. The Nexercise app is a GPS fitness tracker that can be used for a variety of sports activities such as fitness, running and horseback riding, and contains multiple options such as GPS tracking, activity log book, earning points, a competition feature, chat features and linking with social media. This PA app was selected because (1) it was found to include behavior change theories and techniques in a recent review, such as prompting goal setting, prompting self-monitoring, providing feedback on performance, providing rewards and planning social support (Middelweerd et al., 2014); (2) it was freely available from both iTunes and Google Play and thus was compatible with both iPhones and Android smartphones; (3) it enabled tracking a variety of PA behaviors, so it was not focused on only one sport or activity; and (4) the app consisted of multiple features, including GPS tracking, rewarding, ranking, chat and the possibility of sharing results. Participants were asked to use the app when engaging in PA and to post their accomplishments on their social media pages. Use of the app and sharing was completely voluntary, and participants were informed that they could participate in the focus group discussions whether or not they used the app.

The study was approved by the Medical Ethical Committee of the VU Medical Center, Amsterdam.
4.2 Methods

4.2.2 Recruitment

This study was conducted using Dutch bachelor’s and master’s students at the VU University, Amsterdam between April and June 2013. Eligibility required the participants be current students (bachelor’s or master’s) aged between 18 and 25 years, healthy and without contraindications for sports participation, own a smartphone with internet access, be a member of Facebook or Twitter, and have mastery of the Dutch language. The recruitment took place in person by distributing flyers and through online social media advertisements, and eligible persons were informed that they could receive an incentive for their participation (i.e., an arm holder for a smartphone and voucher for free entrance to the university sports center). An effort was made to include participants who were at various PA levels because this might affect their preferences for specific features of a PA app. Participants were divided based on whether or not they met the Dutch PA guidelines. The PA levels were assessed using the Dutch short version of the International Physical Activity Questionnaire (IPAQ) (Craig et al., 2003). Participants were considered to meet the Dutch PA guidelines if they reported at least 30 minutes of moderately intense PA daily for at least five days per week or at least 20 minutes of vigorous activity daily for at least three days per week (Hildebrandt et al., 2013). An effort was made to create homogeneous focus groups based on the participants’ PA levels according to the IPAQ, resulting in three groups comprising participants who met the guidelines and two groups comprising participants who did not meet the guidelines.

4.2.3 Procedures

To streamline the focus group discussions, a discussion guide was developed which included open-ended questions and prompts (statements) to encourage participants to share their opinions. The prompts aimed to provoke discussion about topics that were not yet covered and were used at the end of each discussion. Three prompts were used: 1) “I enjoy using a smartphone app during my sports activities, but only a couple of times. After a while I do not use the app anymore.”; 2) “Positive feedback on my physical activity achievements from my friends encourages me to be more physically active.”; 3) “It really annoys me when my friends on Facebook post their sports activities on their timeline.”. Table 4.1 provides an overview of the topics included in the discussion guide. An example of an open-ended question is: “When do you usually use the app and for what kinds of activities?”.

To obtain demographic characteristics, participants were asked to complete a short online questionnaire prior to the focus group discussion. The first page of this questionnaire contained information about the study and included an informed consent for the questionnaire, ensuring anonymity and confidentiality, and which required the participant’s signature before the remainder of the questionnaire could be completed.

Written informed consents for the focus group discussions were obtained prior to the discussions, which spanned one hour each and were led by a trained moderator (DMvdL) who was an age peer of the participants. Prior to the first focus group discussion, the moderator attended a workshop on qualitative research and pilot-tested the discussion guide under the supervision of a qualitative research expert (MS). During the discussions, the moderator assured that participants were aware of the purpose and procedures, noted that they were audio and video recorded and ensured confidentiality and anonymous transcriptions. Two additional researchers (TV and AM, TV and MMvS or TV and JSM) assisted with the
Table 4.1: Main topics of the focus group discussion guide.

<table>
<thead>
<tr>
<th>Number</th>
<th>Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>General smartphone application usage</td>
</tr>
<tr>
<td>2</td>
<td>General impression of Nexercise a</td>
</tr>
<tr>
<td>3</td>
<td>Usage and appreciation of Nexercise a</td>
</tr>
<tr>
<td>4</td>
<td>Usage and appreciation of various features</td>
</tr>
<tr>
<td>5</td>
<td>Preferences for various features</td>
</tr>
<tr>
<td>6</td>
<td>Social support through an application</td>
</tr>
<tr>
<td>7</td>
<td>Sharing through social media (e.g., Facebook or Twitter)</td>
</tr>
</tbody>
</table>

a Nexercise = fun and weight loss: physical activity smartphone application for iOS and Android.

discussions by acting as practical assistants and observers and took the opportunity to ask the participants questions, clarifying any remaining concerns at the end of the discussions. At the close of each discussion, participants were given forms thanking them for participating and asking for written comments and then were awarded with the incentive. The comments could include issues they wanted to share but did not mention during the discussion and any comments regarding the topics that were discussed or topics they thought should have been discussed. The members of the research team attending the discussion evaluated it by sharing first impressions and assessing the role of the moderator.

4.2.4 Data management and analysis

The recordings were transcribed verbatim (DMvdL) with pseudonyms for each respondent. The transcripts were checked for quality and completeness by another researcher (TV) and were analyzed according to conventional content analysis, generally used when little research has been done in the subject area and little is known (Hsieh and Shannon, 2005). Atlas.ti 6.0, software for qualitative analysis, was used to perform the analyses. First, the transcripts were read verbatim independently by two researchers (DMvdL and TV) to select relevant fragments based on the discussion guide. Various codes and subcodes were created with these fragments. Second, the codes were reviewed and split, combined, added or removed (DMvdL) if overlapping codes or better coding names were discovered. Third, the codes and subcodes were clustered and sorted into general themes (DMvdL), and a tree diagram was composed to provide a visual representation of the codes. Several meetings of the research team (TV, SJtV, MMvS, JSM, MS) were arranged so that consensus could be reached. All fragments were split according to the focus group discussion, which implies splitting data based on PA level (whether or not the participants met PA guidelines). This was done with the aim of finding remarkable differences between the groups. Once these differences were found and described, the data were combined for analysis as one dataset.
4.3 Results

4.3.1 General characteristics

Fifty-seven participants agreed to participate, yet 30 (53%) attended the focus group discussions. Figure 4.1 shows a flowchart of the study procedure, participant dropout rates and reasons.

![Flowchart](Image)

Figure 4.1: Overview of the number of participants initially agreeing to participate and drop-outs at various stages, resulting in 53% attending focus group discussions.

*Reasons for not completing the study included: no response \((n = 12)\), lack of time or commitment \((n = 6)\), technical problems with the Nexercise app or smartphone \((n = 2)\) and schedule conflicts for attending the focus group discussions \((n = 7)\).

The participants \((n = 30)\) were aged \(21 \pm 2\) years and were primarily female \((67\%)\) living in Amsterdam \((50\%)\) and did not meet PA guidelines \((57\%)\). The focus groups ranged from 4 to 7 members each. Participant characteristics are listed in Table 4.2. Within the groups comprising participants with higher PA levels, four participants (two male) were very active, reporting vigorously activity for at least 20 minutes 5 to 9 times weekly.

Table 4.2: Focus group characteristics per focus group discussion and for participants who dropped out of the study.

<table>
<thead>
<tr>
<th>Focus group</th>
<th>Number of participants</th>
<th>Gender (number of females)</th>
<th>NNGB(^1) (number meeting the guidelines)</th>
<th>Fit norm(^2) (number meeting the guidelines)</th>
<th>Number meeting NNGB and/or Fit norm(^3)</th>
<th>Sports</th>
</tr>
</thead>
<tbody>
<tr>
<td>NNGB</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fit norm</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number meeting NNGB and/or Fit norm</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Chapter 4. User preferences for apps that promote physical activity

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>7</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gym workout ($n = 5$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cycling ($n = 3$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Climbing ($n = 1$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Inline skating ($n = 1$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Pole dancing ($n = 1$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Physiotherapy exercises ($n = 1$)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>2</th>
<th>5</th>
<th>3</th>
<th>5</th>
<th>5</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>7</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Running ($n = 4$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Ice skating ($n = 2$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Field hockey ($n = 1$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Swimming ($n = 1$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gym workout ($n = 3$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Climbing ($n = 1$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Indoor soccer ($n = 1$)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>3</th>
<th>5</th>
<th>5</th>
<th>0</th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No sports ($n = 2$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Running ($n = 2$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gym workout ($n = 1$)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>4</th>
<th>2</th>
<th>0</th>
<th>1</th>
<th>1</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Running ($n = 3$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gym workout ($n = 2$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mountain biking ($n = 1$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Surfing ($n = 1$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Inline skating ($n = 1$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Spinning ($n = 1$)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>5</th>
<th>7</th>
<th>7</th>
<th>1</th>
<th>1</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>7</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No sports ($n = 3$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Running ($n = 2$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gym workout ($n = 1$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Handball ($n = 1$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Field hockey ($n = 1$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Spinning ($n = 1$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Box training ($n = 1$)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dropouts</th>
<th>27</th>
<th>14</th>
<th>18</th>
<th>17</th>
<th>20</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No sports ($n = 4$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gym workout ($n = 7$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Running ($n = 4$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Soccer ($n = 3$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Field hockey ($n = 2$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cycling ($n = 1$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Water polo ($n = 1$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Basketball ($n = 1$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Badminton ($n = 1$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Pole dancing ($n = 1$)</td>
</tr>
</tbody>
</table>

1. **NNGB** = Dutch health guidelines for physical activity (30 minutes of moderate activity at least 5 days per week.
2. **Fit norm** = 20 minutes of vigorous activity at least 3 days per week.
3. Number of participants who met the Dutch health guidelines for physical activity and/or the Fit norm.
4. The number of participants who dropped out during the study.
5. Engage in a fitness programme that is provided and supervised by a physiotherapist.
4.3 Results

4.3.2 General themes

Five general themes emerged in all focus group discussions: general app usage, technical aspects, PA assessment, coaching aspects and sharing through social media. Figure 4.2 provides an overview of the themes and subthemes. In general, the same topics were discussed, and similar themes emerged in all groups.

![Figure 4.2: Overview of the themes and subthemes discussed in all focus groups.](image)

4.3.3 App usage

App use in general

The types of smartphone apps most often used by the participants were social networking apps such as Facebook and Twitter, communication apps (e.g., WhatsApp) and content apps (e.g., news reports, weather forecasts or public transport information). Game apps were not very popular; respondents stated that such apps were amusing or pleasant for short-term use only. Some participants, primarily those meeting the PA guidelines, already used a comparable PA app such as RunKeeper, Endomondo or Strava.

Use of the Nexercise app

In all focus group discussions except that of Group 4, the participants had a positive attitude toward PA apps. Group 4 comprised participants with lower levels of PA, and these participants stated they did not need such an app. Participants with higher PA levels clearly believed that the Nexercise app would be useful for inactive people, to raise awareness that they need to be more physically active, for example. However, they perceived the Nexercise app to be less useful for themselves. Self-reported app use revealed that all participants used
the Nexercise app at least once with one exception. On average, it was used eight times (range 0–29).

The frequency of Nexercise app use varied among participants. For some, it became routine to start the tracking feature when they intended to exercise. “When the app was still working on my phone, it became natural to use the app” (male who did not meet PA guidelines). “For me the app usage decreased over time. I used to be a fanatical user by always starting the app, but after a while I couldn’t care anymore” (male who did not meet PA guidelines). For others, it was something they would easily forget. Often, the preceding mind-set appeared to determine the frequency of use (e.g., participants who found the app unnecessary, time consuming and not useful beforehand did not use it often or at all). Some participants mentioned that they exercised only for themselves and they did not require the support of a smartphone app.

The majority of the participants reported becoming more aware of their PA level, such as their activity patterns and duration of their activities. They also saw that short distances could make a difference. “I became more aware of the fact that I am actually pretty active. I thought that I was doing nothing, but afterwards I was not that bad as I thought I was” (female who did not meet PA guidelines).

During all focus group discussions, the participants were presented with the following statement at the end of the discussion: “I enjoy using a smartphone app for a short period, but after that I do not use the app anymore”. Almost all participants agreed with this statement, saying they experienced this feeling with almost all smartphone apps they used. At some point, the novelty disappeared, and when they experienced a problem with the app, such as a stuttering mobile phone, lack of storage or battery problems, they often stopped using it. “It does not personally add anything for me... so you will quit using it, you can see it as useless and something you want to get rid of” (male who did not meet PA guidelines). Participants mentioned that if the app was more tailored to their needs and if they gained added value from it, they would probably still use it.

### 4.3.4 Technical aspects

#### Design

In general, the participants preferred to have a simple and well-ordered app design. They wanted to have a structured layout with only a few important features which could easily and effortlessly log activities and obtain a clear overview of the results. Some participants wanted to customize it themselves. “Maybe you could adapt the starting page of the application and choose a quick forward button, so you could easily go to the option you prefer... with the possibility to add or remove additional options” (male who did not meet PA guidelines).

Most participants appreciated the enormous list of activities included in the Nexercise app, but they found the list to be rather limited for some activities, such as fitness trainings.

#### Calendar

Participants with higher levels of PA liked the idea of a calendar within the app, providing them with an overview of their accomplishments. Some participants did not need such a schedule because they could use their own agenda. “I used it most often to log afterwards,... like trips to school and stuff like that...” (male who met PA guidelines). “What I like as well is the calendar.... I don’t work out on a regular basis, so if you look back you have an overview on which days you did what kind of sports...” (male who met PA guidelines).
“But I usually write it down in my regular agenda, thus it is twice as much work to keep that diary as well” (female who met PA guidelines). “Those agendas should be merged” (male who did not meet PA guidelines). “Yes” (female who met PA guidelines). The participants who did not meet guidelines thought of the app in terms of replacement for a coach telling them what to do. Thus, they wanted the calendar function to make a training schedule and to set tasks for them.

**Reminders**

Most of the participants who had lower levels of PA perceived the reminders as annoying. One reason for this was the feeling that they were able to decide for themselves when they wanted to exercise. Another reason was the potential feeling of guilt that did not work for them; a third was that they did not want to be bothered with notifications reminding them to exercise. However, participants with higher levels of PA more often appreciated the reminders, although they also highlighted that they did not always come at appropriate moments. For instance, they came after recent PA or when it was late in the evening. Some participants reported that the reminders were tools that triggered them to make time to exercise or to fill out more activity results. “For me it was more like a reminder,... I need to fill in my diary today,... that I really got the feeling I need to work out tonight” (male who met PA guidelines).

**Companion website**

Almost all participants preferred having an account on a website in addition to the app. Reasons for this were that they could add data more easily and that it could present much more information about their activities, such as progress bars, activity schedules, graphics, maps and routes. “...that it finds a route...; I want 5 km and then the website tells you which route” (male who did not meet PA guidelines). Instructional videos, tips and forums came up in the discussions only among those with lower levels of PA. They indicated that such a website should be an additional support system where they could access more detailed and in-depth information and tips on how to perform a workout. In contrast, the participants with higher levels of PA preferred a website that presented additional and more detailed information about their workout (e.g., average speed and heart rate) and their progress. A few participants noted that such a website could be a barrier for use of the app. “Yes, an additional website for support with the option to set a goal and to reach it. But it is an additional barrier to go and visit the website” (female who did not meet PA guidelines).

### 4.3.5 Physical activity assessment

**GPS tracking vs. activity logging**

Nexercise provided various options for logging activities. Some participants consistently logged their activities after exercise because they knew the exact information they needed or because it was not possible to track their activities with GPS. Examples of such activities are swimming or playing soccer. “Actually, I only logged my activity afterwards and once I took it for a run, but for everything else like spinning I wouldn’t take it with me” (male who met PA guidelines). Additionally, some participants admitted that they often forgot to start the tracking feature, so they could log recent activities only. For others, it became routine to log activities at the end of the day. A small number found it very annoying to carry their smartphones with them during exercise. It was often mentioned that because tracking with
GPS consumes battery power, participants felt they had no choice and would log activities only after exercise. Some found it unnecessary to log activities because they had already completed their exercise. Some found it was more convenient to track their activities with GPS because the application automatically measured detailed information about speed and distance and showed real-time data. However, they reported that they often forgot to start the GPS tracking.

**Intensity and satisfaction**

Participants highlighted the importance of reporting the intensity of their activities afterwards: they found a big difference between having a training session or a match and doing an exercise just for fun. The intensity also had to be taken into account when calculating the points that could be earned. “You should fill out the intensity...; when I am following a spinning lesson, then a specific amount of points are rated for that activity, but I can exercise to the maximum or I can exercise at ease” (male who met PA guidelines). Participants who did not meet PA guidelines wanted to be able to add information after completing their activities, such as how they felt during the activity. “Maybe when you have finished running, you could indicate with a smiley how you felt during the exercise” (female who did not meet PA guidelines).

**Extra device**

A couple of participants wanted to use the app in combination with another device, such as a pedometer or heart rate monitor. Most who mentioned this functionality were already physically active. Others found it unnecessary and were not willing to pay extra for such a device.

**4.3.6 Coaching aspects**

**Coach**

A coaching feature generally was seen as a huge advantage in a PA smartphone app. Some participants preferred the attention from a live personal coach or the support of friends during their activities. However, they recognized that if this was not possible, a coaching feature in an app is the second-best option. Opinions differed as to whether this coaching feature should provide support during or after PA. Some said they would find it annoying to hear a coach during their activities, primarily because they felt a device should not speak to them. However, most preferred to hear a motivating and enthusiastic voice giving information about their speed, distance or progress and making encouraging statements during PA. “A coach who really encourages you, who is saying that you are doing a good job and who tells you to see you the next time, that is really nice” (female who did not meet PA guidelines).

All participants liked the idea of a coaching feature, but depending on whether they met PA guidelines, they wanted it presented in a slightly different way. Those who did not meet PA guidelines wanted a coaching feature that would stimulate them to reach their goals, encourage them to keep going and provide them with tips about healthy exercising. Those who met PA guidelines wanted a coaching feature that would give detailed information about their workouts and tips on how to intensify the exercise as well as information about sporting events in the neighborhood.
4.3 Results

Goal setting
Almost all participants preferred a coach in combination with goal setting. Most preferred to set goals when using the app. They wanted to choose between different goals or to be able to make a new goal, such as losing weight, improving fitness or keeping up with a specific activity schedule. They highlighted that if they could set a goal, they wanted the app to work as a coach by reminding them to exercise or to tell them what their progress was. It was very important to them to make a schedule, to set a task and to work toward reaching goals. Those who did not meet PA guidelines, in particular, preferred a goal-setting feature. They highlighted that they really needed to set goals and to be guided in reaching these goals. “It is very important for me to set goals... with a graphic representation, like a bar, for example, you have a guideline to exercise a specific amount of hours per week, then it would be very good to see, ‘oh, right now I am in the red zone or the orange zone’, and when I am progressing, ‘I am in the green zone’” (female who did not meet PA guidelines). Those who met PA the guidelines reported that goal setting was unnecessary.

Feedback and motivational features
Most participants would have liked to have some personal feedback from a coach after completing their activities. Examples of such personal feedback included compliments, reporting their progress and helping them with their schedule and reaching their goals. Adding tips to the app about how to reach goals, how to make activities more fun, how to exercise safely and when it is best to exercise would be desirable assets, according to the participants.

In addition, most reported a desire to add more information about themselves before using the app, such as their motivation level, experience level, desired goals and weights and heights. “Maybe you could first fill out something about yourself, for instance how motivated you are and whether you are feeling good at the moment” (female who did not meet PA guidelines).

In addition, they wanted to receive more detailed information about their activities afterwards. For instance, graphic visualizations of their progress, burned calories, a map of the route taken and speed and distance information. The group of participants who did not meet PA guidelines, in particular, preferred information about the number of calories burned during a workout. Information about the environment, such as operating hours of sports facilities, was identified as less important because they already knew it or could search for that type of information on the Internet. Opinions as to whether the app should offer information about the weather were diverse: for some, it would be helpful if the app could take the weather forecast into account when scheduling activities, but for others, it made the application less clear, and they could use the Internet just as easily. Some stressed that information about sporting events in their neighborhood was appreciated.

Some participants suggested a music feature during their activities that could be interrupted by the coach. When doing a good job, this music feature could reward them with a ‘power song’, motivating them to keep going.

Competition
Most participants found the ranking feature interesting and motivating. They experienced this ranking as a match in which they did not want to be inferior to their friends. “Yes it is a little bit shocking when you noticed that your friends did a good job” (female who did
not meet PA guidelines). “Haha, I would go for a workout because it is confronting and because I want to be physically active...” (female who did not meet PA guidelines). “Yes, it encourages me. A friend of mine is jogging quite often, so when I see she did some exercise, it motivates me to go exercising again” (female who did not meet PA guidelines).

A few participants reported the ranking feature as unimportant and something they did not need. They found it only interesting to compare their results with themselves and not with others. Additionally, because of their lack of time, they wanted to focus only on exercising and not on playing a game. Some participants who did not meet PA guidelines found it confrontational, leading to either a decrease or increase in PA.

A couple participants intended to continue using the app after this study. Their reasons were that (1) they started a competition with their friends that they wanted to continue or (2) they used the app to document their exercise progress.

Rewards
Most participants liked earning points according to their exercise. Receiving an award was perceived as motivational and as input for a competition with friends. “But it motivates me to log my activity, if you are going to the next level when you are filling in your activities. ... Yes, I like that” (male who met PA guidelines).

For some of the participants, it was unnecessary to get rewarded with points for being physically active. “Yes, it doesn’t mean anything to me” (female who met PA guidelines).

What bothered most participants was that if they were rewarded with points, it was unclear how these points were calculated. They preferred that the number of points represent the type and intensity of the activity. Most wanted to receive real rewards instead of virtual rewards, such as discount vouchers for sporting goods stores, gift cards or tickets to sporting events. Some wanted to earn points that reflected their burned calories.

Chat
Participants were clear about whether the app should have a chat feature. “The idea is okay, but nobody uses it, so, yes, you don’t need it” (female who did not meet PA guidelines). They were unanimous that the chat function was a needless feature and a waste of time. They highlighted that if they wanted to chat, they would use other apps.

4.3.7 Sharing through social media
Reason for sharing through social media
Some participants reported that they occasionally shared their PA achievements through social media (primarily Facebook). The main reasons for sharing their results were that they were proud of their accomplishments or that they wanted to share information about, for example, their running or cycling routes with friends. The perceptions of their feelings if posts were liked or responded to were diverse. Some reported that it would motivate or support them, while others reported that it would not make a difference. Those who did not meet PA guidelines, in particular, acknowledged that they liked getting Facebook likes for their achievements, and they stated it would make a difference in their PA behaviors.

If participants shared their achievements through social media, they preferred doing so with personally typed messages, maps of their routes or photos. They also highlighted that sharing their achievements via the app seemed unnecessary, because they could share it via Facebook themselves.
4.4 Discussion

Though some participants shared some of their achievements through social media using other smartphone apps, almost nobody shared them via the Nexercise app. In each focus group discussion, there was strong agreement that people should post only exceptional results, such as winning a match, becoming a champion, participating in a marathon or reaching a desired goal. The main reason for this preference was their annoyance at people who post all types of information (e.g., training results or walking to the bus stop), and they did not want to be perceived as that type of person. “Yes indeed, why do others need to know,... it is like, oh I did some sports... It is a little bit stupid. That’s when I think to myself, nobody needs to know...” (female who did not meeting PA guidelines).

Other reasons included being physically active for oneself, being embarrassed by the results, feeling it was not worth mentioning and feeling it was just as easy to tell friends that type of information in person.

Private community

Many participants reported that they found most posts of others in their social media communities as annoying and something they were not interested in. They highlighted that this feeling depended on who shared the information (e.g., close friends or training buddies). They also reported that information shared by others about an exceptional accomplishment or reaching a goal was seen as something interesting to read. Therefore, in almost every focus group discussion, sharing achievements in a private community through social media was raised. Almost all participants reacted quite positively to the idea, and they were willing to form such a group with their closest friends, friends who were interested, people with the same goals, people using the same application, people with the same fitness level, people from the same sports club or people from their area of study. They envisioned that they would receive social support when part of a community with similar interests.

4.4 Discussion

This study explored the use and appreciation of and the preferences for various features of a PA app by Dutch students (aged 18–25 years). As expected, based on the popularity of health and fitness apps, participants expressed positive attitudes toward a PA app. In general, they liked the idea of a PA app. Those who met PA guidelines thought that it was more useful to others than to themselves, stating that PA apps such as Nexercise could raise awareness for those who are not physically active, but that they are not suitable for themselves. Those who did not meet PA guidelines highlighted a desire for a personal coach function to help them achieve their self-determined goals, whereas those who met the guidelines preferred detailed training information, such as how to intensify their training sessions. Almost all participants preferred a companion website that could give detailed and general information about their behaviors.

The preferences for motivational features agree with those found in previous research; participants preferred self-monitoring and goal-setting features. Ehlers and Huberty (Ehlers and Huberty, 2014) note that middle-aged women (mean age, 40.7 years; SD, 10.3 years) prefer a smartphone app that includes features to track their behavior and to set goals; however, these women are less interested in motivational features or features to overcome barriers. Rabin and Bock (Rabin and Bock, 2011) report similar results based on their study of fourteen adults (aged 23–60 years) who used three PA apps and felt that the ideal app
should apply to different types of activities, be easy to use, track activity automatically and set goals. Features that target self-regulatory principles (e.g., self-monitoring, goal setting, behavioral feedback and problem solving features to overcome barriers) have been used successfully in PA promotion interventions. King and colleagues (King et al., 2014) demonstrate in a small study population that an app using self-regulatory principles is able to increase overall moderate-vigorous PA in aging adults. Kirwan and colleagues reported significant differences in daily step activity in favor of the intervention group, which monitored their step activity (Kirwan et al., 2012). These results suggest that a PA app that uses self-regulatory principles could successfully increase PA in young adults.

Although 98% of Dutch young adults (aged 18–25 years) actively use social media (Centraal Bureau voor de Statistiek, 2013), our study participants were not willing to share all their accomplishments on Facebook, suggesting that linking to social networking sites should not be a primary feature in PA app interventions. These results agree with those of Cavallo and colleagues (Cavallo et al., 2012), who conclude that their intervention among students aiming to increase social support for PA with online social networking did not improve perceptions of social support. Our participants reported that Facebook is not an appropriate platform to share their achievements because everybody is able to read their status updates. A private community could offer the possibility of sharing goals and achievements with peers. Further research is needed to explore whether such private social media communities could enhance social support, therefore enhancing the effects of PA apps in young adults.

**Implications for future interventions**

This study suggests the need for an app in the form of a virtual coach to guide users who do not meet PA guidelines and to help them to overcome barriers, reach self-determined goals or monitor their progress. Feedback that is normally provided face-to-face by a personal coach should be integrated into the virtual coach. Besides the personalized and tailored feedback, the feedback should be rated as credible and trustworthy. Translating face-to-face feedback into a virtual coach requires a highly detailed diagnostic assessment for translating the information in a series of “if-then” messages that are linked to feedback messages and techniques for increasing PA.

In addition to an initial diagnostic assessment, the participants preferred ongoing assessments to adjust the feedback messages over time. The initial diagnostic assessment should be based on self-reported data to assess PA level, to identify barriers and to assess daily emotions. However, it should also be based on objective measures to assess the behaviors using GPS and/or an accelerometer. For future interventions, researchers and programmers will be challenged to build an appealing and engaging app that includes a diagnostic assessment able to gain detailed information with minimal burden on the participant and that will be used over a long period of time. However, because the majority of participants perceived the app to be enjoyable for a short period of time, more research is needed to examine whether a PA app alone is a promising tool for achieving long-term behavior change or if it should be combined with other channels, such as a face-to-face program. All participants identified features that would enhance the attractiveness of a PA app, such as self-monitoring features, competition features and goal-setting features. Competition may have been less-preferred by those who did not meet PA guidelines because it was perceived as confrontational by some. Therefore, when intervention designers add a competition feature, they should consider who
would participate in the competition, so that the competition will be motivational and not frustrating.

**Strengths and limitations**

A strength of this qualitative study is its ability to explore the students’ opinions, beliefs and experiences regarding PA apps. To our knowledge, this was the first to explore students’ appreciations and preferences, and therefore provides valuable information for future app-based interventions.

A limitation, also related to its qualitative exploratory character, is that findings cannot be generalized, certainly not beyond the population of Dutch university students. To increase the generalizability to the Dutch young adult population, research should examine these appreciations and preferences among young adults in other groups within this age range. Furthermore, quantitative observational research and interventional studies in larger samples of young adults should be conducted to test our findings, including objectively measuring app usage. This study included a small sample size because of a high drop-out rate (47.4%) which may have created selection bias. A more representative sample may have led to different results, thus the app features we found to be desirable may not meet the needs for all potential app users. However, given that no new information was retrieved from the last focus group discussion, data saturation most likely was reached at least for the population of Dutch university students.

**4.5 Conclusions**

In conclusion, this study provides exploratory insights into the preferences of students regarding a PA app. Apps aiming to increase PA in young adults should provide personalized and tailored feedback and include a coaching function. A well-oriented and easy-to-use design must be developed, with the option to customize the application. Preferred features to be included in an application are ranking features, a coaching feature through which users are motivated during the exercise and receive feedback afterwards, and the possibility to set goals and to work with a schedule. In addition, participants prefer a website that accompanies the app to provide overviews of their results and progress. There is little need for a sharing feature to post results through social media.

**Competing interests**

The authors declare that they have no competing interests.

**Authors’ contributions**

AM provided intellectual input to focus group design, assisted with the focus group discussions, drafted the manuscript and incorporated all feedback. DMvdL designed and executed focus group discussions and transcribed them, performed analyses, provided feedback on the manuscript and approved the final version of the manuscript. MMvS provided intellectual input to focus group design and execution of the focus group discussions, provided feedback on the analyses, provided feedback on the manuscript and approved the final version of the manuscript. JSM provided intellectual input, assisted with the focus group discussions, provided feedback on the manuscript and approved the final version of the manuscript.
MS provided intellectual input to focus group design and execution of the focus group discussions, provided feedback on the manuscript and approved the final version of the manuscript. SJtV provided intellectual input to the focus group design and execution of the focus group discussions, provided feedback on the analyses, provided feedback on the manuscript and approved the final version of the manuscript. JB provided intellectual input, provided feedback and approved the final version of the manuscript. All authors read and approved the final manuscript.

Acknowledgments

The authors are grateful to the participants of the focus group discussions. We especially want to thank Tim Velten (TV) for his contributions during the focus group discussions.
References


Breton, Emily R., Bernard F. Fuemmeler, and Lorien C. Abroms (2011). “Weight loss — there is an app for that! But does it adhere to evidence-informed practices?” In: Translational behavioral medicine 1.4, pages 523–529.

Casey, Monica, Patrick S. Hayes, Fergus Glynn, Gearóid ÓLaighin, David Heaney, Andrew W. Murphy, and Liam G. Glynn (2014). “Patients’ experiences of using a smartphone application to increase physical activity: the SMART MOVE qualitative study in primary care”. In: British Journal of General Practice 64.625, e500–e508.


Chapter 4. User preferences for apps that promote physical activity


Part Three

5 Computational model of influences on physical activity .......................... 97
  5.1 Introduction
  5.2 Social cognitive computational agent model for exercise behavior
  5.3 Simulations
  5.4 Verification of computational agent model
  5.5 Conclusions

6 Empirical validation of computational model 111
  6.1 Introduction
  6.2 Computational model of psychosocial influences on physical activity
  6.3 Methods
  6.4 Results
  6.5 Discussion

7 Exploring parameter tuning for analysis of a computational model .......... 125
  7.1 Introduction
  7.2 Background
  7.3 Use case: model of psychosocial influences on physical activity
  7.4 Methods
  7.5 Results
  7.6 Discussion
  7.7 Conclusion
5. Computational model of influences on physical activity

Abstract

A computational agent model of social and cognitive influences on physical activity based on Bandura’s social cognitive theory is proposed. The utility of this model is twofold. First, it is used to run simulations of many different scenarios, that cannot be manipulated easily in reality, and that can possibly lead to new hypotheses about how social and cognitive factors influence physical activity. Second, as a next step, this computational model will be deployed in a real world coaching agent. The coach will use the current model to reason about the social and cognitive influences on the user’s physical activity and derive which coaching strategy fits the user best.
This chapter is an extended version of the article published as:

5.1 Introduction

Unhealthy lifestyles are a major public health issue, as they increase the risk of disease and fatal illness. For instance, smoking and physical inactivity account for 18% and 12% of all deaths in developed societies respectively (Murray and Lopez, 1999). The increasingly sedentary nature of Western culture cultivates unhealthy lifestyles. It is important to develop novel and effective means to stimulate people to improve their lifestyle.

Physical activity is an important means to lower the risk of disease and premature death (Warburton, 2006). Physical activity can be increased significantly through self-monitoring and the use of pedometers, but whether these changes are durable over the long term is still undetermined (Bravata et al., 2007; Lubans et al., 2009). Furthermore, internet- and e-mail-based systems that promote physical activity, or other healthy lifestyles, can increase cost-effectiveness and accessibility of an intervention, but often have low adherence rates (Davies et al., 2012; Eysenbach, 2005; Wangberg et al., 2008). If no efficient methods are developed that stimulate physical activity and other healthy lifestyles on the long term, developed societies are threatened by rapidly increasing disease prevalence and related health care costs.

The current research serves as the first step towards such an innovative method to effectively increase physical activity. We present a computational model of the social and cognitive factors that influence one’s level of physical activity, according to Albert Bandura’s social cognitive theory (Bandura, 1998, 2004). The purpose of this work is the future integration of this domain model in a smartphone app coaching agent. This agent will gather physical, cognitive and social data about the user, and apply the current computational model and intelligent reasoning techniques to these data, in order to predict the most effective coaching strategy to stimulate the user to exercise more.

Up to our knowledge, this work is the first computational model of the social cognitive theory. It could lead to new testable hypotheses, that are interesting for other disciplines such as Social Sciences and Psychology. Our aim is to simulate many scenarios of different personalities and other factors influencing physical activity and to find emerging properties leading to new hypotheses for future experiments. As a case study, we use the model to investigate how impediments and facilitators influence one’s exercise behavior.

The paper is organized as follows: Section 5.2.1 describes the proposed computational model. Section 5.3 describes the simulations performed with the model. Section 5.4 addresses the analysis of the model through automated property verification. Finally, Section 5.5 concludes the paper and discusses possible refinements and future work.

5.2 Social cognitive computational agent model for exercise behavior

In this section, the computational model based on the social cognitive theory is described both conceptually and formally.

The social cognitive theory addresses both social and cognitive factors that influence health behavior (Bandura, 1998, 2004). The key cognitive factor is the concept of self-efficacy: the confidence in one’s own ability to achieve goals. It plays a fundamental role in achieving motivation and action for healthy behavior. The behavior stands for the level of physical activity that someone is engaged in. However, people may have different perceptions of their behavior. This subjective notion is called the satisfaction. It depends on whether someone’s behavior has met the related intentions, and on the impediments
and facilitators they experience. Another factor contributing to engagement in physical activity concerns the outcome expectations for the behavior, which comes in three types: the expected social outcomes, the expected personal outcomes and the expected physical outcomes.

The motivation to base the current computational agent model on Bandura’s social cognitive theory is threefold. First of all, this theory is well-established in the literature of behavior change: it has shown to explain a large part of the variance observed in physical activity (Rovniak et al., 2002) and it served as basis for many studies investigating the determinants of physical activity (Dzewaltowski et al., 1990; Petosa et al., 2003). Second, in (Bandura, 1998) and (Bandura, 2004), the theory was specifically applied to health promotion and health behavior. Considering the objective of this research, the promotion of physical activity, the theory is very suitable to this particular endeavor. Third, the social cognitive theory explains health behavior by a combination of self-regulative processes and social context. Both of these factors are particularly relevant to the final aim of this research: the former is perfectly suited to be individually supported through a smartphone app, and the latter is available through information and interactions on social media.

5.2.1 Computational model of physical activity behavior

The dynamic relationships between all concepts are depicted in graphical form in Figure 5.1 and formalized with the differential equations below. All concepts are modeled numerically, as real values in the interval [0,1].

![Figure 5.1: Graphical overview of relations between concepts.](image)

**Self-Efficacy (SE).** The self-efficacy is largely based on the behavioral satisfaction. A positive evaluation of one’s own behavior helps to build trust in the efficacy, whereas a
5.2 Social cognitive computational agent model for exercise behavior

feeling of failure undermines it. These effects are formalized by updating the self-efficacy beliefs with a fraction of the difference between the self-efficacy and the satisfaction: if the value of the satisfaction is higher than the self-efficacy, then this will lead to increased feelings of efficacy, and vice versa. In order to guarantee that the resulting value stays within the interval from 0 to 1, the adjustment is multiplied with $SE(t)$ in case of a decrease and $(1 - SE(t))$ in case of an increase.

\[
\text{if } SE(t) \geq Sat(t) : \\
SE(t + \Delta t) = SE(t) \pm \beta_{Sat,SE} \cdot (Sat(t) - SE(t)) \cdot SE(t) \cdot \Delta t
\]

\[
\text{if } SE(t) < Sat(t) : \\
SE(t + \Delta t) = SE(t) + \beta_{Sat,SE} \cdot (Sat(t) - SE(t)) \cdot (1 - SE(t)) \cdot \Delta t
\]

**Impediments (Imp).** The self-efficacy plays a role in how insurmountable one views his/her impediments. Therefore, the ‘objective’ impediments ($Imp^*$) are adjusted based on the difference between the values of the self-efficacy and the impediments.

\[
\text{if } SE(t) \geq Imp^*(t) : \\
Imp(t) = Imp^*(t) - \beta_{SE,Imp} \cdot (SE(t) - Imp^*(t)) \cdot Imp^*(t) \cdot \Delta t
\]

\[
\text{if } SE(t) < Imp^*(t) : \\
Imp(t) = Imp^*(t) - \beta_{SE,Imp} \cdot (SE(t) - Imp^*(t)) \cdot (1 - Imp^*(t)) \cdot \Delta t
\]

**Facilitators (Fac).** Similar to the impediments, the ‘objective’ facilitators ($Fac^*$) are adjusted according to the self-efficacy.

\[
\text{if } SE(t) \geq Fac^*(t) : \\
Fac(t) = Fac^*(t) + \beta_{SE,Fac} \cdot (SE(t) - Fac^*(t)) \cdot (1 - Fac^*(t)) \cdot \Delta t
\]

\[
\text{if } SE(t) < Fac^*(t) : \\
Fac(t) = Fac^*(t) + \beta_{SE,Fac} \cdot (SE(t) - Fac^*(t)) \cdot Fac^*(t) \cdot \Delta t
\]

**Intentions (Int).** The intentions are updated by the difference between the intentions and the self-efficacy and between the intentions and the outcome expectations: the higher the self-efficacy and/or the outcome expectations, the more ambitious the intentions. Also, the intentions are adjusted for the facilitators and the impediments.

\[
\text{Change}_\text{Int}(t) = \\
\beta_{SE,Int} \cdot (SE(t) - Int(t)) + \beta_{SOE,Int} \cdot (SOE(t) - Int(t)) + \beta_{Fac,Int} \cdot Fac(t) - \beta_{Imp,Int} \cdot Imp(t)
\]

\[
\text{if } \text{Change}_\text{Int}(t) \geq 0 : \\
Int(t + \Delta t) = Int(t) + \text{Change}_\text{Int}(t) \cdot (1 - Int(t)) \cdot \Delta t
\]

\[
\text{if } \text{Change}_\text{Int}(t) < 0 : \\
Int(t + \Delta t) = Int(t) + \text{Change}_\text{Int}(t) \cdot Int(t) \cdot \Delta t
\]
Behavior (Beh). The behavior indicates one’s activity level: a value of 0 denotes that someone is not physically active at all and a value of 1 denotes that someone is maximally active. The value is updated by comparing its previous value to the self-efficacy, the outcome expectations and the intentions, and by adjusting it for the facilitators and the impediments that a person is facing.

\[
\text{Change}_{\text{Beh}}(t) = \\
\beta_{SE,\text{Beh}} \cdot (SE(t) - \text{Beh}(t)) + \beta_{OE,\text{Beh}} \cdot (OE(t) - \text{Beh}(t)) + \\
\beta_{\text{Int},\text{Beh}} \cdot (\text{Int}(t) - \text{Beh}(t)) + \beta_{\text{Fac},\text{Beh}} \cdot \text{Fac}(t) - \beta_{\text{Imp},\text{Beh}} \cdot \text{Imp}(t)
\]

if \(\text{Change}_{\text{Beh}}(t) \geq 0\):
\[
\text{Beh}(t + \Delta t) = \text{Beh}(t) + \text{Change}_{\text{Beh}}(t) \cdot (1 - \text{Beh}(t)) \cdot \Delta t
\]

if \(\text{Change}_{\text{Beh}}(t) < 0\):
\[
\text{Beh}(t + \Delta t) = \text{Beh}(t) + \text{Change}_{\text{Beh}}(t) \cdot \text{Beh}(t) \cdot \Delta t
\]

Satisfaction (Sat). The satisfaction, i.e. the evaluation of one’s behavior, is implemented by updating its value with the difference between the intentions and the behavior, and by accounting for the presence of facilitators and impediments.

\[
\text{Change}_{\text{Sat}}(t) = \beta_{\text{Int} & \text{Beh}, \text{Sat}} \cdot (\text{Beh}(t) - \text{Int}(t)) + \beta_{\text{Imp}, \text{Sat}} \cdot \text{Imp}(t) - \beta_{\text{Fac}, \text{Sat}} \cdot \text{Fac}(t)
\]

if \(\text{Change}_{\text{Sat}}(t) \geq 0\):
\[
\text{Sat}(t + \Delta t) = \text{Sat}(t) + \text{Change}_{\text{Sat}}(t) \cdot (1 - \text{Sat}(t)) \cdot \Delta t
\]

if \(\text{Change}_{\text{Sat}}(t) < 0\):
\[
\text{Sat}(t + \Delta t) = \text{Sat}(t) + \text{Change}_{\text{Sat}}(t) \cdot \text{Sat}(t) \cdot \Delta t
\]

Social Norm (SN). The social norm is implemented as the weighted average of the behavior of all relevant friends, where closer and more influential friends contribute more to the social norm than more distant or less influential friends.

\[
\text{SN}(t) = \frac{\sum_{i=1}^{n} \text{Beh}_i(t) \cdot \omega_i}{\sum_{i=1}^{n} \omega_i}
\]

Personal Norm (PN). The personal norm is partly based on the social norm and partly on a personality trait, which is called the static personal norm.

\[
\text{PN}(t) = \alpha_{\text{SN,PN}} \cdot \text{SN}(t) + (1 - \alpha_{\text{SN,PN}}) \cdot \text{Static}_\text{PN}
\]

Expected Social Outcomes (SOE). By comparing the behavioral satisfaction with the social norm, the social outcome expectations are calculated. The second half of the formula allows for outcome expectations that are not dependent on the satisfaction. The parameter \(\alpha_{\text{Sat,SOE}}\) specifies to what extent someone is influenced by the social norm.

\[
\text{SOE}(t) = \alpha_{\text{Sat,SOE}} \cdot \min(\text{Sat}(t)/\text{SN}(t)/2, 1) + (1 - \alpha_{\text{Sat,SOE}}) \cdot \text{Static}_\text{SOE}
\]
Expected Personal Outcomes (POE). The personal outcome expectations are calculated similarly to the expected social outcomes.

\[ POE(t) = \alpha_{Sat,POE} \cdot \min(Sat(t)/PN(t)/2,1) + (1 - \alpha_{Sat,POE}) \cdot Static_{POE} \]

Expected Physical Outcomes (PhOE). The physical outcome expectations are formalized by combining two parts as well: one part is determined by the satisfaction, and the other part allows for expectations based on new experiences.

\[ PhOE(t) = \alpha_{Sat,PhOE} \cdot Sat(t) + (1 - \alpha_{Sat,PhOE}) \cdot Static_{PhOE} \]

Outcome Expectations (OE). The three types of outcome expectations are aggregated into one concept: the outcome expectations. First, the three sets of expected outcomes are combined in a weighted average. Subsequently, the outcome expectations are adjusted based on the feelings of self-efficacy.

\[ OE^*(t) = (\omega_{SOE} \cdot SOE(t) + \omega_{POE} \cdot POE(t) + \omega_{PhOE} \cdot PhOE(t))/(\omega_{SOE} + \omega_{POE} + \omega_{PhOE}) \]

\[ i f OE^*(t) \geq SE(t) : \]
\[ OE(t + \Delta t) = OE^*(t) + \beta_{SE,OE} \cdot (SE(t) - OE^*(t)) \cdot OE^*(t) \cdot \Delta t \]

\[ i f OE^*(t) < SE(t) : \]
\[ OE(t + \Delta t) = OE^*(t) + \beta_{SE,OE} \cdot (SE(t) - OE^*(t)) \cdot (1 - OE^*(t)) \cdot \Delta t \]

The values of all parameters (\( \alpha \)'s, \( \beta \)'s and \( \omega \)'s) can be adjusted by the modeler and appropriate settings were chosen by the authors.

5.2.2 Expert opinion on the computational model
The first step towards validation of our computational model was to ask an expert in the field of physical activity research her opinion on our model. She is familiar with many psychosocial theories about physical activity, which she applies in her research to validate coaching strategies to increase the level of physical activity in children and young adults. She proved to be able to reason about the concepts and the dynamic relationships at the conceptual level, and provided us with insightful feedback. After careful analysis of our chosen concepts, relations and formulas, the expert agreed on all of our relations and the way they were modeled. She evaluated our model as very plausible. A possible refinement based on her advice is discussed in Section 5.5.

5.3 Simulations
The computational model, which was implemented in Matlab, contains 21 parameters. These are the result of representing Bandura’s theory in a computational model without losing details of the original model. A total of 160 simulation scenarios were carefully set up by the authors to investigate many hypotheses about Bandura’s model.
Four characteristic person types for physical activity were determined beforehand, with the advice of an expert in the field of physical activity. The initial values of the relevant concepts are either 0.5 (neutral person), 0.9 (active person) or 0.1 (inactive person), with some exceptions. The person recovering from an injury is similar to an active person, but has low values for behavior and satisfaction. See Table 5.1 for specific parameter settings. All simulations were analyzed manually and interpreted by the authors (in this section), and also verified automatically with the help of a computer tool (in Section 5.4). All data is available upon request.

Table 5.1: Simulation settings for four person types.

<table>
<thead>
<tr>
<th></th>
<th>Neutral</th>
<th>Active</th>
<th>Inactive</th>
<th>Injury</th>
</tr>
</thead>
<tbody>
<tr>
<td>SE_Start</td>
<td>0.5</td>
<td>0.9</td>
<td>0.1</td>
<td>0.9</td>
</tr>
<tr>
<td>Beh_Start</td>
<td>0.5</td>
<td>0.9</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Sat_Start</td>
<td>0.5</td>
<td>0.9</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Int_Start</td>
<td>0.5</td>
<td>0.9</td>
<td>0.1</td>
<td>0.9</td>
</tr>
<tr>
<td>OE_Start</td>
<td>0.5</td>
<td>0.9</td>
<td>0.1</td>
<td>0.9</td>
</tr>
<tr>
<td>Static_PN</td>
<td>0.5</td>
<td>0.9</td>
<td>0.5</td>
<td>0.9</td>
</tr>
<tr>
<td>Static_OE</td>
<td>0.5</td>
<td>0.9</td>
<td>0.1</td>
<td>0.9</td>
</tr>
</tbody>
</table>

5.3.1 Scenario 1–4: Default runs

The first set of simulations investigates whether the levels of physical activity change during the course of the simulation if there are no facilitators or impediments. It was shown that the behavior of the neutral person and the inactive person does neither increase nor decrease. The behavior of the active person decreases slightly, because the expected outcomes decrease below 0.9. The person with the injury starts with unstable settings: the goals and self-efficacy do not match the behavior. Therefore, the behavior first increases to approximately 0.45 and then stabilizes around 0.35, and the goals and self-efficacy decrease to 0.2 and 0.35, respectively. See Figure 5.2.

5.3.2 Scenario 5–28: Impediments only

Impediments were varied in three different strengths over all four person types with either a high or low value for $\beta_{\text{Imp,Beh}}$, resulting in a set of $4 \times 3 \times 2 = 24$ simulations. The simulations show that for an active person and during the presence of impediments: the higher the impediments, the lower the behavior. This effect is more apparent for a high $\beta_{\text{Imp,Beh}}$. On the contrary, the presence of impediments causes a boost in the behavior of an inactive person. (See Figure 5.3.) The effect is greater for a low $\beta_{\text{Imp,Beh}}$. This suggests that encountering (and overcoming) obstacles might in some cases lead to an increase in confidence and as a result an increase in physical activity.
5.3 Simulations

Figure 5.2: Scenario 4: Injured person, no facilitators, no impediments.

Figure 5.3: Scenario 18: Inactive person, average impediments, low $\beta_{\text{Imp,Beh}}$. 
5.3.3 Scenario 29–52: Facilitators only

Facilitators were varied in three different strengths over all four person types with either a high or low value for $\beta_{\text{Fac,Beh}}$, resulting in a set of $4 \times 3 \times 2 = 24$ simulations. The key finding is that when facilitators are present, the physical activity behavior of all person types always increases. This is different than the presence of impediments, which do not always lead to an increase in behavior. When facilitators disappear, the self-efficacy of all person types increases, and behavior always decreases, although never below the value it increased to during facilitators. See Figure 5.4 for an example.

When looking at the whole duration of the simulation, the final behavior value, compared to the initial value, always increases for the active persons and for the neutral persons with a high value for $\beta_{\text{Fac,Beh}}$, but it always ends lower for the inactive and injured persons. This could be accounted for by their high initial levels of self-efficacy, followed by decreases in self-efficacy during facilitators. This result seems to indicate that on the long term, active and injured persons should not experience too many facilitators, even though their behavior increases on the short term.

Another interesting finding is that during facilitators, the self-efficacy decreases for all persons, except for inactive or injured persons with a high value for $\beta_{\text{Fac,Beh}}$. It seems that one normally lowers his/her self-efficacy when experiencing facilitators, because one is not fully responsible for his/her own behavior. The inactive and injured person with a high value for $\beta_{\text{Fac,Beh}}$ seem to still experience their behavior as mastery experiences during facilitators, which could point towards giving these persons the ‘right challenge’ to create mastery experiences.

Figure 5.4: Scenario 34: Neutral person, high facilitators, high $\beta_{\text{Fac,Beh}}$. 
5.3.4 **Scenario 53–160: Impediments and facilitators**

For the active, inactive and injured person type, impediments and facilitators were varied in three orders (simultaneously or sequentially), in three different strength combinations and with high or low values for $\beta_{\text{Imp,Beh}}$ and $\beta_{\text{Fac,Beh}}$, resulting in $3 \times 3 \times 3 \times 2 \times 2 = 108$ simulations. An interesting finding is that no matter if impediments and facilitators are of equal strength or one is stronger than the other, the inactive persons always show that their self-efficacy and physical activity behavior increase over the whole duration of the simulation, while both decrease for the active persons. See Figure 5.5 and Figure 5.6.

![Figure 5.6: Scenario 56: Active person, impediments and facilitators, high $\beta_{\text{Imp,Beh}}$, high $\beta_{\text{Fac,Beh}}$.](image)

This means that when impediments and facilitators are of equal strength, they do not compensate for each other’s effect. For example, for active persons, the self-efficacy decreases twice: first during impediments and later after facilitators are gone. For inactive persons, the self-efficacy increases twice: first during impediments and later after facilitators have disappeared.

When facilitators and impediments arise simultaneously, their effects cancel each other out if they have equal strengths and equal values for $\beta_{\text{Imp,Beh}}$ and $\beta_{\text{Fac,Beh}}$. If one of both factors is stronger than the other, or if the $\beta$s have unequal values, then the effects resemble the presence of either facilitators or impediments.

### 5.4 Verification of computational agent model

The presented computational model was analyzed by specification and verification of properties expressing dynamic patterns that are expected to emerge. The purpose of such verification is to automatically check whether the model behaves correctly, by running a
large number of simulations and verifying such properties against the simulation traces. This process would be very time consuming if done by hand, and it enables verifying complex properties that require deep logical thinking automatically.

Several dynamic properties have been identified to check basic model issues (for example, do all concepts stay within the range of \([0,1]\)) or were based on hypotheses from the researchers or from literature. The properties were formalized in the Temporal Trace Language (TTL) and checked automatically (Bosse et al., 2009). The language TTL is built on atoms referring to states of the world, time points and traces. For example: state(\(\gamma, t\)) \(\models p\) denotes that \(p\) holds in trace \(\gamma\) (a trajectory of states over time). Dynamic properties can be formulated using quantifiers over time and traces and first-order logical connectives, such as \(\neg, \land, \lor, \Rightarrow, \forall, \exists\). Below, we present an example of such a property, both in semi-formal and in formal notation. It checks whether the physical activity level is increased at the end of the simulation.

**P1: Increase of physical activity during simulation**

There exists a timepoint \(t1\) at the beginning of the trace, at which behavior has value \(x\), and a timepoint \(t2\) at the end of the trace, at which behavior has value \(y\) and \(y > x\).

\[
\exists m: \text{TRACE}, \exists t1, t2: \text{TIME}, \exists x, y: \text{REAL} \\
\text{state}(m, t1) = \text{Beh}(x) \& \text{state}(m, t2) = \text{Beh}(y) \& y > x
\]

Property P1 can be used to verify if the physical activity is increased at the end of the simulation. This is interesting to find out for many simulation scenarios at the same time. One can find out which combination of the interaction between the social and cognitive...
processes with the impediments and/or facilitators leads to an increase or decrease in physical activity. This property can be verified for all concepts, simply by changing $\text{Beh}(x)$ into another concept, such as $\text{SE}(x)$. This property was verified for many simulation traces. For example, the property did not succeed for all active persons that experience impediments and injured persons experiencing facilitators, meaning their behavior decremented or remained stable during the simulation. The property was successful for all inactive persons experiencing facilitators, meaning that for all of them their physical activity increased during the simulation. For some of the simulation traces of inactive persons experiencing impediments, the property showed to be successful as well.

5.5 Conclusions

The aim of this research was to develop a computational model of Bandura’s social cognitive theory. The strength of this model is twofold. First, it can be used to simulate numerous scenarios to test hypotheses known from literature and to find emerging properties leading to new hypotheses for future experiments. Second, the model is designed as a basis for a coaching agent that will gather cognitive, social and physical data and use intelligent reasoning capabilities to apply the most efficient coaching strategy to the user. This coach will predict the user’s exercise behavior based on the measured data and it will reason with the current computational model to apply different coaching strategies on the user in order to stimulate him/her to exercise more.

In order to draw strong conclusions or make predictions based on the current computational model, it should be validated. Although an expert in the field of physical activity found our computational model plausible, and the model was also verified successfully through checking many dynamic properties over all simulations, we plan to gather empirical data of the model’s concepts over time to tune the parameters and to test whether the patterns generated by the model are supported by real world data.

A possible improvement of the model identified by the expert was preventing the self-efficacy to drop below the value for the actual behavior. It can be questioned whether it is possible in real life to be physically active, but have low feelings of self-efficacy. A property was created to test this, which revealed that the model does enable it. The empirical data gathered for the validation of the model could be used to find indications whether this artefact is desirable, or whether it should be avoided.

Acknowledgments

The authors would like to thank dr. Saskia te Velde for her insightful feedback on our computational model.
Chapter 5. Computational model of influences on physical activity

References

Abstract

The adoption and maintenance of a healthy lifestyle is a fundamental pillar in the quest towards a healthy society. Modern (mobile) technology allows for increasingly intelligent systems that can help to optimize people’s health outcomes. One of the possible directions in such mHealth systems is the use of intelligent reasoning engines based on dynamic computational models of behavior change. In this work, we investigate the accuracy of such a model to simulate changes in physical activity levels over a period of two to twelve weeks. The predictions of the model are compared to empirical physical activity data of 108 participants. The results reveal that the model’s predictions show a moderate to strong correlation with the actual data, and it performs substantially better than a simple alternative model. Even though the implications of these findings depend strongly on the application at hand, we show that it is possible to use a computational model to predict changes in behavior. This is an important finding for developers of mHealth systems, as it confirms the relevance of model-based reasoning in such health interventions.
This chapter appeared as:


### 6.1 Introduction

It is well known that engaging in sufficient physical activity has many beneficial effects on physical and mental health (Conn et al., 2011; Eime et al., 2013). On the contrary, low levels of physical activity have been associated with increased risks of cardiovascular diseases, cancer, diabetes, and mental illness (Lee et al., 2012). Despite these prominent advantages, a large proportion of the Western population does not meet the guidelines of being moderately to vigorously active for at least 30 minutes on at least five days a week (Haskell et al., 2007).

It is believed that mobile technology provides an opportunity to support people with increasing their level of physical activity (Knight et al., 2015; Payne et al., 2015; Sieverdes et al., 2013), and there is also some initial evidence that this is effective (Stephens and Allen, 2013). However, just monitoring physical activity is not sufficient to achieve a durable improvement (Gierisch et al., 2015). There is a need for development of evidence-informed mobile apps that apply advanced technological features in order to yield long-term effects (Knight et al., 2015).

In (Klein et al., 2015), we described how we developed an intelligent system to stimulate physical activity for young adults. Part of the intelligence lies in the fact that it uses a computational model of behavior change to predict the effect of different intervention strategies on the activity level of the users. The model consists of temporal-dynamic relations between determinants of behavior change. The model predictions are used for deciding on the support messages for specific users in each phase of the intervention. We believe that this can result in a highly tailored and personalized intervention.

The model is based on a number of different theories on behavior change (Michie and Johnston, 2012). Most of these theories have been validated independently, and there also have been validations of integrated frameworks (Cane et al., 2012). However, these validations usually look at correlations between the different constructs in the theories. Hence, we do not yet know whether the dynamic computational representation of our integrated model provides a valid description of the process of behavior change. Therefore, we would like to know to what extent the model is a valid way to predict the most effective coaching strategies. This paper describes a first step towards such a validation. We use empirical data collected in the effectiveness study to compare the prediction of the model based on the initial questionnaire with the actual change in physical activity that has been measured in the study. In addition, we compare the actually measured behavior with predictions of an alternative simple model. The results provide an initial answer about the validity of the model.

The remainder of this paper is organized as follows. In Section 6.2, the details of the computational model under consideration are presented. Section 6.3 describes the methods used to provide a validation of the computational model. The results are presented in Section 6.4, and reflected upon in Section 6.5.

### 6.2 Computational model of psychosocial influences on physical activity based on social cognitive theory

The computational model investigated in this paper was designed in context of the development of an intelligent behavior change support system (Klein et al., 2015). The reasoning engine of this system uses the model to predict what available coaching strategies are the
most promising to improve the user’s behavior. (For more information, see (Klein et al., 2015).)

The model captures the dynamics between psychosocial influences on physical activity behavior. It describes the relations between several psychological determinants, such as self-efficacy, intentions and social norms, and their influence on physical activity behavior. The model under investigation in this paper is an adaptation of the computational model presented in (Mollee and van der Wal, 2013). Therefore, the concepts and the relations of the model described below are explained in more detail in the original publication. The revision was motivated by a decrease in conceptual detail and computational complexity of the model, and by suggestions of experts in behavior change.

The computational model is largely based on the social cognitive theory by Albert Bandura 1998. This is a well-established theory of behavior change, with high applicability in the domain of health behavior (Bandura, 2004). The theory has proven to account for a large proportion of the variance in physical activity (Rovniak et al., 2002), and is therefore very suitable as a basis to describe the dynamics underlying physical activity behavior.

All concepts are modeled numerically, as real values in the interval [0,1], and the relations are formalized as differential equations. The relations express the influence of the source concept on the target concept by increasing or decreasing the value of the target concept in the direction of the source target, moderated by a parameter named $\beta_{\text{source}\_\text{target}}$ (or occasionally $\beta_{\text{source1+source2}\_\text{target}}$). The increase or decrease of the target concept is also relative to its current value: e.g., $SE(t)$ in case of a decrease and $(1 - SE(t))$ in case of an increase. This consideration of the current value also ensures that the concept values stay in the interval [0,1]. The constant $\Delta t$ indicates the step size of the model, and is set at 0.1 to ensure smooth results. Figure 6.1 shows a graphical representation of the dynamic relations between all concepts in the model.

Below, the meaning of the concepts in the model are explained and the formal relations are specified.

**Self-Efficacy:** The self-efficacy ($SE$) is a key element of the process described by the social cognitive theory. It represents the confidence in one’s own ability to achieve certain goals, which plays a fundamental role in the acquisition and maintenance of some desired behavior. The self-efficacy increases with high satisfaction of the current behavior, and it decreases if one is dissatisfied with his/her behavior.

$$
\text{if}(SE(t) \geq Sat(t)):
\quad SE(t + 1) = SE(t) + \beta_{Sat,SE} \cdot (Sat(t) - SE(t)) \cdot \Delta t \cdot SE(t)
$$

$$
\text{if}(SE(t) < Sat(t)):
\quad SE(t + 1) = SE(t) + \beta_{Sat,SE} \cdot (Sat(t) - SE(t)) \cdot \Delta t \cdot (1 - SE(t))
$$

**Impediments:** Impediments ($Imp$) are the (personal, situational or systemic) factors that form an obstacle to the desired behavior. The self-efficacy plays a role in how insurmountable one views those obstacles. Therefore, the personal impediments ($Input\_Imp$) are adjusted based on the level of the self-efficacy.

$$
\text{if}(SE(t) \geq Input\_Imp(t)):
\quad Imp(t) = Input\_Imp(t) - (\beta_{SE,Imp} \cdot (SE(t) - Input\_Imp(t))) \cdot \Delta t \cdot Input\_Imp(t)
$$
Figure 6.1: Graphical representation of the model.

\[
\text{if}(SE(t) < Input\_Imp(t)) : \\
\quad \text{Imp}(t) = Input\_Imp(t) - (\beta_{SE,Imp} \cdot (SE(t) - Input\_Imp(t))) \cdot \Delta t \cdot (1 - Input\_Imp(t))
\]

In the system, these personal ‘input impediments’ are assessed through a questionnaire, so they reflect the user’s overall experience of barriers on a scale of 0 (no impediments) to 1 (very strong impediments).

**Social Norm:** The social norm (SN) represents the behavioral standards that one’s social connections impose on him or her. It is derived directly from information about the user’s social network, and it assumed to be stable for the duration covered by the simulations.

**Long-Term Goals:** The long-term goals (LTG) can be interpreted as the overall motivation to achieve change in the behavior. The levels of self-efficacy can increase or decrease the long-term goals.

\[
\text{Change\_LTG}(t) = (\beta_{SE,LTG} \cdot (SE(t) - LTG(t)))
\]

\[
\text{if}(\text{Change\_LTG}(t) \geq 0) :\\
\quad LTG(t + 1) = LTG(t) + \text{Change\_LTG}(t) \cdot \Delta t \cdot (1 - LTG(t))
\]

\[
\text{if}(\text{Change\_LTG}(t) < 0) :\\
\quad LTG(t + 1) = LTG(t) + \text{Change\_LTG}(t) \cdot \Delta t \cdot LTG(t)
\]
**Intentions:** The intentions (Int) denote the user’s aims for the desired behavior. They provide focus and a measure for evaluation. The intentions are influenced by the self-efficacy, the social norm and the outcome expectations, and adjusted by the perceived impediments.

\[
\text{Change} \_ \text{Int}(t) = \\
(\beta_{SE, \text{Int}} \cdot (SE(t) - \text{Int}(t)) + \beta_{LTG, \text{Int}} \cdot (LTG(t) - \text{Int}(t)) + \beta_{OE, \text{Int}} \cdot (OE(t) - \text{Int}(t)) \\
+ \beta_{SN+Sat, \text{Int}} \cdot (Sat(t) - SN(t)) - \beta_{Imp, \text{Int}} \cdot \text{Imp}(t))
\]

if \((\text{Change} \_ \text{Int}(t) \geq 0)) : 
\[ \text{Int}(t + 1) = \text{Int}(t) + (\text{Change} \_ \text{Int}(t)) \cdot \Delta t \cdot (1 - \text{Int}(t)) \]

if \((\text{Change} \_ \text{Int}(t) < 0)) :
\[ \text{Int}(t + 1) = \text{Int}(t) + (\text{Change} \_ \text{Int}(t)) \cdot \Delta t \cdot \text{Int}(t) \]

**Behavior:** The behavior (Beh) describes the level of physical activity that someone is engaged in: its value is 0 if someone is not physically active at all, and 1 if someone is maximally active. It is mainly influenced by the self-efficacy, outcome expectations, intentions and the impediments.

\[
\text{Change} \_ \text{Beh}(t) = \\
(\beta_{SE, \text{Beh}} \cdot (SE(t) - \text{Beh}(t)) + \beta_{Int, \text{Beh}} \cdot (\text{Int}(t) - \text{Beh}(t)) \\
+ \beta_{OE, \text{Beh}} \cdot (OE(t) - \text{Beh}(t)) - \beta_{Imp, \text{Beh}} \cdot \text{Imp}(t))
\]

if \((\text{Change} \_ \text{Beh}(t) \geq 0)) :
\[ \text{Beh}(t + 1) = \text{Beh}(t) + (\text{Change} \_ \text{Beh}(t)) \cdot \Delta t \cdot (1 - \text{Beh}(t)) \]

if \((\text{Change} \_ \text{Beh}(t) < 0)) :
\[ \text{Beh}(t + 1) = \text{Beh}(t) + (\text{Change} \_ \text{Beh}(t)) \cdot \Delta t \cdot \text{Beh}(t) \]

**Satisfaction:** The satisfaction (Sat) denotes one’s perception of his/her own behavior, i.e. an evaluation of the behavior. It is based on the difference between one’s intentions and current behavior, and adjusted with the perceived impediments.

\[
\text{Change} \_ \text{Sat}(t) = (\beta_{Int+Beh, \text{Sat}} \cdot (\text{Beh}(t) - \text{Int}(t)) + \beta_{Imp, \text{Sat}} \cdot \text{Imp}(t))
\]

if \((\text{Change} \_ \text{Sat}(t) \geq 0)) :
\[ \text{Sat}(t + 1) = \text{Sat}(t) + (\text{Change} \_ \text{Sat}(t)) \cdot \Delta t \cdot (1 - \text{Sat}(t)) \]

if \((\text{Change} \_ \text{Sat}(t) < 0)) :
\[ \text{Sat}(t + 1) = \text{Sat}(t) + (\text{Change} \_ \text{Sat}(t)) \cdot \Delta \cdot \text{Sat}(t) \]

**Outcome Expectations:** The outcome expectations (OE) represent the anticipated results of performing the behavior, on a physical, personal and social level. They are influenced by one’s satisfaction with the current behavior and the self-efficacy.
$$Change\_OE(t) = (\beta_{Sat,OE} \cdot (Sat(t) - OE(t)) + \beta_{SE,OE} \cdot (SE(t) - OE(t)))$$

if($Change\_OE(t) \geq 0$):

$$OE(t + 1) = OE(t) + (Change\_OE(t)) \cdot \Delta t \cdot (1 - OE(t))$$

if($Change\_OE(t) < 0$):

$$OE(t + 1) = OE(t) + (Change\_OE(t)) \cdot \Delta t \cdot OE(t)$$

The values of all parameters ($\beta$) can be adjusted by the modeler. In the current implementation of the model, the parameters were chosen based on correlations between the concepts found in literature (Plotnikoff, Costigan, et al., 2013; Plotnikoff, Lippke, et al., 2008; Rovniak et al., 2002), in order to keep the ratio between the parameters in accordance with empirical findings. This parameter set is shown in Table 6.1. Additionally, one day was chosen to correspond with 10 time steps.

Table 6.1: Parameter settings.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$\beta_{Sat,SE}$</th>
<th>$\beta_{Imp,Int}$</th>
<th>$\beta_{SE,Int}$</th>
<th>$\beta_{SE,Imp}$</th>
<th>$\beta_{Imp,Sat}$</th>
<th>$\beta_{Int+Beh,Sat}$</th>
<th>$\beta_{SE,Beh}$</th>
<th>$\beta_{Int,Beh}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>0.50</td>
<td>0.08</td>
<td>1.00</td>
<td>0.43</td>
<td>0.25</td>
<td>0.50</td>
<td>0.17</td>
<td>0.60</td>
</tr>
<tr>
<td>Parameter</td>
<td>$\beta_{Imp,Beh}$</td>
<td>$\beta_{SE,LTG}$</td>
<td>$\beta_{LTG,Int}$</td>
<td>$\beta_{Sat,OE}$</td>
<td>$\beta_{SE,OE}$</td>
<td>$\beta_{OE,Int}$</td>
<td>$\beta_{OE,Beh}$</td>
<td>$\beta_{SN+Sat,Int}$</td>
</tr>
<tr>
<td>Value</td>
<td>0.25</td>
<td>0.05</td>
<td>0.20</td>
<td>0.10</td>
<td>0.05</td>
<td>0.02</td>
<td>0.01</td>
<td>0.02</td>
</tr>
</tbody>
</table>

6.3 Methods

In order to assess the validity of the computational model described above, empirical data is collected, preprocessed and analyzed. These steps are described in this section.

6.3.1 Data collection

The data was collected in context of a user study, in which three (versions of) physical activity promotion apps were tested. Each of the participants ($N = 108$) used one of the apps for at least 12 weeks, in the period between March and October 2016.

At the start of the experiment, the participants were asked to fill in an extensive intake questionnaire. This questionnaire included questions about their demographics (e.g., gender, age), about their daily life patterns (e.g., occupation, important locations, travel options), and about psychological concepts underlying behavior (e.g., intentions, self-efficacy). Each of the eight psychological constructs is assessed by a number of items on a four- or five-point Likert scale, or by one item on a scale of 1 to 10. The items were based on extensive, validated questionnaires (e.g., Frank et al., 2009; Rovniak et al., 2002; Sallis et al., 1988).

All participants received a Fitbit One activity tracker that monitored their physical activity and synchronized their data wirelessly to their assigned app. The tracker registers steps, floors climbed, distance, calories burned and active minutes. As mentioned before, the
participants measured their physical activity via the Fitbit One for a minimum of 12 weeks (that is, apart from possible dropouts). The data from the first week was used to assess the initial physical activity, whereas the subsequent eleven weeks were used as ground truth to compare with the model’s predictions.

### 6.3.2 Data preprocessing

In order to obtain the initial values for the concepts of the computational model, the responses to the questionnaire items for assessing the psychological constructs were aggregated per concept. As the model assumes numerical values between 0 and 1, all responses were rescaled and averaged per concept in order to fit in that same range.

The initial value of the behavior concept was based on the Fitbit step data, rather than self-reported questionnaire answers. This value was calculated in three steps. First, the number of steps in the first seven days of participation was averaged, while discarding any days with no recorded steps (e.g., because the participant forgot to wear the activity tracker). Then, the average number of steps was capped off at 15,000 steps per day, as that represents amply complying with the guideline of 10,000 steps per day. Finally, the initial value of the behavior concept was obtained by normalizing the average number of steps. This way, a behavior value of 1.0 corresponds with 15,000 (or more) daily steps, which is regarded as “optimal” behavior. The calculation of this initial behavior value is shown in Equation 6.1.

\[
\text{Beh}(t_0) = \frac{\min(15,000, \text{NumSteps})}{15,000}
\]  

(6.1)

where NumSteps is the average number of steps for the days in the first week, discarding any days with no recorded steps.

### 6.3.3 Analyses

In order to investigate the validity of the computational model, the assessments of the psychological constructs and the normalized average number of steps in the first week were used to initialize the computational model (as described in Section 6.3.2. Starting from these values, the model was run to predict the physical activity behavior in the second week, third week, etc., up to the twelfth week of the experiment. As for the calculation of the initial behavior value, the actual behavior values for these subsequent weeks were determined as well. Then, the actual change in behavior and the predicted change in behavior were calculated.

The accuracy of the model’s predictions was assessed by comparing the predicted differences to the actual differences in the behavior values. This was done by means of calculating Spearman rank-order correlations.

In order to exclude the possibility that the model’s performance relies on certain underlying pattern in the data (e.g., high values will probably decrease, and vice versa), it was evaluated by comparing it to the results of a simple alternative model. In this ‘random model’, for each user a random value in [0,1] is drawn as predicted behavior value. This random model will also show that high values generally decrease and vice versa, as the probability of drawing a value below (for example) 0.8 is higher than drawing a value above 0.8. Spearman correlations were calculated for the random predictions as well. To avoid
flukes, the random model was applied and evaluated 100 times, and the resulting correlation coefficients and p-values were averaged.

Although 108 participants filled out the intake questionnaire and wore the Fitbit One for an intended period of 12 weeks, not all participants had step data for each subsequent week (e.g., dropouts). Those were not considered in the analyses.

### 6.4 Results

The 108 people that participated in the user study were between 18 and 30 years old at the time of the data collection. Of those, 22 were male and 86 were female. However, the number of participants with usable data varied each week. Table 6.2 shows the number of users whose data was included in the analyses.

Table 6.2: Number of users included and excluded in the analyses for each predicted week.

<table>
<thead>
<tr>
<th>Week number</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. included users</td>
<td>92</td>
<td>88</td>
<td>87</td>
<td>85</td>
<td>82</td>
<td>78</td>
<td>79</td>
<td>79</td>
<td>74</td>
<td>72</td>
<td>66</td>
</tr>
<tr>
<td>No. discarded users</td>
<td>16</td>
<td>20</td>
<td>21</td>
<td>23</td>
<td>26</td>
<td>30</td>
<td>29</td>
<td>29</td>
<td>34</td>
<td>36</td>
<td>42</td>
</tr>
</tbody>
</table>

The results of the Spearman rank-order correlation tests are summarized in Table 6.3. It shows the correlation coefficient ($r_s$) and corresponding p-value for the predictions of both the computational model and the random model. The results of the random model are based on 100 draws of a random prediction for each user.

Table 6.3: Results of the Spearman rank correlation for week 2 up to week 12.

<table>
<thead>
<tr>
<th>Week</th>
<th>Computational Model</th>
<th>Model (avg)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Spearman’s $r_s$</td>
<td>p-value</td>
</tr>
<tr>
<td>2</td>
<td>.4134</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>3</td>
<td>.4019</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>4</td>
<td>.3001</td>
<td>.0047</td>
</tr>
<tr>
<td>5</td>
<td>.3957</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>6</td>
<td>.3064</td>
<td>.0051</td>
</tr>
<tr>
<td>7</td>
<td>.5522</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>8</td>
<td>.3173</td>
<td>.0044</td>
</tr>
<tr>
<td>9</td>
<td>.3201</td>
<td>.0040</td>
</tr>
<tr>
<td>10</td>
<td>.2841</td>
<td>.0142</td>
</tr>
<tr>
<td>11</td>
<td>.2319</td>
<td>.0499</td>
</tr>
<tr>
<td>12</td>
<td>.3510</td>
<td>.0039</td>
</tr>
</tbody>
</table>

To illustrate, Figure 6.2 shows a scatter plot of the changes in the behavior values of 85 included users as predicted by the model (on the vertical axis) and the corresponding
changes according to the empirical data (on the horizontal axis). The scatter plot for the random model shows the predictions of one of the 100 repeated runs.

![Figure 6.2: Correlation plots with the predictions of the computational model (left) and the random model (right) on the vertical axis and the empirical data on the horizontal axis for week 5.](image)

6.5 **Discussion**

The results presented in Section 6.4 reveal that the computational model performs quite well in predicting the change in physical activity level. The predictions show a weak to moderate positive correlation with the actual data, which is statistically significant ($p < .05$) for all predicted weeks. In contrast, the random model has both weaker and non-significant correlations with the empirical data. However, the random model performs relatively well on some of the weeks (i.e., week 2 and week 7). This indicates that some characteristics of the data can make it easier to predict the change right. Further investigation into this finding should reveal why this happens.

The work presented in this paper clearly is only a first step in the direction of validating computational models that are applied in mHealth systems. For instance, it would be interesting to see whether the computational model is also able to predict the course of the behavior on a more detailed (i.e., daily) level, rather than considering the errors for its predictions per week. On the other hand, the steady good performance for each week suggests that the model captures the dynamics of the behavior over time quite well. Also, the current work is limited to 12 weeks, which might not be transferable to longer periods. The model’s performance on predicting the underlying psychological constructs would be an interesting further exploration as well. Moreover, as Table 6.2 shows, the dataset contained a substantial number of missing observations, for example because of dropouts or participants forgetting to wear the activity tracker. This could have affected the results, and therefore replicating the analyses on a more complete dataset would be another valuable endeavor.

Several directions for further analysis could reveal whether the computational model is able to perform even better than the results found in this work. For example, the current model uses global parameters (based on indications from literature), but it is plausible that better results could be obtained when the parameters are tuned to the users in the dataset,
either globally or individually. In addition, the computational model does not account for
the fact that the participants in the study were exposed to an intervention during the data
collection: they were using a physical activity promotion app (see Section 6.3.1). By taking
a potential effect of the intervention into account, the model’s predictions could arguably be
improved even further.

Validation of dynamic computational models is an important endeavor, as it allows
researchers to better understand the dynamics of the modeled behavior through simulations.
Moreover, this work presents a step in the direction of more reliable and effective mHealth
systems. After all, if the computational models underlying their reasoning engines are
proven trustworthy, this increases the dependability of the support provided by the mHealth
systems.
References


Plotnikoff, Ronald C., Sonia Lippke, Kerry S. Courneya, Nick Birkett, and Ronald J. Sigal (2008). “Physical Activity and Social Cognitive Theory: A Test in a Population Sample of Adults with Type 1 or Type 2 Diabetes”. In: Applied Psychology 57.4, pages 628–643.


Abstract

Computational models of human processes are used for many different purposes and in many different types of applications. A common challenge in using such models is to find suitable parameter values. In many cases, the ideal parameter values are those that yield the most realistic simulation results. However, there are situations in which the goodness of fit is not the main or only criterion to evaluate the appropriateness of a model, but where other aspects of the model behavior are also relevant. This is often the case when computational models are employed in real-life applications, such as mHealth systems. In this paper, we explore how parameter tuning techniques can be used to analyze the behavior of computational models systematically and to investigate the reasons behind the observed behavior. We study a computational model of psychosocial influences on physical activity behavior as an in-depth use case. In this particular case, an important measure of the feasibility of the model is the diversity in the simulation outcomes. This novel application of parameter tuning techniques for analysis and understanding of model behavior is transferable to other cases, and is therefore a valuable new approach in the toolset of computational modelers.
This chapter appeared as:

7.1 Introduction

Computational models of human processes are used for many different purposes and in many different types of applications. A common use case of computational models is the analysis and understanding of the modeled processes. Another application area of such models is human support systems, in which the model provides the understanding of the user. Examples of such systems are behavior change support systems (Oinas-Kukkonen, 2010) or systems that provide support during demanding tasks.

A challenge when developing these models is to find parameter sets that result in adequate behavior of the model. Usually, background knowledge and (psychological) literature is used to determine the values in these models. In other situations, especially when empirical data about actual human behavior is available, automated parameter estimation techniques are used to find suitable parameter values. To test the appropriateness of the parameterized model, the goodness of fit is determined by comparing the generated values with the observed values.

However, there are situations in which the goodness of fit is not the only criterion to evaluate a model. When the data is noisy, goodness of fit might lead to incorrect outcomes (Pitt and Myung, 2002), but it is also possible that other aspects of model behavior are important. For example, in this paper we use a model (that is deployed in an mHealth system) as case study in which the variety in the outcomes is an important measure of the feasibility of the model (see Section 7.3.2 for further explanation). When this aspect of the model behavior is suboptimal, the question arises how this can be explained. This is a non-trivial problem to which analytical techniques could contribute.

In this paper, we exploit parameter tuning techniques to investigate the behavior of the model and to investigate the reasons behind the observed behavior. Specifically, we do an in-depth analysis of a computational model that is expected to show diverse outcomes. We investigate the question about the cause of this lack of diversity in outcomes: is it due to the data that is used (for initial values) or due to (an inadequate choice of) parameter values? The application of the parameter tuning techniques leads to a better understanding of the model behavior and provides answers to these questions. It can be used to gain more insight in the structural properties of the model.

7.2 Background

In this section, we describe related work on optimization problems and parameter tuning, as well as the novelty of the current work.

7.2.1 Parameter tuning

Parameter tuning is a widely used optimization approach. It is often used in machine learning applications, to find optimal parameters for the learning process (Chapelle et al., 2002), but also in evolutionary computing applications (Eiben and Smit, 2011). In addition, these algorithms are broadly used in dynamic modeling, in order to fit model predictions to actual data. We can find applications of parameter tuning strategies in several domains. For instance, in hydrology, parameter tuning strategies are used in hydrologic models, which are defined by parameters and states. Here, parameters are physical and generally time-invariant descriptions of surface and subsurface characteristics, while states are fluxes and storages
of water and energy that are propagated in time by the model physics (Moradkhani et al., 2005).

Parameter tuning techniques also have many applications in the simulation of human systems or agent-based models (Bonabeau, 2002). Simulations of crowd behavior (Bosse et al., 2012; Sun and Wu, 2011), organizations, and emotion contagion (Tsai et al., 2011) are examples of scenarios for parameter tuning. Individual and group behaviors rely on many aspects, which can be captured in models that will need to be fitted to real data obtained from empirical experiments.

7.2.2 Simulated annealing

The computational model investigated in this paper has many continuous variables. The presence of continuous variables in an optimization problem increases the complexity of the problem solving mechanism due to the infinite number of possibilities in the solution space, which makes it an NP-hard combinatorial problem. Often, parameter tuning tasks with continuous variables are limited by time constraints, making it impossible to identify the globally best result. Therefore, many optimizations algorithms are designed to yield good solutions in a limited time period, but do not guarantee that the solution found is the globally best solution. Examples of such algorithms include simulated annealing (SA), gradient descent, and evolution-based, swarm-based and ecology-based algorithms (Bertsimas and Tsitsiklis, 1993; Binitha and Sathya, 2012; Černý, 1985; Kirkpatrick et al., 1983).

For our analysis, we use the simulated annealing algorithm. This technique is able to find the global optimum when the variables are discrete. In case of continuous variables, SA is preferable over alternatives (like gradient descent), because of its fast convergence and easy implementation. This is the reason that NP-hard combinatorial optimization problems can be successfully addressed with SA (Bouleimen and Lecocq, 2003).

7.2.3 Novelty

In this work, we also apply parameter tuning in the domain of human behavior models. However, in contrast with the research described above, we do not use parameter tuning techniques with the aim to find a best fitting model (as is done in most other applications of parameter optimization in computational models). Instead, we use parameter tuning as a means to analyze the behavior of the model. The tuning algorithm is used to investigate to what extent a model can produce different simulation outcomes. We compare the outcomes of the simulations with the structural aspects of the model. This provides us with an additional tool that helps to increase our understanding of the behavior of the model in relation to the structural characteristics.

7.3 Use case: model of psychosocial influences on physical activity

The use case studied in this paper concerns a computational model of psychosocial influences on physical activity behavior based on the social cognitive theory (Bandura, 1998, 2004). It describes the relations between several psychological determinants, such as self-efficacy, intentions and social norms, and their influence on physical activity behavior. The model under investigation is an adaptation of the computational model presented in (Mollee and van der Wal, 2013). Therefore, the concepts and the relations of the model described in Section
7.3 Use case: model of psychosocial influences on physical activity

7.3.1 Detailed specification of the model

The computational model is largely based on the social cognitive theory by Albert Bandura 1998. This is a well-established theory of behavior change, with high applicability in the domain of health behavior (Bandura, 2004). The theory has proven to account for a large proportion of the variance in physical activity (Rovniak et al., 2002), and is therefore very suitable as a basis to describe the dynamics underlying physical activity behavior.

All concepts are modeled numerically, as real values in the interval [0,1], and the relations are formalized as differential equations. The relations express the influence of the source concept on the target concept by increasing or decreasing the value of the target concept in the direction of the source target, moderated by a parameter named $\beta_{\text{source, target}}$. The increase or decrease of the target concept is also relative to its current value (e.g., for concept SE, a decrease is relative to $SE(t)$ and an increase is relative to $(1 - SE(t))$, in order to ensure that the concept values stay in the interval [0,1]. The constant $\Delta t$ indicates the step size of the model, and is set at 0.1 to ensure smooth results. An example of the formal relation between the concepts is given below, for the concept Behavior ($Beh$). Figure 7.1 shows a graphical representation of the dynamic relations between all concepts in the model.

$$Change_{-}Beh(t) = (\beta_{SE,Beh} \cdot (SE(t) - Beh(t)) + \beta_{Int,Beh} \cdot (Int(t) - Beh(t)) + \beta_{OE,Beh} \cdot (OE(t) - Beh(t)) - \beta_{Imp,Beh} \cdot Imp(t))$$

if($Change_{-}Beh(t) \geq 0$) :
$$Beh(t + 1) = Beh(t) + (Change_{-}Beh(t)) \cdot \Delta t \cdot (1 - Beh(t))$$

if($Change_{-}Beh(t) < 0$) :
$$Beh(t + 1) = Beh(t) + (Change_{-}Beh(t)) \cdot \Delta t \cdot Beh(t)$$

The values of all parameters ($\beta$) can be adjusted by the modeler. In the current implementation, the parameters were chosen based on correlations between the concepts found in literature (Plotnikoff, Costigan, et al., 2013; Plotnikoff, Lippke, et al., 2008; Rovniak et al., 2002), in order to keep the ratio between the parameters in accordance with empirical findings. This original parameter set is shown in Table 7.1.

7.3.2 Practical application of the model

The computational model described above was created in the context of a behavior change system for encouraging physical activity among young adults (Klein et al., 2015). The system monitors the users’ behavior through an activity tracker, and combines this with
Chapter 7. Exploring parameter tuning for analysis of a computational model

Figure 7.1: Graphical representation of the model.

Table 7.1: Parameter settings.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$\beta_{\text{Sat,SE}}$</th>
<th>$\beta_{\text{Imp,Int}}$</th>
<th>$\beta_{\text{SE,Int}}$</th>
<th>$\beta_{\text{SE,Imp}}$</th>
<th>$\beta_{\text{Imp,Sat}}$</th>
<th>$\beta_{\text{Int+Bh,Sat}}$</th>
<th>$\beta_{\text{SE,Bh}}$</th>
<th>$\beta_{\text{Int,Bh}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>0.50</td>
<td>0.08</td>
<td>1.00</td>
<td>0.43</td>
<td>0.25</td>
<td>0.50</td>
<td>0.17</td>
<td>0.60</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$\beta_{\text{Imp,Bh}}$</th>
<th>$\beta_{\text{SE,LTG}}$</th>
<th>$\beta_{\text{LTG,Int}}$</th>
<th>$\beta_{\text{Sat,OE}}$</th>
<th>$\beta_{\text{SE,OE}}$</th>
<th>$\beta_{\text{OE,Int}}$</th>
<th>$\beta_{\text{OE,Bh}}$</th>
<th>$\beta_{\text{SN+Sat,Int}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>0.25</td>
<td>0.05</td>
<td>0.20</td>
<td>0.10</td>
<td>0.05</td>
<td>0.02</td>
<td>0.01</td>
<td>0.02</td>
</tr>
</tbody>
</table>

various other sources of information (e.g., location data, weather forecasts, personality questionnaires, social network information) to provide tailored coaching. The role of the model is to run simulations to estimate the effect of different coaching strategies on the behavior. Thus, the model is run for each user and each possible strategy, and outcomes for the behavior are compared. The strategy that yields the highest result for the behavior concept ($\text{Beh}$) is considered to be the most promising.

In context of developing the behavior change system, eight coaching strategies were defined. Each of the strategies targets one of the psychological concepts in the model, and consists of coaching messages that are sent to the user through a smartphone app during one week. The presumed effect of the possible coaching strategies is implemented as a subtle boost of 5% in the first three days of the simulation. For example, if the simulated coaching...
strategy targets the self-efficacy, the value of the self-efficacy concept is increased with 5% of the potential improvement (i.e., its distance to the maximum), as in Equation 7.1.

$$if(coaching\_strategy == SE) : SE(t) = SE(t) + 0.05 \cdot (1 - SE(t))$$ (7.1)

Since the outcomes of these simulations are used in a real-life application, their closeness to reality is not the only relevant measure. Specifically, for the behavior change intervention to be both effective and engaging to the user, the outcomes (which determine the activated coaching strategy) should be diverse.

Figure 7.2 and Figure 7.3 both show an example of two simulations for a certain person from our dataset, with on the left side the concept targeted by the simulated coaching strategy and on the right side the change in the behavior based on that coaching strategy. The simulated effect of the coaching strategy is visible in the first half of the graphs on the left hand side. Since the behavior value increases most in Figure 7.2, that coaching strategy (self-efficacy) would be preferred over the other (intentions).

Figure 7.2: Example of a model simulation, showing the concept targeted by the coaching strategy (left) and its effect on the behavior (right). The targeted concept is self-efficacy \(SE\). The vertical axis represents the simulation values; the horizontal axis represents the time.

Figure 7.3: Example of a model simulation, showing the concept targeted by the coaching strategy (left) and its effect on the behavior (right). The targeted concept is intentions \(Int\). The vertical axis represents the simulation values; the horizontal axis represents the time.
Chapter 7. Exploring parameter tuning for analysis of a computational model

7.4 Methods

When investigating the model behavior with respect to its application in a behavior change intervention, it is important to base this analysis on realistic combinations of values for concepts in the model. Therefore, we collected data about the state of the concepts from ten potential users with varying levels of physical activity. These initial values were obtained from extensive validated questionnaires for assessing the psychological constructs (e.g., Frank et al., 2009; Rovniak et al., 2002). That way, we collected a set of reliable assessments for the combinations of initial values of the concepts. Four users in the data set were male and six were female. The average age was 30.1 years (range [21, 42]).

When running the model simulations for this user set with the parameter values based on correlations found in literature (see Section 7.3.1), the results yield very little diversity. When sorting the eight coaching strategies based on the predicted outcome of the behavior concept, only three different strategies appear in the first two positions for all ten users. Overall, it appears that some strategies are more often among the best options, while others occur only towards the end of the order (see Figure 7.4).

The surprising lack of diversity among these findings raises the question: what factors cause the stability of the model’s simulation outcomes? In order to find an answer to that question, we took a numerical approach, and explored the possible underlying causes in two directions.

First, it is theoretically possible that the initial values are too uniform to yield diverse outcomes. This possibility can be examined by running the model with the original parameter values (based on indications from literature) on many combinations of random starting values. Second, it is possible that the structure of the model in combination with the parameter values implies a certain importance of the concepts, which is stronger than the individual differences. This possibility can be examined by searching for a parameter set that does yield diverse results among the collection of realistic initial values.

7.4.1 Quantifying the lack of diversity

In order to evaluate the (lack of) diversity of the simulation outcomes, it has to be quantified systematically. In this research, the lack of diversity is computed by comparing all coaching strategy sequences for each pair of users, and applying a 0.01 penalty each time a strategy occurs in the same position. This choice is motivated by the particular application of
the computational model in our use case. Other operationalizations of diversity or other measures (e.g., entropy, variance) could be more relevant in other applications. The currently adopted approach implies a maximum cost of \( \frac{N(N - 1)}{2} \cdot 8 \cdot 0.01 = 3.6 \). For the results shown in Figure 7.4, the cost is 1.8.

7.4.2 Analyzing the influence of the initial values

To test whether the set of combinations of initial values plays a role in the constancy of the simulation results, we generate a set of 1,000 times 10 different combinations of random starting values for the concepts of the model, to mimic sets of 10 users. Then, we run the computational model on each of these initializations with the original parameter settings (see Table 7.1). For each of these simulations, we compute the cost as specified in Section 7.4.1.

7.4.3 Analyzing the influence of the parameter set

To investigate whether the lack of diversity in the simulation results is caused by the model’s parameter settings, we explore the ability of the model to produce diverse outcomes for the dataset of real users. This exploration is done by searching the parameter space with a simulated annealing algorithm (Kirkpatrick et al., 1983), in which we optimize for a reduction in the cost as specified in Section 7.4.1.

The solution space in this instance of the simulated annealing algorithm is represented by the possible values of the 16 parameters. Each solution is a set of parameter values (see Table 7.1), which describe the strength of the relations between the concepts. A neighboring solution is generated by adding a (positive or negative) change drawn from a normal distribution to each parameter value, while bounding them to the range [0,1]. The costs corresponding to each solution are calculated by running the model with the parameter values, and computing the costs of the outcomes according to Section 7.4.1. The probability of accepting a solution with a higher cost depends on the difference between the current cost and the new cost and the “temperature” \( T \), as in Equation 7.2. The temperature decreases with a factor 0.9 with each 100 iterations, and stops the algorithm when the initial temperature of 1.0 has decreased to a value below 0.00001. The final parameter set is used to run the model once more and calculate the corresponding costs, and the results are stored. The entire search process was repeated 75 times.

\[
acceptance\_probability = e^{(old\_cost - new\_cost)/T} \tag{7.2}
\]

7.4.4 Hypothesized outcomes

The experimental setup described above yielded a number of anticipated outcomes, enumerated below.

Hypothesis H1:

Because of the diverse combinations of values for the ten users and the extraordinary stability of the initial simulation results (in Figure 7.4), we expect that these starting values do not cause the stable outcomes. Therefore, we expect similar or lower levels of diversity when running the model on random sets of input values (see Section 7.4.2).
Hypothesis H2:
Since the parameter values determine the influence of the different concepts, we expect that they play a key role in the stability of the initial simulation results. Therefore, we expect more diversity when running the model on parameter sets found through simulated annealing (see Section 7.4.3).

Hypothesis H3:
Because of the computational model’s complexity, and the corresponding large number of degrees of freedom, we expect that many different parameter sets that are obtained through the simulated annealing search can lead to similar outcomes in terms of cost.

Hypothesis H4:
However, we still expect to see some global pattern in the parameter sets on average, i.e. that some parameters generally end up in the lower part of the range and some parameters in the higher part of the range.

Hypothesis H5:
If indeed such overall patterns are found, we expect that we will be able to explain them based on the underlying meaning of the concepts and the structure of the model.

As mentioned before, the combination of the dataset of real users with the original parameter values yielded a cost of 1.8 out of a possible 3.6. (See Section 7.4.1.)

7.5 Results
As mentioned before, the combination of the dataset of real users with the original parameter values yielded a cost of 1.8 out of a possible 3.6. (See Section 7.4.1.)

7.5.1 Influence of initial values
When running the model with the original parameter set on 1,000 different sets of 10 combinations of random initial values, the average cost of the simulation results is 2.21 out of 3.6. Even though in the set of 1,000 different outcomes there are some runs that produce more diversity than in the original situation, 92.5% of the results produce less diversity and correspondingly have a higher cost than the 1.8 of the initial solution. See Table 7.2 for an overview of the results.

Table 7.2: Results of original parameter settings and random initial values over 1,000 runs.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average cost</td>
<td>2.21459</td>
</tr>
<tr>
<td>Range of costs</td>
<td>[1.33, 2.96]</td>
</tr>
<tr>
<td>Standard deviation of cost</td>
<td>0.261168</td>
</tr>
<tr>
<td>Percentage where cost &gt; 1.8</td>
<td>92.5%</td>
</tr>
</tbody>
</table>
7.6 Discussion

7.5.2 Influence of parameter set

After 75 runs of the simulated annealing algorithm, the average cost of the model on the dataset of real users is 0.85 out of 3.6. The standard deviation is quite low, indicating that most of the outcomes are close to the average. See Table 7.3 for an overview of the results.

Table 7.3: Results of real initial values and parameters found through simulated annealing.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average cost</td>
<td>0.853867</td>
</tr>
<tr>
<td>Range of costs</td>
<td>[0.66, 0.99]</td>
</tr>
<tr>
<td>Standard deviation of cost</td>
<td>0.033898</td>
</tr>
</tbody>
</table>

When further exploring the parameter sets that yield these results, we see a large variety in the found parameter values for largely similar cost outcomes. Most parameters cover (almost) the full range of [0,1] in the 75 runs, with 14 out of 16 parameters approaching or reaching the limits in both directions within 0.1. The average parameter value is 0.5268, and the average standard deviation of each parameter is 0.2564, further supporting the observation that their values vary widely. Hence, there is no clear pattern of the parameter values for each individual solution found.

However, when looking at the parameter values averaged over the 75 runs, it appears that there are some patterns visible. That is, for some of the parameters, the average values deviates from the overall average of 0.5268. Table 7.4 shows the parameter values averaged over all 75 runs.

Table 7.4: Parameter settings.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$\beta_{\text{Sat,SE}}$</th>
<th>$\beta_{\text{Imp,Int}}$</th>
<th>$\beta_{\text{SE,Int}}$</th>
<th>$\beta_{\text{SE,Imp}}$</th>
<th>$\beta_{\text{Imp,Sat}}$</th>
<th>$\beta_{\text{Int+Beh,Sat}}$</th>
<th>$\beta_{\text{SE,Beh}}$</th>
<th>$\beta_{\text{Int,Beh}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg. value</td>
<td>0.5932</td>
<td>0.2760</td>
<td>0.5102</td>
<td>0.5321</td>
<td>0.8258</td>
<td>0.8320</td>
<td>0.3319</td>
<td>0.3086</td>
</tr>
<tr>
<td>Parameter</td>
<td>$\beta_{\text{Imp,Beh}}$</td>
<td>$\beta_{\text{SE,LTG}}$</td>
<td>$\beta_{\text{LTG,Int}}$</td>
<td>$\beta_{\text{Sat,OE}}$</td>
<td>$\beta_{\text{SE,OE}}$</td>
<td>$\beta_{\text{OE,Int}}$</td>
<td>$\beta_{\text{OE,Beh}}$</td>
<td>$\beta_{\text{SN+Sat,Int}}$</td>
</tr>
<tr>
<td>Avg. value</td>
<td>0.3677</td>
<td>0.6865</td>
<td>0.7381</td>
<td>0.4348</td>
<td>0.3042</td>
<td>0.5703</td>
<td>0.4610</td>
<td>0.6563</td>
</tr>
</tbody>
</table>

7.6 Discussion

The results provide grounds to investigate the hypothesized outcomes from Section 7.4.4. As expected, the results presented in Section 7.5.1 show that the lack of diversity in the original simulation is not caused by the initial values in our dataset. Although some of the sets of initial values produce more diversity than in the original simulation, the vast majority (92.5%) yields higher costs. Therefore, hypothesis H1 can be confirmed. On the contrary,
the costs found in Section 7.5.2 are considerably lower than in the original simulations. This confirms hypothesis H2, and indicates that the potential for more diversity in the simulation results is related to the parameter values.

The results described in Section 7.5.2 also showed a large variety in the parameter sets found through the optimization algorithm. This is in line with hypothesis H3, and can be explained by the model’s complexity. The high number of degrees of freedom in the model implies that many different parameter sets may produce similar results in terms of cost. However, when analyzing the found parameters in more detail, we do discern patterns in their values: some parameter values are generally high (e.g., $\beta_{\text{Imp},\text{Sat}}$), while others are low (e.g., $\beta_{\text{SE},\text{Beh}}$). This finding confirms hypothesis H4.

By looking at the meaning of the parameters with relatively low or high values, we attempt to explain why they end up with these values. Since we’re optimizing for high diversity in the effect on the behavior, it can be expected that concepts with a structurally high influence on the behavior will be dampened, while concepts with a structurally low influence will be increased. Indeed, we see that the parameters in Table 7.4 tend to weaken the path from the self-efficacy (the central notion of the social cognitive theory) to the behavior. For example, $\beta_{\text{SE},\text{Beh}}$ and $\beta_{\text{Int},\text{Beh}}$ are relatively low on average, thus decreasing the effect of the self-efficacy on the behavior. Similarly, $\beta_{\text{Imp},\text{Sat}}$ has a relatively high value, but since the impediments have a negative effect on the satisfaction, this will ultimately lead to a decrease in the self-efficacy. At first sight, the low average value of $\beta_{\text{Imp},\text{Int}}$ is surprising: one would expect that a strong (negative) influence on the intentions will transfer to the behavior. However, when taking a closer look at the data, the impediments seem to have generally low values (average: 0.29, minimum: 0.10, maximum: 0.39), so a high parameter would in fact cause an increase in the intentions and consequently a stronger influence on the behavior. In summary, hypothesis H5 is confirmed as well.

The observation of the surprising value for $\beta_{\text{Imp},\text{Int}}$ indicates that the outcomes are still dependent on the values in the dataset of users. Therefore, although we demonstrated that the relatively small set of users does not cause the lack of diverse outcomes, it would be interesting to see whether the findings scale to larger populations of users, or that the global patterns of the parameters change. Another limitation to the generalizability of this work concerns the dependency of the results on the structure of our model and the specific implementation of the relations between the concepts. Also, we have restricted ourselves to the simulated annealing algorithm, and further research should reveal whether the results are dependent on our choice of algorithm. However, our use case has demonstrated that applying parameter tuning techniques to analyze and better understand models has indeed given us insight in the model behavior. This novel use of parameter tuning is also transferable to other applications, with different underlying models and different evaluation (cost) measures.

In the context of the application of the model (see Section 7.3.2), we strive for both a close fit to reality and diverse outcomes. In other words, we look for a balance between keeping the parameters close to the indications found in literature and searching for a parameter set that yields diverse results. Therefore, we ran the algorithm described in Section 7.4.3 again, but with constraints to the generation of neighboring solutions, forcing them to stay within a distance of $\pm 0.1$ from the original parameters. This approach allowed us to increase the diversity of the outcomes (reducing the cost from 1.8 to 1.06), while keeping the literature-based parameters to some extent intact.
7.7 Conclusion

This research has established that applying parameter tuning techniques to analyze and better understand the behavior of dynamic computational models has indeed the potential to provide more insight in the structural properties of the models. In our use case, where we tried to increase the diversity in the results of the model simulations, we successfully demonstrated the cause of the initial lack of diversity. Subsequently, we were able to show that diverse outcomes can be achieved by finding suitable parameter values through simulated annealing, and that the global patterns of these parameter sets provide information about (or can be explained by) the structure of the model and the meaning of its concepts and relations. This novel application of parameter tuning techniques is transferable to other cases, with different underlying models and different evaluation (cost) measures, and is therefore a valuable new approach in the toolset of computational modelers and designers of mHealth systems.
Chapter 7. Exploring parameter tuning for analysis of a computational model

References


Plotnikoff, Ronald C., Sonia Lippke, Kerry S. Courneya, Nick Birkett, and Ronald J. Sigal (2008). “Physical Activity and Social Cognitive Theory: A Test in a Population Sample of Adults with Type 1 or Type 2 Diabetes”. In: *Applied Psychology* 57.4, pages 628–643.


<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Sections</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>Effect of changes in a social network on social contagion</td>
<td>8.1 Introduction, 8.2 Behavior change interventions, 8.3 Emotion contagion in social networks, 8.4 Experiments, 8.5 Discussion, 8.6 Conclusions and future work</td>
</tr>
<tr>
<td>9</td>
<td>Analysis of social contagion of physical activity</td>
<td>9.1 Introduction, 9.2 Social contagion model, 9.3 Experimental setup, 9.4 Results, 9.5 Discussion, 9.6 Conclusion</td>
</tr>
<tr>
<td>10</td>
<td>Effect of community membership on impact of a health program</td>
<td>10.1 Introduction, 10.2 Related work, 10.3 Methods, 10.4 Results, 10.5 Discussion, 10.6 Conclusion</td>
</tr>
<tr>
<td>11</td>
<td>Explaining changes in physical activity through social contagion</td>
<td>11.1 Introduction, 11.2 Related work, 11.3 Methods, 11.4 Results, 11.5 Conclusions</td>
</tr>
<tr>
<td>12</td>
<td>Effectiveness of social comparison on physical activity</td>
<td>12.1 Introduction, 12.2 Background, 12.3 Methods, 12.4 Results, 12.5 Discussion, 12.6 Conclusion</td>
</tr>
</tbody>
</table>
Abstract

Agent-based support systems are used to help people with developing and maintaining a healthy lifestyle. Interventions on the social network of an individual could play a role in achieving behavior change. In this paper, a method for finding effective network interventions to influence specific individuals is proposed. The effect of these interventions was analyzed by simulating the diffusion of emotional values about intentions and goals in a social network. Experiments showed that changing connections closer to the target have a larger influence than changing connections further from the target node. A comparison of the effect of the proposed interventions with all possible interventions showed that they are among the most optimal possible interventions. Finally, it was shown that nodes with fewer connections are easier to influence. The proposed interventions could form the basis for a support system that focus on affecting the social interaction between people in an online social network.
This chapter is based on:

8.1 Introduction

Developing and maintaining a healthy lifestyle is often a major challenge; e.g. see (Quinn et al., 2010). One of the problems is that decisions for unhealthy behaviors are mainly made in an unconscious manner and based on habits. Nowadays, agent-based intelligent systems are developed for monitoring and supporting people in changing their behavior, for example, of the type as addressed in Ambient Intelligence; e.g., (Cook et al., 2009). The main characteristics of Ambient Intelligence applications are that (1) they are nonintrusive, hidden in a person’s environment, (2) they are able to monitor using sensor systems, and (3) their interventions have a high extent of sensitivity to the context and state of the person.

In this paper, it is investigated how an agent-based system can exploit the social network around individuals to support them in their behavior change. It has been found since long that people are more successful in developing and maintaining a healthy lifestyle when they function in a supportive social context; e.g., (Wing and Jeffery, 1999; Zimmerman and Connor, 1989). For example, the number of friends or family members and the frequency of social contact were found to be positively associated with higher physical activity levels (Groenewegen et al., 2012; Spanier and Allison, 2001). As observed in (Kendall et al., 2011), such findings indicate that social networks can provide a leveraging mechanism for achieving and maintaining a healthy lifestyle; see also (McNeill et al., 2006).

The area of Social Networks has a tradition of more than 40 years, starting in Social Science, but more recently it has gradually developed stronger and moved into other disciplines as well, such as Biology, Neuroscience, Mathematics, Physics, Economics, Informatics, Artificial Intelligence, and Web Science; e.g., (Boccaletti et al., 2006; Valente, 2010). An important role of networks is that they form a basis for diffusion or contagion processes for various matters, for example, diseases, information, innovations, opinions, emotions, behaviors, and lifestyles; e.g., (Coviello et al., 2014). Monitoring and analyzing the dynamics of given diffusion or contagion processes is one thing. However, having such means of analysis available, it can be used for prediction as well, and in particular for what-if simulation: predicting what will happen if some action is undertaken, such as changing the strength of a certain connection. More generally, methods can be developed to determine what types of network intervention actions can be undertaken under which conditions in order to achieve some specific goal. Examples of such goals are: avoiding that an epidemic will develop, achieving that many persons will know about a new product that you bring out, achieving that more people will adopt a healthy lifestyle, and avoiding social isolation among elderly people.

Usually, the network interventions considered in such literature aim at affecting the whole network in a positive way. However, a more focused and less intrusive approach can be achieved when in a social network specific persons are selected that are most in need of being affected positively, and (only) network interventions are generated that have positive effects on these persons specifically. Such network interventions can be supported by an agent-based intelligent system that has knowledge about the structure of the network and the effect of changes in the structure. The system analyzes and affects social interactions between people in a network to stimulate healthy behavior of some selected persons.

The aim of such an agent-based support system is to achieve a behavioral change of an individual by changing the structure of the network around a person, which results in a different propagation of emotional values about intentions and goals.

In this paper, it is described how a number of specific network intervention strategies have
been designed, and how their effectiveness has been tested and evaluated in a comparative manner in a developed simulation environment. The adopted approach combines the predictive value of simulating computational models of social processes with social network analysis to investigate how social network interventions could be made more effective.

The remainder of this paper is organized as follows. Section 8.2 describes the traditional approaches towards behavior change and the role that social networks can play. In Section 8.3, the emotion contagion processes in social networks is described. This section also introduces the set of interventions in the structure of the social network, including heuristics to select an intervention based on a goal for a specific individual. The hypotheses, experimental framework and the experiment themselves are described in Section 8.4. Section 8.5 contains an analysis of the results and finally, Section 8.6 concludes the paper.

### 8.2 Behavior change interventions

#### 8.2.1 Traditional behavior change interventions

There are many different situations in which there are active attempts to change the behavior of people. For example, governments try to promote a more active lifestyle, to reduce the alcohol consumption, or to stimulate people to eat healthily. General practitioners and clinicians aim at changing the behavior of their patients to let them adhere to their therapeutic advices and to take their medication in the right manner. Interventions to change the behavior of people in these scenarios either target people as a group via mass media (e.g., in the form of public health campaigns), or try to focus on individuals. In the first case, one tries to change the opinions and intentions of all people at the same time by providing information via advertisements. In the second case, people get personalized support, either provided by a human coach, or as computer-tailored information (Smeets et al., 2007).

Nowadays, there is also an increasing number of (mobile) support systems that can help individuals that desire to change their behavior, for example to increase their physical activity, or to quit smoking. These systems focus on the cognitions and behavior of individuals as well. The most important elements are usually the monitoring of behavior, self-regulation (i.e., people can specify personal goals), and persuasive messaging (e.g., based on the principles of motivational interviewing).

The interventions that are studied in this paper are based on a different approach. Instead of targeting groups of people or individuals, the structure of the social network around people is the subject of the intervention.

#### 8.2.2 Social network interventions

There is quite some literature available on interventions that target the network of a group of people to achieve certain kinds of behavior change. In most of these papers, the goal is to find the nodes in the network that should change to achieve a change in the whole group. For example, the authors in (Valente and Pumpuang, 2006) suggest various kinds of strategies to find the agents in a network that are important to achieve a behavior change in the group. These strategies focus on persons in the network with larger numbers of connections, as they may affect many others. Such nodes could represent various types of people, such as celebrities, opinion leaders or experts in a certain field. In another paper (Borgatti, 2006), the network structure is exploited to find the key players to make sure that an innovation diffuses throughout the network. Similarly, in a strongly clustered network, influence can
take place according to a kind of repeated sequence of waterfalls, where at each step some time is passing to get a cluster affected, after which a next cluster is affected. Structural measures are often used to find the important nodes in these processes, which are regarded as bridging nodes. The authors in (Valente and Fujimoto, 2010) argue that these kind of nodes have more potential to act as change agent and these individuals are also able to adopt new behaviors faster than rest of the network. In (Valente, 2010, 2012), more can be found on the area of network interventions.

Often, the overall goal of network interventions is to apply an intervention at the group level rather than choosing an intervention based on an individual’s need. The latter is important, as every person is different and he/she requires personalized support based on his/her own behavioral characteristics. In this paper, the aim is to achieve a behavioral change of an individual by changing the network structure.

### 8.3 Emotion contagion in social networks

#### 8.3.1 Emotion contagion model

The underlying assumption of the structural network interventions is that the behavior of individuals is influenced by their emotional values about intentions and goals, and that these values can diffuse through the social network (Coviello et al., 2014). To analyze the effect of changes in the network structure, a model of contagion of emotions is used. In (Bosse et al., 2009), the authors used two ‘flavors’ of the contagion model (i.e., absorption and amplification) to study and simulate emotion contagion processes in groups. Furthermore, the computational model was used to study the role of emotion contagion to motivate people in a Twitter network to exercise regularly (Breda et al., 2012). The model is generic and does not only fit emotion contagion, but also covers behavior contagion. In this work, this contagion model (Bosse et al., 2009) is used to find the appropriate set of interventions.

The approach presented in this paper differs from the work in (Breda et al., 2012) in at least three ways. First, the social network is much bigger (34 nodes compared to 4 nodes) and is based on a realistic network. Second, the approach is based on indirect influence rather than direct influence, i.e. the change is realized by weakening or strengthening a path rather than through a particular person in the environment. Third, a much richer collection of interventions is explored.

The contagion process is modeled in terms of the contagion strength $\gamma_{BA}$ from node B to node A. It indicates the way in which the actions of node A are affected by the corresponding actions of node B. For example, if node B expresses his/her intention to visit a fitness center, node A would also feel the motivation to accompany node B. The contagion strength depends on three traits: the expressiveness of node B, $\varepsilon_B$, the openness of node A, $\delta_A$, and the channel strength from node B to node A, $\alpha_{BA}$. In addition to these personality traits, it is also possible to let other factors have an influence on the connection strength. Such factors could include the type of relationship, the frequency of interactions, the physical distance between two people and medium of communication, e.g. whether it is by phone, social media or direct. This will be included in future work. The formalization for the combined contagion strength is as follows:

$$\gamma_{BA} = \varepsilon_B \cdot \alpha_{BA} \cdot \delta_A$$ (8.1)
Chapter 8. Effect of changes in a social network on social contagion

For the sake of simplicity and to avoid more intricate computations, in the current work, the three values are not treated separately, but rather only a single connection strength value between two nodes is considered. Furthermore, each node in the graph is tagged with an emotion value, that represents the trait/state that is spread through contagion. The overall contagion strength $\gamma_A$ from the rest of the group towards node A is computed as in Equation 8.2.

$$\gamma_A = \sum_{B \neq A} \gamma_{BA}$$  \hspace{1cm} (8.2)

The aggregated impact $q_{SA^*}^*$ of all these nodes on state S (emotion level) of node A is computed by means of the weighted average as in Equation 8.3.

$$q_{SA^*}(t) = \sum_{B \neq A} \gamma_{BA} \cdot q_{SB}(t) / \gamma_A$$  \hspace{1cm} (8.3)

In the remainder of this paper, only one “emotional value” per node is used; thus, the specific interpretation of the emotion value (e.g., whether it is a goal, intention, etc.) is abstracted from.

8.3.2 Selection of target connections

An important question is to find out which types of interventions are more effective and which are less effective. A simple approach is to only consider interventions affecting direct connections to a specific person in need of change. However, perhaps such persons in the direct social environment are also not in a good state, as it happens often that direct contacts strongly affect each other. Therefore, it may be better to look further in the network for more positive sources, and aim at affecting the given person in a cascade-like, more indirect manner by an intervention directed to other connections. The interventions that are investigated in this study all target the strength of certain connections between two nodes in the network: either to decrease the strength of negative influences on the target node, or to increase the strength of positive influences on the target node.

In order to do so, first all strong paths to the target are identified. This is achieved by doing an exhaustive search of paths leading to the target node, up to a certain strength threshold. The path strength is calculated by multiplying the connection strengths of all connections in the path. In this search, cyclic paths are allowed, as emotion contagion effects also spread through circular paths. Then, the path is selected that has the highest combination of negativity (i.e., inverse positivity) of its nodes and strength of its connections. The pseudo code for finding this strong negative path is shown in Algorithm 8.1.

In the strong negative path, two connections are selected as locus for the interventions: (1) the first connection, that is closest to the target node, and (2) the strongest connection in the path. Once these two target connections are identified, four degrees of interventions are applied: (1) decreasing the connection strength to 75% between the current value and 0 (i.e., multiplying by 0.25), (2) decreasing the strength to 50% between the current value and 0, (3) decreasing the strength to 25% between the current value and 0, and (4) ‘cutting’ the connection in the path by decreasing the strength to 0. These four degrees of the intervention are considered, in order to be able to study the effects of different ‘sizes’ of the intervention.
Algorithm 8.1: Strong negative path strategy.

```
1 $g \leftarrow$ input graph
2 $t \leftarrow$ target node
3 $s \leftarrow$ strength threshold
4 $i \leftarrow$ interventions
5 function findStrongNegPaths($g, t, s$):
6     // Find all strong paths through exhaustive search,
7     // up to a certain path strength threshold $s$.
8     paths = findAllStrongPaths($g, t, s$)
9     // Find the path with the highest combination of path strength
10     // and path negativity in paths.
11     foreach path in paths do
12         // Calculate combination of strength and negativity.
13         path.score = calcStrength($g, path$) $\cdot$ calcNegativity($g, path$)
14         // If the score is higher than the score for the other paths,
15         // then store it.
16         if path.score > maxScore then
17             maxScore = path.score
18             result = path
19         end
20     end

A similar approach is followed for the increasing the strength of positive influences
on the target node. First, all weak paths (up to a certain length threshold) to the target are
listed. Then, the path with the highest combination of positivity of its nodes and weakness
(i.e., inverse strength) of its connections is identified. A comparable algorithm is used to
find the weak positive paths as the one described above. In the weak positive path, again
two connections are selected as locus for the interventions: (1) the first connection, that
is closest to the target node, and (2) the weakest connection in the path. On these target
connections, again four degrees of interventions are applied: (1) increasing the connection
strength to 25% between the current value and 1, (2) increasing the strength to 50% between
the current value and 1, (3) increasing the strength to 75% between the current value and 1,
and (4) maximizing the connection in the path by setting the strength to 1. Table 8.1 shows
an overview of the resulting 16 intervention types.

8.4 Experiments

8.4.1 Research questions

In this paper, several experiments are described that should contribute to the understanding
of the effect of different interventions in the connection strengths of a social network. A
number of hypotheses (H1–H4) about the differences in the effect of the heuristics have
been defined. The first aim of this research is to verify these hypotheses (listed below).
Chapter 8. Effect of changes in a social network on social contagion

H1. The interventions are based on imperfect heuristics, so some interventions may also lead to lower emotion values, compared to not intervening in the network structure.

H2. Given a certain locus of intervention, i.e. the targeted connection: the larger the change in the connection strength, the larger the effect on the target node. More specifically:
   a. The effect of intervention 1Aiv will be stronger than the effect of interventions 1Ai, 1Aii and 1Aiii.
   b. The effect of intervention 1Biv will be stronger than the effect of interventions 1Bi, 1Bii and 1Biii.
   c. The effect of intervention 2Aiv will be stronger than the effect of interventions 2Ai, 2Aii and 2Aiii.
   d. The effect of intervention 2Biv will be stronger than the effect of interventions 2Bi, 2Bii and 2Biii.

H3. The effects of interventions A (targeting the first connection in a path) can be weaker than the effects of interventions B (targeting the strongest/weakest connection in a path).

H4. The most successful interventions will always be at a connection within the three degrees of influence (Christakis and Fowler, 2007) from the target node.

Table 8.1: Overview of intervention types.

<table>
<thead>
<tr>
<th></th>
<th>Interventions to decrease strength of strong negative path (snp)</th>
<th>Interventions to increase strength of weak positive path (wpp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A Decrease strength of first connection in snp.</td>
<td>A Decrease strength of first connection in snp.</td>
</tr>
<tr>
<td></td>
<td>i Decrease connection strength to 75% between current value and 0.</td>
<td>i Increase connection strength to 25% between current value and 1.</td>
</tr>
<tr>
<td></td>
<td>ii Decrease connection strength to 50% between current value and 0.</td>
<td>ii Increase connection strength to 50% between current value and 1.</td>
</tr>
<tr>
<td></td>
<td>iii Decrease connection strength to 25% between current value and 0.</td>
<td>iii Increase connection strength to 75% between current value and 1.</td>
</tr>
<tr>
<td></td>
<td>iv Decrease connection strength to 0.</td>
<td>iv Increase connection strength to 1.</td>
</tr>
<tr>
<td></td>
<td>B Decrease strength of strongest connection in snp.</td>
<td>B Increase strength of weakest connection in wpp.</td>
</tr>
<tr>
<td></td>
<td>i Decrease connection strength to 75% between current value and 0.</td>
<td>i Increase connection strength to 25% between current value and 1.</td>
</tr>
<tr>
<td></td>
<td>ii Decrease connection strength to 50% between current value and 0.</td>
<td>ii Increase connection strength to 50% between current value and 1.</td>
</tr>
<tr>
<td></td>
<td>iii Decrease connection strength to 25% between current value and 0.</td>
<td>iii Increase connection strength to 75% between current value and 1.</td>
</tr>
<tr>
<td></td>
<td>iv Decrease connection strength to 0.</td>
<td>iv Increase connection strength to 1.</td>
</tr>
</tbody>
</table>
The second research question is to what extent the heuristics find the most optimal connection to be changed. In order to answer this question, the results from the intervention based on the heuristic search will be compared to the results of a brute-force test of all possible interventions.

As a third research question, it is investigated whether the outcomes of the heuristics are dependent on the connectedness of the target node, or more specifically, on the degree of the target node. To this end, the experiment is run on three different target nodes, each with a different level of connectedness, and compare the results.

### 8.4.2 Zachary social network

The network that is used in the experiments in this paper is based on the classical Zachary karate club network (Zachary, 1977). The data were collected from the members of a university karate club in 1977 and represent friendship relations. Originally, the network consists of 34 nodes and 78 undirected and unweighted edges. However, for our experiments, the network was transformed into a weighted directed version by randomly assigning weights to each of the directions of all edges. Additionally, random “emotion” values have been assigned to the nodes in the network.

Figure 8.1 depicts the resulting network. The emotion value of the nodes is indicated by their color: the lighter the color, the higher the emotion value. The degree of the nodes is shown by their size: the larger the node, the higher the degree.

An important property that needs to be validated before using this network for further experiments is the scale-free power-law distribution. It is known from literature that many real-world networks follow the power-law distribution: the fraction of nodes with degree $k$, denoted as $P(k)$ is approximated by a function $f(k) = ck^{-\gamma}$ for some $\gamma$, which typically lies
between 2 and 3, and a constant $c$.

In practice, it is often the case that not all data concerning a phenomenon follow the power-law distribution. Therefore, it is sometimes necessary to remove nodes which can be regarded as outliers, because they may prevent the network to follow the distribution. In the case of the Zachary network, one such outlier, a node with a single connection, was removed for analysis of its adherence to the power-law distribution. Furthermore, a parameter estimation method called sum of squared errors was used to find the most suitable values for $\gamma$ and $c$ (constant of proportionality), which are 1.9 and 45 respectively. As can be seen in Figure 8.2, our empirical data (blue curve) follow the ideal power-law pattern (red curve) quite well.

![Figure 8.2: Degree distribution of the Zachary social network vs. ideal power-law distribution.](image)

8.4.3 Experimental framework

The experiments consist of two parts. In the first part, the target connections for the interventions are determined, based on the heuristics described in Section 8.3.2, and the networks with the adjusted connection strength are generated. In the second part, these adjusted networks serve as input for the emotion contagion model described in Section 8.3.1. All scripts were developed in Python.

In order to run the experiments, several simulation parameters need to be set. First, one should determine on which time point to evaluate the results of the interventions. Because of the absorptive nature of the emotion contagion model, the emotion values tend to converge (i.e. regression to the mean), and are therefore difficult to compare at later time points. Exploratory simulations revealed that the emotion values at time step 15 were already more or less stabilized, but not yet converged completely, as illustrated in Figure 8.3. Therefore, $t = 15$ was chosen as the time point for the evaluation.

Two other simulation parameters that should be determined are the length threshold and the strength threshold for the path search algorithm, in order to limit the search of paths to the target to paths of a certain strength (for strong paths) or length (for weak paths). These thresholds were set at 0.05 and 5, respectively, because these values led to sufficiently good results within very acceptable computation time.
8.4 Experiments

8.4.4 Experiments and results

To answer the research questions, several experiments have been performed. The set of heuristics has been applied to three different target nodes: a very well connected node, an averagely connected node, and a weakly connected node. To guarantee that emotion value of a targeted node can increase, the nodes are selected from the subset of nodes with the lowest emotion value (i.e., 0.1). Otherwise, the regression to the mean would cause the emotion values of target nodes to always decrease. As measure for connectedness, the degree has been used. The following target nodes have been selected:

- Agent $n_1$ (degree 16)
- Agent $n_4$ (degree 6)
- Agent $n_{21}$ (degree 2)

For each of these nodes, all possible strong negative paths and weak positive paths were calculated, following the procedure described in Section 8.3.2. The exhaustive search for strong negative paths for the three target nodes $n_1$, $n_4$, $n_{21}$ resulted in respectively 33,279, 10,168, and 3,570 paths. From these, the ones with the highest combination of strength and negativity are chosen. For the weak positive paths, the exhaustive search resulted in 86, 53, and 32 different weak positive paths. The ones with the highest combination of weakness and positivity were chosen. The strength of the selected connections was then adjusted.

For the networks resulting from these interventions, a simulation of the emotion contagion process has been performed, which results in an emotion value of the target node at time step 15.

Table 8.2, Table 8.3 and Table 8.4 list the outcomes for the three different target nodes. The first column shows the intervention that was applied (cf. Table 8.1). The next column shows the two nodes between which the connection is altered. The third column shows the emotion value of the target node at time step 15, and the final column shows the difference between that emotion value and the emotion value at time step 15 when there was no

Figure 8.3: Emotion values of all nodes over time.
intervention. The shading of the cells containing the emotion values indicates their relative position: the higher the emotion value, the lighter the cell shading. In addition, the highest emotion value is printed in boldface.

Table 8.2: Results for target n4.

<table>
<thead>
<tr>
<th>Intervention</th>
<th>Connection</th>
<th>Emotion</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>from to</td>
<td>0.51389</td>
<td>0</td>
</tr>
<tr>
<td>1A-i</td>
<td>n3 n4</td>
<td>0.51471</td>
<td>0.000825</td>
</tr>
<tr>
<td>1A-ii</td>
<td>n3 n4</td>
<td>0.5154</td>
<td>0.001515</td>
</tr>
<tr>
<td>1A-iii</td>
<td>n3 n4</td>
<td>0.51589</td>
<td>0.002004</td>
</tr>
<tr>
<td>1A-iv</td>
<td>n3 n4</td>
<td>0.51609</td>
<td>0.002202</td>
</tr>
<tr>
<td>1B-i</td>
<td>n4 n3</td>
<td>0.514</td>
<td>0.000119</td>
</tr>
<tr>
<td>1B-ii</td>
<td>n4 n3</td>
<td>0.51414</td>
<td>0.000251</td>
</tr>
<tr>
<td>1B-iii</td>
<td>n4 n3</td>
<td>0.51428</td>
<td>0.000399</td>
</tr>
<tr>
<td>1B-iv</td>
<td>n4 n3</td>
<td>0.51445</td>
<td>0.000566</td>
</tr>
<tr>
<td>2A-i</td>
<td>n8 n4</td>
<td>0.51503</td>
<td>0.001149</td>
</tr>
<tr>
<td>2A-ii</td>
<td>n8 n4</td>
<td>0.51617</td>
<td>0.002282</td>
</tr>
<tr>
<td>2A-iii</td>
<td>n8 n4</td>
<td>0.51728</td>
<td>0.003399</td>
</tr>
<tr>
<td>2A-iv</td>
<td>n8 n4</td>
<td>0.51839</td>
<td>0.0045</td>
</tr>
<tr>
<td>2B-i</td>
<td>n3 n8</td>
<td>0.5049</td>
<td>-0.00899</td>
</tr>
<tr>
<td>2B-ii</td>
<td>n3 n8</td>
<td>0.4976</td>
<td>-0.01629</td>
</tr>
<tr>
<td>2B-iii</td>
<td>n3 n8</td>
<td>0.49167</td>
<td>-0.02222</td>
</tr>
<tr>
<td>2B-iv</td>
<td>n3 n8</td>
<td>0.48684</td>
<td>-0.02704</td>
</tr>
</tbody>
</table>

In addition to the simulations of the effect of the specific changes in the network, a set of simulations of the effect in all possible changes in the connection strengths in the network has been performed in a brute force manner. The Zachary network, on which the experiments are based, has 34 nodes. Initially, the graph was undirected with 78 connections but to apply the emotion contagion model in a more realistic way, these undirected connections were changed into directed ones. Hence the graph has 156 edges (connections). The graph is implemented as a key-value pair, where a key denotes a particular node and value is either a connection to another node or an emotion value. The algorithm for generating all possible changes loops through all pairs of nodes and changes all those connections one by one, ultimately storing the results in a CSV file for further processing. The pseudo code for the brute force strategy is given in Algorithm 8.2.

The brute force algorithm resulted in 1,248 possible changes in the network: the eight different interventions have been applied to all 156 connections in the network. These adapted structures together with the original network have been used as basis for the simulation of the emotion contagion. Finally, this produced the final emotion values at time step 15 of all 34 nodes in 1,249 situations. The brute force analysis allows for comparing
Table 8.3: Results for target $n_1$.

<table>
<thead>
<tr>
<th>Intervention</th>
<th>Connection</th>
<th>Emotion</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>from to</td>
<td>0.46534</td>
<td>0</td>
</tr>
<tr>
<td>1A-i</td>
<td>$n_{11}$ $n_1$</td>
<td>0.46795</td>
<td>0.00261</td>
</tr>
<tr>
<td>1A-ii</td>
<td>$n_{11}$ $n_1$</td>
<td>0.47062</td>
<td>0.00528</td>
</tr>
<tr>
<td>1A-iii</td>
<td>$n_{11}$ $n_1$</td>
<td>0.47336</td>
<td>0.00803</td>
</tr>
<tr>
<td>1A-iv</td>
<td>$n_{11}$ $n_1$</td>
<td>0.47618</td>
<td>0.01084</td>
</tr>
<tr>
<td>1B-i</td>
<td>$n_{11}$ $n_1$</td>
<td>0.46795</td>
<td>0.00261</td>
</tr>
<tr>
<td>1B-ii</td>
<td>$n_{11}$ $n_1$</td>
<td>0.47062</td>
<td>0.00528</td>
</tr>
<tr>
<td>1B-iii</td>
<td>$n_{11}$ $n_1$</td>
<td>0.47336</td>
<td>0.00803</td>
</tr>
<tr>
<td>1B-iv</td>
<td>$n_{11}$ $n_1$</td>
<td>0.47618</td>
<td>0.01084</td>
</tr>
<tr>
<td>2A-i</td>
<td>$n_{32}$ $n_1$</td>
<td>0.46586</td>
<td>0.00053</td>
</tr>
<tr>
<td>2A-ii</td>
<td>$n_{32}$ $n_1$</td>
<td>0.46638</td>
<td>0.00105</td>
</tr>
<tr>
<td>2A-iii</td>
<td>$n_{32}$ $n_1$</td>
<td>0.4669</td>
<td>0.00156</td>
</tr>
<tr>
<td>2A-iv</td>
<td>$n_{32}$ $n_1$</td>
<td>0.4674</td>
<td>0.00207</td>
</tr>
<tr>
<td>2B-i</td>
<td>$n_{19}$ $n_{33}$</td>
<td>0.46547</td>
<td>0.00013</td>
</tr>
<tr>
<td>2B-ii</td>
<td>$n_{19}$ $n_{33}$</td>
<td>0.46559</td>
<td>0.00025</td>
</tr>
<tr>
<td>2B-iii</td>
<td>$n_{19}$ $n_{33}$</td>
<td>0.46571</td>
<td>0.00037</td>
</tr>
<tr>
<td>2B-iv</td>
<td>$n_{19}$ $n_{33}$</td>
<td>0.46582</td>
<td>0.00048</td>
</tr>
</tbody>
</table>

Algorithm 8.2: Brute force strategy.

1. $g \leftarrow$ input graph
2. $n \leftarrow$ nodes
3. $e \leftarrow$ edges
4. $i \leftarrow$ interventions
5. \textbf{foreach} node $n$ in graph $g$ do
6. \hspace{1em} \textbf{foreach} incoming edge $e$ in ($n, n'$) do
7. \hspace{2em} \textbf{foreach} intervention $i$ do
8. \hspace{3em} temp = $e$.weight
9. \hspace{3em} \text{// Modify the current edge weight with the next intervention.}
10. \hspace{3em} $e$.weight = $i$.next()
11. \hspace{3em} \text{// Apply the contagion model.}
12. \hspace{3em} emotionContagion($g$)
13. \hspace{3em} \text{// Save the results in CSV format.}
14. \hspace{3em} \text{// Return to initial value.}
15. \hspace{3em} $e$.weight = temp
16. \hspace{2em} end
17. \hspace{1em} end
18. end
Chapter 8. Effect of changes in a social network on social contagion

Table 8.4: Results for target n21.

<table>
<thead>
<tr>
<th>Intervention</th>
<th>Connection</th>
<th>Emotion</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>from to</td>
<td>0.4036</td>
<td>0</td>
</tr>
<tr>
<td>1A-i</td>
<td>n34 n21</td>
<td>0.38036</td>
<td>-0.02324</td>
</tr>
<tr>
<td>1A-ii</td>
<td>n34 n21</td>
<td>0.35198</td>
<td>-0.05163</td>
</tr>
<tr>
<td>1A-iii</td>
<td>n34 n21</td>
<td>0.31738</td>
<td>-0.08622</td>
</tr>
<tr>
<td>1A-iv</td>
<td>n34 n21</td>
<td>0.27532</td>
<td>-0.12828</td>
</tr>
<tr>
<td>1B-i</td>
<td>n34 n21</td>
<td>0.38036</td>
<td>-0.02324</td>
</tr>
<tr>
<td>1B-ii</td>
<td>n34 n21</td>
<td>0.35198</td>
<td>-0.05163</td>
</tr>
<tr>
<td>1B-iii</td>
<td>n34 n21</td>
<td>0.31738</td>
<td>-0.08622</td>
</tr>
<tr>
<td>1B-iv</td>
<td>n34 n21</td>
<td>0.27532</td>
<td>-0.12828</td>
</tr>
<tr>
<td>2A-i</td>
<td>n33 n21</td>
<td>0.42412</td>
<td>0.02052</td>
</tr>
<tr>
<td>2A-ii</td>
<td>n33 n21</td>
<td>0.44071</td>
<td>0.03711</td>
</tr>
<tr>
<td>2A-iii</td>
<td>n33 n21</td>
<td>0.45411</td>
<td>0.0505</td>
</tr>
<tr>
<td>2A-iv</td>
<td>n33 n21</td>
<td>0.46489</td>
<td>0.06129</td>
</tr>
<tr>
<td>2B-i</td>
<td>n19 n33</td>
<td>0.40426</td>
<td>0.00066</td>
</tr>
<tr>
<td>2B-ii</td>
<td>n19 n33</td>
<td>0.40488</td>
<td>0.00128</td>
</tr>
<tr>
<td>2B-iii</td>
<td>n19 n33</td>
<td>0.40548</td>
<td>0.001874</td>
</tr>
<tr>
<td>2B-iv</td>
<td>n19 n33</td>
<td>0.40604</td>
<td>0.002439</td>
</tr>
</tbody>
</table>

the effect of the changes suggested by the heuristics with the optimal effect.

8.5 Discussion

In this section, the results of the experiments are discussed and the research questions that have been formulated are answered.

8.5.1 Comparison of different strategies

The results of the different interventions, summarized in Table 8.2, Table 8.3 and Table 8.4, show a number of interesting findings.

First, for two of the target nodes, some interventions lead to a lower emotion value, compared to not intervening in the network structure. This finding is in line with our expectations as stated in hypothesis H1. For example, with agent n4 as target node, all interventions that increase the strength of the weakest connection in a weak positive path (<n3, n8>), i.e. interventions 2B-i to 2B-iv, lead to lower emotion values. A possible explanation for this finding is that agent n3 is connected to a number of nodes, which have relatively low emotion values. By increasing the influence of n3 on n8, the influence of the negative nodes on n4 is also enhanced. So, even though this connection is found in a weak and positive path, it may still be a connection to a ‘hub’ of negative nodes.

Second, for all target nodes and all targeted connections, the results show that larger changes in the connection strength lead to larger effects in the resulting emotion value. This outcome is consistent with our second hypothesis, H2. It holds for all subhypotheses, H2a,
H2b, H2c and H2d. This finding is also reflected in the fact that the interventions with the
highest resulting emotion value are either cutting a connection by decreasing its strength to
0 (target \textit{n1}), or increasing a connection strength to 1 (target \textit{n4} and target \textit{n21}).

Third, the results indicate that the interventions with the highest resulting emotion
value all act upon a direct connection of the target node. This outcome is in line with
the expectations in our final hypothesis, H4, but it contradicts the third hypothesis, H3.
This finding suggests that the biggest improvement can be found by considering only the
first-order connections of a certain individual. However, additional research should be done
to find out whether this is a structural finding, or an artefact of our particular dataset.

### 8.5.2 Optimal connection found

The results show that for each target node, an intervention was found that leads to an
improvement in the resulting emotion value, compared to not intervening in the network
structure. However, in order to evaluate the quality of the underlying heuristics, the results
were compared to the ordered list of all 1,249 possible results, as determined by the brute-
force search.

Table 8.5 gives an overview of the rank of the selected interventions and the situation
with no intervention, together with the corresponding percentile. As shown in the final
column, for each of the three target nodes, it was found that about half of all possible
interventions will lead to lower emotion values than when not applying an intervention. The
selected interventions, however, lead to emotion values in the 98\textsuperscript{th} or 99\textsuperscript{th}
percentile. For target node \textit{n21}, our heuristics were even able to find the best intervention out of the 1,249
possibilities.

<table>
<thead>
<tr>
<th>Selected intervention</th>
<th>No intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rank</td>
</tr>
<tr>
<td>Target \textit{n4}</td>
<td>15</td>
</tr>
<tr>
<td>Target \textit{n1}</td>
<td>7</td>
</tr>
<tr>
<td>Target \textit{n21}</td>
<td>1</td>
</tr>
</tbody>
</table>

Histograms of the data show that the selected interventions are far into the right tail of
the distribution of the possible outcomes.

Figure 8.4 illustrates the relative position of the selected intervention for target \textit{n1} in the
set of all possible outcomes for this node. The intervention is denoted by a green triangular
mark and can be found in the right tail. The resulting emotion value in the situation without
intervention is marked with a red triangle, and can be found around the mean.

Additional simulations revealed that deeper searches, i.e. with looser thresholds for
the length or strength of the investigated paths, still found the same target connections.
Furthermore, deeper searches led to an enormous increase in computation time, even with
only a marginal difference in the thresholds. For example, increasing the length threshold
from 5 links to 8 links, or decreasing the path strength threshold from 0.05 to 0.03, increased the computation time from mere seconds to hours/days.

8.5.3 Effect of connectedness

The third aim of this research was to investigate whether the connectedness of the target node affects the outcomes of the heuristics. As the results show, the strongly connected node, target \( n1 \), benefited most from cutting a connection to one of its neighbors. The less connected nodes, target \( n4 \) and target \( n21 \), benefited most from increasing a connection to one of their neighbors. However, more network data would be required to determine whether these findings are generalizable.

Another interesting finding lies in the range of outcomes of the interventions for each of the target nodes. For target \( n1 \), with degree 16, the difference between the highest and the lowest emotion value is approximately 0.01. For target \( n4 \), with degree 6, this difference is approximately 0.06 and for target \( n21 \), with degree 2, this difference is approximately 0.19. This could be explained by the fact that the effect of applying an intervention on one connection is diminished if the target node has many other connections. This implies that individuals with few connections are easier to influence than highly connected individuals.

8.6 Conclusions and future work

In this paper, a number of network interventions have been introduced that focus on achieving an effect on a specific individual in a social network. The interventions exploit the structure of the network around the individual to find strong transitive connections to people with a negative influence and weak transitive connections to people with a positive influence. Via simulation experiments, the effect of the changes in the structure of the network on the individual have been studied.

The aim of the proposed social network interventions is to investigate whether it is
possible to design a support system that can influence a person by affecting the social interaction with other persons. The simulation experiments have shown that this is the case: it is possible to identify and change specific connections in a network such that it has a positive effect on the targeted individual.

There are no clear conclusions yet on the type of change to connections. The results indicate that changing the strength of nearby connections seems to have more influence than changing very strong/weak connections in a path. Larger changes in the connection strengths have larger effects.

With respect to the selection of connections that should be changed, it can be concluded that the proposed heuristics performed very well. For all targets that have been considered, it turned out that the connection that was selected to be changed was among the best possible options. Another finding was that changes in connections to sparsely connected nodes have larger effects than changes in connections to highly connected nodes.

More research is required to validate the initial findings, preferably with larger and real network data. Future work includes plans to perform the heuristics on data of an existing social network of people that aim at increasing their physical activity.

The proposed network interventions can be used in the context of a behavior change support system. By emphasizing or filtering the information about behavior, intentions or goals that is shared between people in an online social network, it is possible to alter the social interactions and thus the spread of influences throughout the network. Another application would be to suggest a person to make strong ties with more positive people around him or her or restrain completely or to reduce communication with negative people.

The research presented in this paper combines computational modeling of social processes with social network analysis in order to steer interventions in a social network. Thereby, this approach uses the predictive value of model simulations to investigate the suitability of a set of heuristics for identifying the most effective change in a network to positively influence an individual. It therefore lays a valuable foundation for network interventions that aim to stimulate health behavior change. In addition, it can serve as a reference point for further investigation of contagion processes in social networks.

**Acknowledgments**

The authors would like to thank Aart van Halteren for his useful comments.
Chapter 8. Effect of changes in a social network on social contagion

References


Abstract
It is known that opinions, attitudes and emotions spread through social networks. Several of these cognitions influence behavioral choices. Therefore, it is assumed that the level of physical activity of people is influenced by the activity levels of the people in their social network. We have performed an experiment with 20 participants between 19 and 28 years old, measuring their physical activity levels for 30 days, in order to observe if there is a contagion effect due to the relationships in the social network. Using our social contagion model, we investigated if people will become more or less active according to the contacts with their peers within the network. Our model correctly predicts the direction of the change (increasing or decreasing) in 80% up to 87% of the cases investigated.
This chapter appeared as:

9.1 Introduction

Physical inactivity is a major public health challenge in the developed world and is recognized as a global epidemic (Allender et al., 2006). Insufficient physical activity is a risk factor for cardiovascular diseases and other conditions. The amount of physical activity of a person is usually represented by the physical activity level (PAL). It refers to “any bodily movement produced by skeletal muscles that results in energy expenditure” (Caspersen et al., 1985).

The global recommendation for daily exercise is an accumulated 30 minutes of moderate intensity activity, such as cycling, brisk walking or swimming, in segments of at least 10 minutes per activity (Vandelanotte et al., 2007). Research has shown that a large part of the Western population does not meet these guidelines (Hallal et al., 2012). Sports medicine and public health constituencies also acknowledged a concern about the deleterious health consequences of insufficient physical activity (Dunstan et al., 2012).

It is known that social influences play a key role in lifestyles and are fundamental to whomever wants to maintain healthy behavior (Zimmerman and Connor, 1989). Several aspects underlying a lifestyle, such as emotions, opinions and behaviors, can spread through a social network, in a process called “social contagion”. Social contagion theories explain how one’s social network influences these aspects and how this social environment can provide support in changing them (Breda et al., 2012).

This research builds on the belief that the social environment can be used as an unobtrusive, even unconscious and therefore suitable way of supporting people to become more physically active (Aral and Walker, 2010; Medicine (US). Committee on Health and Practice, 2001). To develop practical applications of lifestyle interventions based on social influence, it is important to have a thorough comprehension of the dynamics underlying the social contagion process. To contribute to this understanding, we have performed an experiment in which we compared the predictions of a model that describes social contagion in a community (Bosse et al., 2009, 2015; Duell et al., 2009) with the actual change in physical activity level. Our assumption is that the model can be applied to describe the spread of behavior, considering the willingness to be more active is led by the emotions, attitudes and motives of each person. In the experiment, we constructed a graph of the social network of a group of young adults between 19 and 28 years old, using the strength of the relations between the participants. In addition, we assessed the important characteristics of each participant, such as their openness and expressiveness. For all participants, we collected the PAL data during a period of 30 days. The change of the physical activity per participant was compared with the change predicted by the model.

The paper is organized as follows: in Section 9.2, the social contagion model is explained in more detail. Section 9.3 describes the setup of the conducted experiment. In Section 9.4, we present the results obtained, and we discuss the results in Section 9.5. In Section 9.6, we conclude our explorations and discuss some ideas about possible future work.

9.2 Social contagion model

In this section, we briefly summarize the computational model of social contagion and explain how it is used to predict the change in the physical activity level (Bosse et al., 2009, 2015; Duell et al., 2009).
In the context of this research, the factor that is assumed to spread through the social network is the physical activity level of the people in the network. The extent to which people express themselves, which affects the strength of their influence on others, is captured by the concept of expressiveness. Similarly, the extent to which people are open to receive influence is represented by the openness. The strength of the relation between two people in the network is described by the connection strength. These concepts form the key parameters of the contagion model, see Table 9.1. They are formalized as real numbers between 0 and 1.

Table 9.1: Parameters for personal and social characteristics.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical activity level of person A</td>
<td>$q_A$</td>
</tr>
<tr>
<td>Expressiveness of person A</td>
<td>$\varepsilon_A$</td>
</tr>
<tr>
<td>Openness of person A</td>
<td>$\delta_A$</td>
</tr>
<tr>
<td>Connection strength between sender B and receiver A</td>
<td>$\alpha_{BA}$</td>
</tr>
</tbody>
</table>

The contagion process is modeled in terms of the contagion strength $\gamma_{BA}$ between sender B and receiver A. This contagion strength represents the influence of sender B on receiver A. The contagion strength depends on three aspects of the relationship: the expressiveness of person B, $\varepsilon_B$, the openness of node A, $\delta_A$, and the connection strength between person B and person A, $\alpha_{BA}$. The contagion strength between sender B and receiver A is calculated as in Equation (9.1).

$$\gamma_{BA} = \varepsilon_B \alpha_{BA} \delta_A$$ (9.1)

The overall contagion strength $\gamma_A$ represents the total contagion strength of all connections of person A in the network. It is computed as in Equation (9.2).

$$\gamma_A = \sum_{B \neq A} \gamma_{BA}$$ (9.2)

The proportional weight of the contagion from sender B to receiver A is computed as in Equation (9.3).

$$\omega_{BA} = \varepsilon_B \alpha_{BA} / \sum_{C \neq A} \varepsilon_C \alpha_{CA}$$ (9.3)

The aggregated impact $q_A*$ of all connections of person A is calculated by means of a weighted average as in Equation (9.4).

$$q_A^{*} = \sum_{B \neq A} \omega_{BA} q_B$$ (9.4)
Experimental setup

The goal of the experiment is to compare the actual change in the activity level of people in a network with the change predicted by the computational model explained in Section 9.2. To do so, an empirical experiment with people that were part of a social network was conducted for 30 days. Characteristics of the persons and their relations were gathered via a questionnaire and objective data about their physical activity was collected with an electronic activity monitor.

The network consisted of 25 participants, all between the age of 19 and 28. The participants were recruited from one person’s social network, thus, every participant has at least one connection to another node in the network. Five of the 25 participants provided less than 25 days of useful data and were taken off the experiment, which left a number of 20 participants.

As we use a stable network with relationships that have been established before the start of the experiment, we expected the changes of the physical activity level would be small. This is due to the fact that no external trigger was introduced, e.g., a support system or an encouragement program for doing more activities. Nevertheless, the fact that people are participating in this experiment could intensify their awareness of others’ physical activity levels. Therefore, changes could still occur in a smaller ratio.

At the start of the experiment, an intake questionnaire was administered using an online survey software tool. Via this questionnaire, information was obtained about the participants: (1) physical activity level, (2) personal characteristics, (3) level of friendship with the other participants, and (4) frequency of contact with other participants. This information was used to determine the values for the parameters of the computational model.

During the period of 30 days, the participants wore an activity monitor (Fitbit One*) that

\[
\Delta q_N(t + \Delta t) = q_N(t) + \gamma N(q^*_N(t) - q_N(t))\Delta t
\]

(9.5)

This computational model of social contagion has been used in several studies. For example, it was applied to predict the emotion levels of team members, in order to maintain emotional balance within the team (Duell et al., 2009). If the team’s emotion level was found to become deficient, the model, which was embedded in an ambient agent, provided support to the team by proposing the team leader to give his employees a pep talk (Duell et al., 2009). Another study experimented with simulations of changes in the social network structure in order to guide the contagion process in a certain direction (Klein et al., 2014).

In the current research, the contagion model that was developed and simulated on emotions is applied on the contagion of physical activity, which to our knowledge has not been done in earlier research. The involvement of the main aspects, such as expressiveness, openness and connection strength are based on the model of social contagion. A detailed method for determining these parameter values for the computational model is described in Section 9.3.

*http://www.fitbit.com/one
kept track of their daily physical activity. In addition to the data obtained from the Fitbit, short questionnaires were used regularly to collect additional data about their exercise.

The data about the participants’ characteristics and their relations is used as basis for the parameters values in the model; the activity data is used as initial input for the model simulations and to compare the outcome of the simulations with. In the next sections, we explain the specific steps that were taken to convert the collected data to numerical values that are suitable for the computational model.

9.3.1 Physical activity level

The participants’ physical activity data that was collected by the Fitbit activity monitor was automatically stored in the Fitbit servers. After the period of 30 days, this data was exported from the participants’ personal Fitbit accounts. The exported data consists of the number of steps taken per day, the number of minutes that the participant was fairly or highly active per day and the number of floors climbed per day. These numbers were divided by the number of recommended steps, fairly/highly active minutes and floors, which are 10,000, 30 and 10 respectively. Days that contained less than 1,500 steps were considered as days that the participant (partly) forgot to wear the Fitbit, so these days were discarded.

Finally, weights were assigned to each aspect, according to their importance. The number of steps taken gives the best estimation of the amount of physical activity and is therefore the most important. The number of fairly/highly active minutes is chosen to be slightly more important than the number of floors, because meeting the recommendation of 30 active minutes per day contributes more to a physically active lifestyle than climbing 10 floors. Therefore, the PAL is calculated as in Equation (9.6).

$$\text{PAL} = \left( \frac{\text{steps}}{10000} \times 0.7 \right) + \left( \frac{\text{am}}{30} \times 0.2 \right) + \left( \frac{f}{10} \times 0.1 \right),$$

where $am$ is the active minutes and $f$ is the number of floors climbed.

9.3.2 Tie characteristics

The strength of connections between people is a combination of the amount of time, emotional intensity, intimacy, frequency of contact and reciprocal services (Soares and Lopes, 2014). In this model, a distinction between different types of interaction was added because of the assumption that contact in real life and one-to-one communication (also through private chats) both contribute to a higher level of contagion than contact that takes place in group conversations through smartphones or social media and by observing someone’s public posts on social media. Therefore, a combination of the level of friendship and the frequency and type of interaction is used to calculate a value for the strength of the connection.

The parameter used to represent the tie strength from node B to node A is the connection strength ($\alpha_{BA}$). This parameter was operationalized by a combination of the type of relation and the frequency of contact. These aspects were measured through questions included in the intake questionnaire. The levels of friendship were measured using a scale on which each participant rated all other network members as: unknown (0.0), acquaintance (0.2), good acquaintance (0.4), friend (0.6), best friend (0.8) and partner (1.0). In addition, two
questions concerning the frequency of interaction were included, distinguished by type. The participants gave an estimation, only for the participants who they stated to be connected to in the previous question, of how often they interact with them. These two questions about contact in real life or in private conversations and contact in groups or on social media were answered by the following scale, using the accompanied assigned values: less than once a month (0.0), 1–2 times a month (0.2), once a week (0.4), 2–5 times a week (0.6), once a day (0.8) and more than once a day (1.0). The formula for the tie strength is shown in (9.7).

$$\alpha_{BA} = (fl \times 0.6) + (crl \times 0.25) + (cg \times 0.15),$$

where $fl$ is the friendship level, $crl$ is the amount of contact in real life (i.e., private conversations), and $cg$ is contact in groups and social media.

### 9.3.3 Personality traits

Personality traits of a person were measured by statements that give an indication of the expressiveness and openness of this person. We formulated a number of statements based on the aspects extraversion, openness to new experience and agreeableness from the Big Five Inventory (Gosling et al., 2003) that were taken as a measure of the values of the participants’ expressiveness and openness. The statements that were used to assess these values are listed in Table 9.2. For each domain, three out of six statements were reversed. When using questionnaires with only positive (or negative) sentences, the subjects may be biased in a positive (or negative) way.

Participants were asked to assess how strongly they agreed or disagreed with each statement. A value was assigned to each answer as follows: strongly disagree (0.0), disagree (0.25), neutral (0.5), agree (0.75) and strongly agree (1.0). Some statements were reversed and therefore, the score was subtracted from 1 to obtain the right score for these answers.

We assumed that the expressiveness in our model depends on the Extraversion domain from the Big Five Inventory, and that the Openness and Agreeableness can be used for the degree of openness of the receiver in our model. Thus, statements 1 to 6 determined the value for expressiveness, which is the average of the calculated score over these six questions. The average value of statements 7 to 18 represents the overall value for openness.

### 9.3.4 Network structure

Figure 9.1 shows the social network of the participants. Each node represents a participant in the experiment, and the arrows are the direction of the connection, according to the questionnaires filled out by the participants. The thickness of the edges represents the weight of the connection. An overview of the structural characteristics of the network can be found in Table 9.3.

The degrees of the nodes in the graph follow a normal distribution, according to the Shapiro-Wilk test ($p = .723, H_0 =$ normal distribution), Anderson-Darling test ($p = 0.585, H_0 =$ normal distribution), and Jarque-Bera test ($p = .629, H_0 =$ normal distribution).

The network forms one connected component, which is due to the nature of the selection of the participants. Despite the fact that all of the participants are from the same class, the density of the network is not too high (47.9% of the possible edges existing).
Table 9.2: Statements measuring expressiveness and openness. (Reversed statements are marked by an asterisk (*).)

<table>
<thead>
<tr>
<th>Extraversion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I keep my feelings and thoughts to myself.*</td>
</tr>
<tr>
<td>2. I am assertive.</td>
</tr>
<tr>
<td>3. I am outgoing and enthusiastic.</td>
</tr>
<tr>
<td>4. I think carefully before I speak.*</td>
</tr>
<tr>
<td>5. I am shy and do not like to be the center of attention.*</td>
</tr>
<tr>
<td>6. I often post things on social media (Facebook and Instagram).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Openness to new experience</th>
</tr>
</thead>
<tbody>
<tr>
<td>7. I have a strong opinion.*</td>
</tr>
<tr>
<td>8. I am interested in what others do and think.</td>
</tr>
<tr>
<td>9. My decisions are thoughtful.*</td>
</tr>
<tr>
<td>10. I rather remain in my current safe habits and environment than trying and exploring new things.*</td>
</tr>
<tr>
<td>11. I am open for suggestions, ideas and opinions of others.</td>
</tr>
<tr>
<td>12. I am a curious person.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Agreeableness</th>
</tr>
</thead>
<tbody>
<tr>
<td>13. I have a rigid personality.*</td>
</tr>
<tr>
<td>14. I am easily influenced by what others think or do.</td>
</tr>
<tr>
<td>15. I have a strong feeling of empathy.</td>
</tr>
<tr>
<td>16. I am difficult to persuade by other people.*</td>
</tr>
<tr>
<td>17. I have a distant personality towards others.*</td>
</tr>
<tr>
<td>18. I feel sorry for other people very quickly.</td>
</tr>
</tbody>
</table>

Table 9.3: Structural characteristics of the network.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size</td>
<td>20</td>
</tr>
<tr>
<td>Edges</td>
<td>182</td>
</tr>
<tr>
<td>Minimum degree</td>
<td>3.0</td>
</tr>
<tr>
<td>Maximum degree</td>
<td>38.0</td>
</tr>
<tr>
<td>Mean degree</td>
<td>18.2</td>
</tr>
<tr>
<td>Std. deviation degree</td>
<td>9.082</td>
</tr>
<tr>
<td>Graph density</td>
<td>0.479</td>
</tr>
</tbody>
</table>

9.3.5 Data preparation and analysis

This study aims to investigate whether the computational model of social contagion is able to predict the change in the PALs of the participants correctly. In order to derive the direction of change of the PAL of each person, the trendlines for each participant were calculated. They were used to determine whether there was an increase or decrease in their physical activity.
activity during the experimental period. This is later compared with the direction of change predicted by the model.

The computational model needs an initial value for the activity level of each of the participants to start the simulation with. Therefore, the initial value is highly important and needs to be chosen carefully. As there is a relatively large variation in the PAL values for different days, it would be unrealistic to use only the PAL from one single day as the starting point for the model. Figure 9.2 shows the PAL values for one participant during the 30 days of the experiment. Since trendlines of the real data were calculated and the value of the trendline at the start provide a good aggregation of the first days of activities, we used the first point of the trendline for each participant as the initial value for the model. The model was then used to simulate the activity level of all participants for 30 days.

The computational model of contagion (Bosse et al., 2009, 2015; Duell et al., 2009) was implemented in Python, which performed calculations following the formulas discussed in Section 9.2. It takes three matrices as input for computing the change of PAL over time for each person. The first matrix contains the initial PALs for each participants, the second one contains values of the connection strengths in the network, and the third one contains values for each person’s expressiveness and openness. The values in the latter two matrices were the same for all days, as it was assumed that the network structure and the personality
characteristics did not change in the period of 30 days.

Before the model was run for the first set of experiments, some tuning was performed. To achieve a realistic speed of change, all contagion strengths were divided by 10, which resulted in realistic levels of contagion over time and in plausible values for changes in the levels of physical activity.

9.4 Results

The trendlines of the physical activity levels (based on the real data) can be seen in Figure 9.3(a). Mann-Kendall tests showed that the trendlines are not significant in 75% of the cases. That fits in our assumption that big changes should not occur in a relatively short period of time within a stable network.

Figure 9.3(b) shows the simulation results for each participant, using a simulation period of 30 days and with parameter values calculated as described in Section 9.3.2 and Section 9.3.3.

Table 9.4 shows that in 80% of the cases, the slopes of the trendlines have the same direction (increasing or decreasing) as the model predicts. It shows that, for stable networks and for a short period of time, the model can predict the direction of the change of the activity level of people with high precision.

As not all of the trendlines presented a clear slope, we took off the slopes with a tau value less than 0.03, i.e. the lines that were almost flat. This value is obtained by a Mann-Kendall test for analyzing the significance of the trendlines for the experiment. The Mann-Kendall test statistically assesses if there is a monotonic upward or downward trend over time of the variable under analysis. In our case, this variable is the PAL of each person. The test has the initial assumption that there is no monotonic trend as its null hypothesis ($H_0$). If the null hypothesis is rejected, then we have a reasonable indication that there is a trend. The tau value is the variable that assesses the slope of the trend, in case $H_0$ is rejected. Table 9.4 shows an 87% accuracy after taking out the trendlines that were almost flat.

Another way to determine the adequacy of the model is to compare the error of the model predictions with the error of the trendlines. However, this comparison is not fair, as the trendlines consider all the data points in the period, fitting the best linear regression...
(a) Trendlines for each participant during 30 days.

(b) Simulation results of the social contagion model.

Figure 9.3: Graphs with trendlines of real data and simulation results of the model.

to the entire data set, while the model simulations only use one initial value (based on the start value of the trendline). To make the predictions of a regression line comparable to the model, we created linear graphs using only the first seven days of data, so it uses the same amount of input data as the model.
Table 9.4: Comparison of line tendency between contagion model and real data\(^a\) and comparison of line tendency between contagion model and real data after removing trendlines with a slope less than 0.03 \(^b\). Correct predictions\(^c\) are shown in the last row.

<table>
<thead>
<tr>
<th>Model</th>
<th>Real data</th>
<th>Matches(^a)</th>
<th>Matches(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXP01</td>
<td>Up</td>
<td>Down</td>
<td>No</td>
</tr>
<tr>
<td>EXP02</td>
<td>Down</td>
<td>Down</td>
<td>Yes</td>
</tr>
<tr>
<td>EXP03</td>
<td>Up</td>
<td>Up</td>
<td>Yes</td>
</tr>
<tr>
<td>EXP04</td>
<td>Up</td>
<td>Up</td>
<td>Yes</td>
</tr>
<tr>
<td>EXP05</td>
<td>Up</td>
<td>Up</td>
<td>Yes</td>
</tr>
<tr>
<td>EXP06</td>
<td>Up</td>
<td>Up</td>
<td>Yes</td>
</tr>
<tr>
<td>EXP07</td>
<td>Up</td>
<td>Up</td>
<td>Yes</td>
</tr>
<tr>
<td>EXP08</td>
<td>Down</td>
<td>Down</td>
<td>Yes</td>
</tr>
<tr>
<td>EXP09</td>
<td>Up</td>
<td>Up</td>
<td>Yes</td>
</tr>
<tr>
<td>EXP10</td>
<td>Down</td>
<td>Up</td>
<td>No</td>
</tr>
<tr>
<td>EXP11</td>
<td>Down</td>
<td>Down</td>
<td>Yes</td>
</tr>
<tr>
<td>EXP12</td>
<td>Up</td>
<td>Up</td>
<td>Yes</td>
</tr>
<tr>
<td>EXP13</td>
<td>Up</td>
<td>Down</td>
<td>No</td>
</tr>
<tr>
<td>EXP14</td>
<td>Down</td>
<td>Down</td>
<td>Yes</td>
</tr>
<tr>
<td>EXP15</td>
<td>Down</td>
<td>Down</td>
<td>Yes</td>
</tr>
<tr>
<td>EXP16</td>
<td>Down</td>
<td>Down</td>
<td>Yes</td>
</tr>
<tr>
<td>EXP17</td>
<td>Down</td>
<td>Up</td>
<td>No</td>
</tr>
<tr>
<td>EXP18</td>
<td>Up</td>
<td>Up</td>
<td>Yes</td>
</tr>
<tr>
<td>EXP19</td>
<td>Up</td>
<td>Up</td>
<td>Yes</td>
</tr>
<tr>
<td>EXP20</td>
<td>Down</td>
<td>Down</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Total matches\(^c\) 16 (80%) 13 (87%)

Table 9.5 shows the mean squared error for three situations: trendlines based on all data, linear graphs extrapolated from the first seven days and the predictions of the contagion model. The values of the lines or model predictions in the three cases were compared with the actual data, the differences were squared and the average of these errors was calculated.

Table 9.5: Mean squared errors (MSE) of prediction lines.

<table>
<thead>
<tr>
<th>Method</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trendline</td>
<td>0.3613</td>
</tr>
<tr>
<td>Extrapolation</td>
<td>6.8992</td>
</tr>
<tr>
<td>Contagion model</td>
<td>0.4272</td>
</tr>
</tbody>
</table>
9.5 Discussion

Considering that our network is stable, with no formations of new connections nor additional support systems attached to it, we would not expect major changes in the daily level of activities of the participants. Still, we would expect some changes to happen due to the continuous social contagion effect between people from the same class and the increased awareness because of their participation in the experiment.

Figure 9.3(b) shows that the activity levels in the model simulations tend to converge to an average after some time. This is a consequence of the fact that the simulations assume that neither the network nor the personal characteristics of the participants change.

There is a large difference between the mean squared errors shown in Table 9.5 for the model predictions and the extrapolation of the first seven days. Extrapolating the first seven days of physical activity and creating new trendlines for each participant increased the mean squared error to 6.8992. This shows that our model results in a far better prediction than linear regression using a similar amount of information. The large error for the extrapolation can be explained by the high variation in the PALs of the participants per day, which makes it difficult to predict the trend based on a few days only. If we take a longer period of analysis, we can reduce the error for the statistical trendline.

In this experiment, we assumed that social contagion was the only factor that influences change in physical activity level. However, this is a clear simplification. Factors like weather and daily duties also have an effect on a person’s changes in activity. Similarly, there are also other factors than stated in this research that influence the contagion, such as group norms, physical closeness, age and many more. This makes that our model does not provide a complete description of the contagion process. Related to this, the structure of a social network itself is usually not independent of the characteristics of the participants. This effect is called homophily, which describes the tendency of people to connect with others that have the same lifestyle (Valente, 2010).

9.6 Conclusion

The results of the experiment show changes in physical activity levels of all members of the social network, in smaller or bigger ratios. The computational model of contagion predicted in 80% of cases correctly whether the PAL increases or decreases, which is the vast majority of the participants, and it grows to 87% if only people with a clear change in activity level are considered. We can conclude that the model has a good accuracy in predicting the tendency of the physical activity levels in a small network of 20 people, in a relatively short period of time.

The current experiment aims to understand the behavior of a stable network and answers the question whether it is possible to use the model to predict the direction of change of activity. These outcomes provide valuable insights for further explorations. To be able to draw conclusions about more dynamic situations, additional experiments should be done. These experiments should last longer, have a higher number of participants, and include changes in the network, e.g. changes in connections or applying interventions in order to influence people’s behavior using external triggers.

Future work should also investigate other strategies for parameter tuning and compare the resulting parameters with the values obtained from the questionnaires.
Acknowledgments

The authors would like to thank prof. dr. Jan Treur for his valuable input and feedback on the manuscript.
References


Chapter 9. Analysis of social contagion of physical activity


Abstract

Influence on health behavior from peers is well known and it has been shown that participants in an online physical activity promotion program are generally more successful when they share their achievements through an online community. However, more detailed insights are needed into the mechanisms that explain the influence of a community on physical activity levels (PAL).

This paper discusses a detailed analysis of a data set of participants in an online physical activity promotion program. The analysis focuses on the comparison two groups of participants, namely participants who will join an online community at some point in time and participants who will never join such a community. A well-balanced selection is made to eliminate to a large extent factors that dilute the effect of the willingness to partake in a community. We create statistical models that describe the PAL increase at the end of the program. A comparison of these models shows that participants that will participate in a community not only have a higher PAL at the start of the program, but also that the PAL increase is significantly greater compared to participants that will not become community members.

The results further support the hypothesis that the possibility to share achievements is an important feature of successful health promotion programs. At the same time, it raises the question whether part of the success is caused by a selection bias, as people that are willing to participate in a community are already more active at the start.
This chapter is based on:

10.1 Introduction

Engaging in sufficient physical activity has many beneficial effects on physical and mental health (Conn et al., 2011; Eime et al., 2013), while low levels of physical activity have been associated with increased risks of cardiovascular diseases, cancer, diabetes, and mental illness (Lee et al., 2012). Despite this, a large proportion of the Western population does not meet the guidelines of being moderately to vigorously active for at least 30 minutes on at least five days a week (Haskell et al., 2007). Therefore, physical activity promotion is a priority in most Western countries and many (online) intervention programs exist. It is important to understand which elements of these physical activity program are effective or could potentially accelerate the impact of these health promotion programs.

Previous research has already revealed that being part of an online social network in a health promotion program is correlated with a higher level of physical activity. In an earlier study based on a data set of 4,333 participants (Groenewegen et al., 2012), it was shown that the activity level of people that participated in an intervention aiming at stimulating physical activity (for 14 weeks) who became a member of an online community was significantly higher compared to people that chose not to become a member of that community. However, based on this result, it is not possible to conclude that a higher level of physical activity is the result of being member of an online social network. It is also possible that there is a selection bias: the people that are more willing to participate in such a community are the more active people or the people that have a higher motivation to become more active. Besides that, there are other possibly confounding factors. For example, the people participating in a community could be selected in a season that is more suitable for physical activity, the participants could be recruited in companies with higher average activity levels, or the people that opt in for a community are biased with respect to gender or have psychological traits that makes them more intrinsically motivated to share their achievements.

In this paper, we also compare the level of physical activity of people that become member of an online social network within a physical activity promotion program with people that choose not to. We build on the previous work in two ways. First, we perform a very balanced selection in the two groups of people that we compare: we compensate for possible differences in starting season, in country and in gender. This way, we can rule out a number of alternative explanations for the higher level of physical activity of community members. Second, we compare both the starting activity level and the change in activity level over the time of the intervention for the two groups, rather than focusing on correlations between mean physical activity level and characteristics of participants. In this way, we can investigate whether people that become member of an online community are different with respect to their starting activity level and whether an intervention aiming at increasing physical activity has a different effect on them. In other words: does the possibility to participate in an online community accelerate the impact of the physical activity promotion program, possibly by attracting a different type of person? Note that the people that were selected as community participants are people that became member of a community at some time during their usage of the system, but not necessarily during the active period of the physical activity promotion program. We are therefore not able to conclude whether the actual participation in a community has positive effect.

The remainder of this paper is organized as follows. In Section 10.2, we discuss some related work about the influence of online social networks and (mobile) healthy lifestyle interventions on behavior. Section 10.3 presents the data, its characteristics and our methods.
for processing it. In Section 10.4, we provide statistical analyses to answer our research questions. We conclude with our main findings and directions for future work in Section 10.5 and Section 10.6.

10.2 Related work

Previous analyses (Groenewegen et al., 2012) show that there is a positive relation between being part of the online community of a physical activity intervention and the physical activity level of participants. The online community therefore matters. It has also been shown that the number of contacts in the online community does not have a significant effect on the physical activity level, while network density even has a significant, negative effect. On the other hand, adding online community features to an Internet-mediated walking program did not increase average daily step counts, but did reduce participant attrition (Richardson et al., 2010).

Online social interaction plays an important role in forming or adapting some kind of behavior based on the peer’s behavior. It has been studied recently that online social networks are equally responsible (as offline networks) in the diffusion of one’s emotions to another (Coviello et al., 2014). It is often difficult to adopt new behavior and adhere to it, but it has been shown that close social circles (such as family, friends, and co-workers) are helpful in sustaining a healthy lifestyle (McNeill et al., 2006; Zimmerman and Connor, 1989). In (Breda et al., 2012), the role of online social interactions is discussed in the context of developing and maintaining a healthy lifestyle, e.g. an ambient system can continuously monitor and help people to alter their social ties in order to sustain healthy behavior. Having an infrastructure like a social network already available, social network interventions could be designed to leverage the full potential of a social network (Klein et al., 2014), for example in case of a health behavior change program.

With the rise of mobile technology, there has also been a steep increase in the number of healthy lifestyle interventions that are available through a smartphone. As of May 2016, the number of apps in the Health & Fitness category has grown to 67,552 for the Google Play Store (AppBrain, 2016) and 68,248 for the iTunes App Store (PocketGamer.biz, 2016). A systematic review of apps that promote physical activity has shown that even though most apps apply only a few behavior change techniques (Abraham and Michie, 2008; Middelweerd et al., 2014), a majority of these apps (approximately 58%) provided a form of social support or social change (Middelweerd et al., 2014). This was done, for example, through providing chat possibilities among users or through enabling a link to an external virtual social network, where users could share their goals or achievements (Middelweerd et al., 2014).

It is widely believed that mobile technology can be a useful tool to promote physical activity among a large part of the population. First, average smart phone ownership numbers are high: 68% in the United States, with higher numbers among young adults (86% in ages 18–29 years and 83% in ages 30–49 years) (PewResearchCenter, 2015), and 80% in The Netherlands (GfK, 2015) in the third quarter of 2015. This means that interventions designed for smartphones can theoretically reach a large number of people. Second, mobile interventions are always accessible to the user, and also allow for continuous monitoring and (if applicable) feedback. Also, similar to interventions delivered over the Internet, mobile interventions can reduce stigma and lower the barrier for people to address their (health)
issues (Griffiths et al., 2006). In combination with the relatively high number of apps that enable social support or social change, these advantages of mobile interventions imply that smartphone apps are a very suitable means to guide social influence for behavior change.

10.3 Methods

This section describes the data set that is used for the analysis. In Section 10.3.1, we describe the process of data collection and the resulting data. Section 10.3.2 describes the way in which we processed the data to select suitable subsets, and some of the structural characteristics of the selected social network components are presented in Section 10.3.3.

10.3.1 Data collection

The analysis presented in this paper uses a data set of people \( n \approx 50,000 \) that participated in an online physical activity promotion program. The promotion program has three different phases. The first phase is a one-week assessment period, that is used to evaluate the user’s activity level during his/her daily routine. The assessment is followed by the second phase: a 12-week plan that aims to gradually increase the user’s activity level towards a specified end goal. The goal is determined based on the physical activity reported during the assessment week. After the plan, the members of the program can opt to start a new 12-week plan to further increase their activity level or simply continue with the activity goal set during the last week of their program. This constitutes the third phase.

The activity promotion program provides an online community and joining this social network is optional for the users. Each member of the community can connect to other users (i.e., become online friends), exchange messages and see the relative achievements of themselves and their connections (which are only visible after the participants confirm their connection). Around 5,000 people in the data set opted to join the online community at some point in time during their usage of the system.

The participants in the program wear an activity monitor device that measures their physical activity level (PAL). When they register to the program via the website, the participants fill in their gender, age, and nationality. In addition, the data set contains information about the date that people start the program, the company they work in (if the program is offered via a company), and their friendship connections with other participants. In order to ensure anonymity of the participants, their age was omitted from the data set before the analysis.

10.3.2 Data selection

As our aim is to compare the difference in the physical activity level between people that become part of a community and people that will not become member of a community, we select two subsets of the data. The first subset is the intervention group: participants in the physical activity promotion program that at some point in time opted in for the community. The second subset is the control group: participants in the physical activity promotion program that never opted in for the community. In this section, we describe how we selected those two subsets.

The data is represented in two files. One file is a GEXF (Graph Exchange XML Format) file and represents the network structure of the community. The other file consists of the
PAL values of all participants, and their personal characteristics, such as gender, body mass index (BMI), corporation and country.

As we want to be able to consider the mutual effect of friendship relations on the activity level of participants in future research, we select a number of connected components from the community. A connected component (or just component) of an undirected graph is a subgraph in which any two nodes are connected to each other by edges, and which is connected to no other nodes in the supergraph. In order to extract the components, we use Python’s NetworkX library (Hagberg et al., 2008), which is based on the community detection algorithm Tarjan’s algorithm with Nuutila’s modifications (Nuutila and Soisalon-Soininen, 1994; Tarjan, 1972). The algorithm is based on the principle of strongly connected components, where each node in a graph has a bidirectional connection. The total number of communities that are found by the algorithm is 395. One of them is a large community with 3,926 participants; the second largest community consists of 42 participants. Figure 10.1 shows an overview of the number of participants in each of the components. The components are ordered by size, and the largest component (of 3,962 participants) is left out.

For all connected components, we extract the PAL (physical activity level) values for the individuals in each of the components. Since there are multiple consecutive plans (i.e., periods of twelve weeks in which people are stimulated to increase their activity level), the PAL values used in the analysis represent the first 12-week plan, in order to ensure fair comparisons.

Not all detected network components are used in the analysis. We select only components with (1) a limited number of participants for whom other data is missing and (2) a minimal difference between the plan start dates of the members of the component. For some participants in the online community, no other personal data or PAL data was available. Only components with at most one such participant were eligible for inclusion. For the (earliest and latest) start dates of the plans in the component, the maximum difference is four months. This is done to ensure that the participants in the community were using the program around the same time, so the community was ‘active’. As a result, we discard the largest component, because of the fact that the earliest and latest start dates are three years apart. The second component with 42 nodes is not included in the analysis because a lot of data is missing for that component.

This selection process yielded ten of such connected components, consisting of 109 individuals in total. We left out 25 individuals for whom PAL values were missing for one or more weeks, for instance because they dropped out of the program. Eventually, this resulted in 84 individuals in the intervention group.

For the control group, we select a set of individuals who did not opt for the community, but who are otherwise similar to the participants in each of the components in the intervention group. We balance the data with respect to the following characteristics: the participants work in the same companies, their plan earliest and latest start and date are similar to the corresponding component, and their gender ratio is also similar to the matching component. As the number of non-community individuals is much larger than the number of individuals within a community, we randomly select a set of around five times the size of the number of people in the community component with corresponding characteristics, resulting in a set of 501 people. For this data set, we also avoid including individuals with missing data, i.e. individuals who dropped out of the program or who had missing PAL data for one or more weeks. In total, this resulted in a set of 498 participants. Based on the selection of the
individuals in both groups, the PAL values are extracted for the two subsets.

![Figure 10.1: Number of nodes in each of the components.](image)

A summary of the selected data is given in Table 10.1 and Table 10.2. In Table 10.1, each row shows several meta-data characteristics of the selected components of the network. ‘Component’ is the identifier of the network component, and ‘Number of Participants’ represents the total number of participants in the component. It can be seen that the size of the selected components varies between 7 and 26 participants. ‘Dropouts’ shows the number of people that were omitted from the component, because at least one week of PAL data was missing. The ‘Start (earliest)’ and ‘Start (latest)’ columns show the earliest or the latest date on which people in a component started their first plan.

The last column of Table 10.1 shows the number of individuals with similar characteristics identified in the non-community data set and the number of them that were randomly selected for the control group. As mentioned earlier, since the number of people that did not opt in for the community is much larger than the number of people who did, the size of the control group is about five times the intervention group size. For example, for component A (consisting of 26 participants), 130 participants were randomly selected from a set of 2,735 individuals with similar characteristics. However, for some components, we could not find enough individuals with similar characteristics for the non-community data subset. For example, only six individuals were found for the non-community data subset corresponding to component G.

Table 10.2 illustrates different characteristics of people in the components, such as their nationality. The ‘Country’ column shows that the participants in each of the communities are from the same country, namely Germany, the Netherlands or the United States. ‘Number of corporations’ shows whether all people in a certain component work in the same or different organizations. It is possible that people in a community work in different organizations, like in components C, H and J. In rest of the components, the participants all work in the same company. The column ‘Gender ratio’ provides information about the ratio of male and
Chapter 10. Effect of community membership on impact of a health program

Table 10.1: Meta-properties of selected components.

<table>
<thead>
<tr>
<th>Component</th>
<th>Number of participants</th>
<th>Dropouts</th>
<th>Start (earliest)</th>
<th>Start (latest)</th>
<th>Number of participants non-community</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>26</td>
<td>0</td>
<td>25-01-2010</td>
<td>22-03-2010</td>
<td>130 / 2,735</td>
</tr>
<tr>
<td>B</td>
<td>15</td>
<td>4</td>
<td>15-02-2010</td>
<td>26-04-2010</td>
<td>70 / 838</td>
</tr>
<tr>
<td>C</td>
<td>13</td>
<td>4</td>
<td>18-05-2009</td>
<td>17-05-2010</td>
<td>65 / 178</td>
</tr>
<tr>
<td>D</td>
<td>9</td>
<td>1</td>
<td>16-05-2009</td>
<td>20-07-2009</td>
<td>45 / 74</td>
</tr>
<tr>
<td>E</td>
<td>9</td>
<td>3</td>
<td>25-01-2010</td>
<td>12-04-2010</td>
<td>45 / 2,839</td>
</tr>
<tr>
<td>F</td>
<td>8</td>
<td>0</td>
<td>25-05-2009</td>
<td>19-04-2010</td>
<td>40 / 608</td>
</tr>
<tr>
<td>G</td>
<td>8</td>
<td>6</td>
<td>19-04-2010</td>
<td>21-06-2010</td>
<td>6 / 6</td>
</tr>
<tr>
<td>H</td>
<td>7</td>
<td>6</td>
<td>15-02-2010</td>
<td>07-06-2010</td>
<td>30 / 35</td>
</tr>
<tr>
<td>I</td>
<td>7</td>
<td>1</td>
<td>22-02-2010</td>
<td>22-03-2010</td>
<td>35 / 358</td>
</tr>
<tr>
<td>J</td>
<td>7</td>
<td>0</td>
<td>02-03-2009</td>
<td>27-07-2009</td>
<td>35 / 335</td>
</tr>
</tbody>
</table>

female participants in each of the communities. ‘Average BMI’ represents the average BMI for each of the components.

Table 10.2: Characteristics of participants in selected components.

<table>
<thead>
<tr>
<th>Component</th>
<th>Country</th>
<th>Number of companies</th>
<th>Gender ratio (M/F:%)</th>
<th>Average BMI</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>DE</td>
<td>1</td>
<td>M:88.5, F:11.5</td>
<td>25.85</td>
</tr>
<tr>
<td>B</td>
<td>NL</td>
<td>1</td>
<td>M:100.0, F:0.0</td>
<td>25.74</td>
</tr>
<tr>
<td>D</td>
<td>US</td>
<td>1</td>
<td>M:89, F:11</td>
<td>31.76</td>
</tr>
<tr>
<td>E</td>
<td>DE</td>
<td>1</td>
<td>M:34, F:66</td>
<td>23.97</td>
</tr>
<tr>
<td>F</td>
<td>US</td>
<td>1</td>
<td>M:87, F:13</td>
<td>30.78</td>
</tr>
<tr>
<td>G</td>
<td>NL</td>
<td>1</td>
<td>M:58, F:42</td>
<td>25.60</td>
</tr>
<tr>
<td>H</td>
<td>NL</td>
<td>4</td>
<td>M:100.0, F:0.0</td>
<td>32.10</td>
</tr>
<tr>
<td>I</td>
<td>NL</td>
<td>1</td>
<td>M:86, F:14</td>
<td>28.05</td>
</tr>
<tr>
<td>J</td>
<td>DE</td>
<td>3</td>
<td>M:86, F:14</td>
<td>25.65</td>
</tr>
</tbody>
</table>

10.3.3 Structural analysis of the components

As described in the previous section, we selected 10 components from the community for the analysis, ranging from 7 to 26 participants each and with different configurations. The difference in the structural characteristics between the components can be seen in Figure 10.2. Social network analyses were run on the components in order to understand the structure of the connections.
The components are mostly sparse networks with a low average degree and low clustering coefficient, meaning that the neighbors of each node are not well connected among
themselves. All nodes within each component belong to the same country, the countries being the Netherlands, the USA and Germany. Because of the nature of the online friendship connections, all connections in the network are bidirectional.

Details of two components will be given to further illustrate the data. Component I has the highest average density and clustering coefficient, both more than 60%. It also presents a small diameter, which means that the nodes are very well connected, and are very close to each other. In this network, the degree of the nodes ranges from 2 to 12, with every connection being bidirectional. One of the nodes with the highest degree is connected to all the other nodes in the network, having an important role for the social influences in this component.

Component E is also a well-connected component with a small network diameter, as in most social networks in real life. This component has an average density and clustering coefficient of around 50%, which makes the network well connected, but not very dense. Two nodes have only one connection, and the rest of the network presents a very good clustering coefficient. Table 10.3 shows the detailed characteristics of this component.

Table 10.3: Detailed characteristics of component E.

<table>
<thead>
<tr>
<th>Number of nodes</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edges</td>
<td>32</td>
</tr>
<tr>
<td>Average degree</td>
<td>3.556</td>
</tr>
<tr>
<td>Average path length</td>
<td>1.583</td>
</tr>
<tr>
<td>Network diameter</td>
<td>3</td>
</tr>
<tr>
<td>Density</td>
<td>0.444</td>
</tr>
<tr>
<td>Average clustering coefficient</td>
<td>0.622</td>
</tr>
<tr>
<td>Country</td>
<td>Germany</td>
</tr>
</tbody>
</table>

10.4 Results

We perform several steps to answer our main question: is (the change in) the physical activity level of people that become part of an (online) community different from people that do not become member of such a community?

10.4.1 Visual comparison

Our first analysis is based on a visual comparison of the differences between the two groups. The average PAL values for both groups during twelve weeks (84 days) are shown in Figure 10.3. The figure illustrates that community people are more active, since their average PAL is higher than the average PAL of the non-community participants. It also shows that the linear trendline of both groups has a different slope.

10.4.2 Multiple linear regression model

For a more thorough analysis, we use statistical methods. In the second step of the analysis, a multiple linear regression model is fitted to predict the average physical activity level
at the end of the program (i.e., the last three weeks) based on whether a person is in the community group and the average PAL at the start of the program as predictors. For the average PAL at the start of the program, we consider only the second and third week. The first week is left out, because this week is usually a bit atypical, presumably due to novelty effects of starting the program.

To measure the average difference between groups, a dummy variable (Community) is coded with the value ‘1’ if a person is in the community and ‘0’ if the person did not opt in for the community. The results are illustrated in Table 10.4. A significant regression model was found ($F(2,579) = 227, p < .001$). The model accounts for 44% of the variance in the PAL values of the participants at the end of the program, $R^2 = .4395$. Both predictor variables, Start-PAL and Community, are statistically significant, $p < .05$. The model shows that the predicted PAL for the last three weeks is equal to $0.23562 + 0.05061 \times \text{Community} + 0.85041 \times \text{Start-PAL}$, where Community is 0 or 1. The model signifies that being member of a community is associated with an increase of approximately 0.05 in physical activity level.

Table 10.4: Analysis using multiple linear regression.

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>p-value</th>
<th>95% C.I.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.23562</td>
<td>0.06743</td>
<td>&lt;.001</td>
<td>[0.103183, 0.36806]</td>
</tr>
<tr>
<td>Start-PAL</td>
<td>0.85041</td>
<td>0.04091</td>
<td>&lt;.001</td>
<td>[0.77005, 0.93076]</td>
</tr>
<tr>
<td>Community=1</td>
<td>0.05061</td>
<td>0.02327</td>
<td>.0300</td>
<td>[0.00490, 0.09631]</td>
</tr>
</tbody>
</table>

10.4.3 Linear mixed model

The regression model described in the previous section only compares the PAL at the start with the PAL at the end. A linear mixed model can be used to take into account all days of data (except for the first week, as mentioned above). Since the data are longitudinal by
nature, we follow the approach as outlined in (Bliese and Ployhart, 2002). A sample of the data is shown in Table 10.5. Each row represents one day’s PAL for an individual, and there are 77 rows (eleven weeks) for each individual. As discussed in (Bliese and Ployhart, 2002), we first conduct the test using a simple model based on the generalized least square method and later add random effects to the intercepts in the simple model to see if the two models differ significantly. For this purpose, R’s NLME library is used (Pinheiro et al., 2016). Since we are primarily interested to see whether becoming member of a community makes a difference over time, the model includes an interaction term, i.e. a product of Community and Time. The results of the simple model (without random effects) are shown in Table 10.6. Here, becoming part of the community is taken as the reference group (Community=1), in contrast to the model presented in Table 10.4. The estimates associated with the predictor variables indicate the effect of the program on the PAL. So, the interaction term tests whether the effect of the program on the PAL of the participants is different for people inside or outside the community group. The results show that this is indeed the case: people perform differently in the two groups. In this analysis, not being part of the community is again associated with a lower PAL value (with a difference of approximately 0.06).

Table 10.5: Physical activity level data in long format

<table>
<thead>
<tr>
<th>Id</th>
<th>Time</th>
<th>PAL</th>
<th>Community</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>1.57800</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>9</td>
<td>1.85780</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>1.78080</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>582</td>
<td>82</td>
<td>1.5803</td>
<td>0</td>
</tr>
<tr>
<td>582</td>
<td>83</td>
<td>1.7658</td>
<td>0</td>
</tr>
<tr>
<td>582</td>
<td>84</td>
<td>1.4576</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 10.6: Analysis using generalized linear regression.

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.6933389</td>
<td>0.009814656</td>
<td>172.53167</td>
</tr>
<tr>
<td>Community=0</td>
<td>-0.0586920</td>
<td>0.010610159</td>
<td>-5.53168</td>
</tr>
<tr>
<td>Time</td>
<td>0.0005821</td>
<td>0.000192112</td>
<td>3.02993</td>
</tr>
<tr>
<td>Community=0 : Time</td>
<td>-0.0006525</td>
<td>0.000207683</td>
<td>-3.14177</td>
</tr>
</tbody>
</table>

The results of the more advanced random intercept model are shown in Table 10.7. In this model, we account for the fact that the start PAL (i.e., the intercept) of each of the individuals is different by adding a mixed effect for this value. The results show that
there are only some small differences in the standard error compared to the simple model. Similar to the model in Table 10.6, being part of the community is taken as the reference group (Community=1). The intercept therefore represents the predicted PAL scores for the people in the community, and the estimated coefficient for Community=0 indicates the difference between the predicted PAL for the people in the community group and the people in the non-community group. The coefficient of Time indicates that for every unit of time, there is an increase of 0.0005821 in the PAL for people in the community group. The estimated coefficient for the interaction term represents the difference in the slope for the two groups. In other words, the interaction term tells us that the two groups (community vs. non-community) show a significantly different change in PAL over a period of twelve weeks.

Table 10.7: Analysis using linear mixed effects modeling.

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Error</th>
<th>df</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.6933389</td>
<td>0.02388</td>
<td>44,230</td>
<td>70.901</td>
</tr>
<tr>
<td>Community=0</td>
<td>-0.0586920</td>
<td>0.02581</td>
<td>580</td>
<td>-2.273</td>
</tr>
<tr>
<td>Time</td>
<td>0.0005821</td>
<td>0.00015</td>
<td>44,230</td>
<td>3.791</td>
</tr>
<tr>
<td>Community=0 : Time</td>
<td>-0.0006525</td>
<td>0.00016</td>
<td>44,230</td>
<td>-3.931</td>
</tr>
</tbody>
</table>

The likelihood ratio test is often conducted to test the significance of predictor variables, i.e. to compare the fit of one model (with a reduced set of predictors variables) to the fit of another model (with a complete set of predictor variables). Here, we also use this test to see which model provides a better fit for data. Model 1 is based on a generalized linear regression (Table 10.6) and model 2 is based on a linear mixed effects model (Table 10.7). The latter includes all the variables of model 1, plus an additional mixed effect for the individuals’ intercepts. The results are shown in Table 10.8. The null hypothesis (stating that the between-subject variation in the intercept is equal to zero) is rejected, $\chi^2(1) = 17,882.63$, $p < .001$. This tells us that adding a random effect for the individuals to the model is a significant improvement, therefore the mixed effect model provides better fit for the data.

Table 10.8: Comparison of standard linear regression model with random intercept model.

<table>
<thead>
<tr>
<th>Model</th>
<th>df</th>
<th>logLik</th>
<th>Chi Squared</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>standard (1)</td>
<td>5</td>
<td>-15,712.890</td>
<td></td>
<td></td>
</tr>
<tr>
<td>advanced (2)</td>
<td>6</td>
<td>-6,771.574</td>
<td>17,882.63</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>
Chapter 10. Effect of community membership on impact of a health program

10.5 Discussion

The main research question that is investigated in this paper is whether an intervention aiming at increasing physical activity has a different effect on people that become member of an online community compared to people that choose not to participate in such a community.

Two statistical analyses were performed. In the first analysis, a significant linear regression model was found. Based on the adjusted $R^2$, we conclude that 44% of the variance in the PAL values is explained by this model. In the second analysis, a linear mixed model was fitted on the whole data set (eleven weeks), which shows that there is a significant difference between the increase in PAL of the two groups, even when a random factor for the (different value of the) start PAL is taken into account. It can thus be inferred from the results that on average people that participate in an online community at some time show a larger increase in activity level between the start and end of the program compared to people that will not participate in a community. This answers our question positively.

However, these findings do not yet answer the question why this is the case. First, note that people in the community group were not necessarily member of the community during the active part of the intervention. We can therefore only make statements about people willing to become member of a community, and we are not able to conclude that being member of a community causes the increase in PAL. Even if this would be the conclusion, there are still different social phenomena that could explain the effect. One hypothesis is that it could be caused by social contagion, i.e. the process of influencing others (sometimes unconsciously) via a network of social relations (Schoenewolf, 1990). Another possible explanation is social support, in the sense that community members help each other in performing physical activities (e.g., doing sports together) (Cohen and Syme, 1985). Yet another hypothesis is that social comparison is a driving factor, i.e. that people that choose to share their physical activity level online are stimulated by the achievements of others (Festinger, 1954). These questions require further investigation and provide directions for future research.

The visual representation in Figure 10.3 of the PAL during the period of the intervention shows – apart from a different slope for the two groups – also two other interesting aspects. First, a regular pattern of peaks and dips in both groups can be seen. Since each participant always starts his/her plan on a Monday, the data is aligned per weekday. Our explanation is that the dips correspond to weekends, in which people are on average less active. Second, we see that the PAL of the people that are part of the non-community group does not increase at all, even though they participate in a physical activity promotion program. There is no obvious explanation for this observation, but it seems that the intervention is not effective in increasing the PAL for the participants who do not join the community (during or after the analyzed period of the program). It is possible, however, that the activity levels would have decreased without the intervention. Therefore, a comparison with a control group of people who do not participate in the program at all should reveal whether the intervention is effective in maintaining the PAL.

10.6 Conclusion

The willingness to participate in an online community in a physical activity promotion program does make a difference in the effect of such a program. A data set of approximately 50,000 individuals was used to extract data that ensured a fair comparison between
participants that are willing to participate in an online community and participants that are not. From the set of approximately 5,000 individuals that opted in for the community (consisting of a collection of several smaller connected components), a number of components was selected based on specific inclusion criteria. Based on the characteristics of those sub-communities, similar individuals were found from the set of individuals who were not tied to any community.

The two data sets were analyzed and compared with each other. We were able to conclude that there is a difference in PAL, as the users in the community group are already more active at the start. This confirms findings from earlier work (Groenewegen et al., 2012). Also, we were able to conclude that the PAL of people that are willing to join a community shows an increase that is significantly greater compared to the other users. Since we balanced the data sets for possibly confounding factors like gender, time of the year and corporation, it is very likely that the fact that people are willing to become member of the community is the dominant factor that makes a difference for their increase in physical activity level. We can conclude that the willingness to participate in an online social network for sharing activity data is associated with an increase in physical activity. However, since we also observed that those people are already more physical active at the start, part of the effect might be caused by the fact that people that are willing to participate in such a community are different from others, e.g. more motivated. Further research is needed to see whether active participation in an online community contributes to the effectiveness of an intervention. Still, our findings are a valuable step towards answering the question “does online sharing of physical activity accelerate the impact of a health promotion program”.

In future work, we plan to distinguish actual participation in a community from non-participation. Further, we will use an existing computational model of social contagion (Bosse et al., 2009) to see whether this model can explain and predict the change. Also, it would be interesting to consider the effect of other factors on the physical activity level, such as the community size and structure. That way, research can further uncover phenomena that are at the basis of the beneficial effects of online social networks in health promotion programs.
References


for Adults From the American College of Sports Medicine and the American Heart Association”. In: Circulation 116.9, pages 1081–1093.


Abstract

Social processes play a key role in health behavior. Understanding the underlying mechanisms of such processes is important when designing health interventions with a social component. In this work, we apply a computational model of social contagion to a data set of 2,472 users of a physical activity promotion program. We compare this model’s predictions to the predictions of a simple linear model that has been derived by a regression analysis. The results show that the social contagion model performs better at describing the pattern seen in the empirical data than the linear model, indicating that some of the dynamics of the physical activity levels in the network can be explained by social contagion processes.
This chapter is based on:

Physical inactivity is a major worldwide concern, as it can lead to many long-term health risks (Conn et al., 2011; Eime et al., 2013). These risks can be reduced if an adult fulfills the requirement (according to recommendations of the WHO and other public health organizations) of at least 150 minutes of moderate or 75 minutes of vigorous intensity physical activity per week, or a combination of both (Garber et al., 2011; World Health Organization, 2010). An active lifestyle not only improves a person’s physical health, but it also has positive effects on mental health (Pate et al., 1995).

Maintaining a healthy lifestyle is a difficult task, even in this technological era in which we are surrounded by many types of gadgets that aim at supporting this. If used in innovative ways, eHealth and mHealth hold great potential to steer physical activity promotion programs in the right direction and let greater numbers of people benefit from it. However, this requires the right choices about the way in which technology is embedded in these programs. For example, simply using a wearable device alone will not suffice to achieve sustainable behavior change (Patel et al., 2015). To maintain new behavior for a longer period of time, other important ingredients are needed, e.g. evidence-based techniques such as goal setting and timely feedback, and a supportive social environment.

Social processes play a key role in health behavior. Several aspects of it are relevant for achieving a healthy lifestyle. It has been shown that people become more successful in maintaining a healthy lifestyle when they function within their social context (Wing and Jeffery, 1999; Zimmerman and Connor, 1989). In addition, the social environment enables people to compare their physical activity achievements with their peers or to seek social support from them. Within online social networks, this is commonly implemented via leader boards with achievements, building on the theory of social comparison (Suls and Wills, 1991). In (Klein et al., 2014), we investigated whether this principle can be used to explicitly change the influence of others by changing the visibility of the connections. For example, if a person is vulnerable to adopt a certain kind of negative behavior, more positive people in that person’s friendship network could be shown more prominently. Overall, in the context of health promotion programs, support for social processes can provide a leveraging mechanism to achieve and maintain a healthy lifestyle. Understanding these mechanisms is therefore important.

In this paper, we use a data set about health behavior in a social context to understand the underlying social processes. It is a continuation of earlier work on this subject (Groenewegen et al., 2012; Manzoor et al., 2016). In (Manzoor et al., 2016), a large data set of an online physical activity promotion program was used to compare the physical activity levels of people who (at some point in time) participated in an online social network with those who opted not to join the network. Ten network components were identified from the overall network and based on similar characteristics as the participants in the community data set, individuals were identified who were not part of the online community. One of the conclusions was that participants who chose to join the online community (either during the period of the collection of the analyzed data or later) had significantly higher activity levels and a higher increase in activity compared to participants who chose not to become part of the community.

In this work, we try to answer the question whether the changes in physical activity can be explained by social contagion. Social contagion can play a large role in shaping a certain kind of behavior in a social network (Christakis and Fowler, 2013). After a rigorous process
of cleaning and filtering the data, 2,472 community members were extracted for the analysis. The research question is addressed by comparing the activity data of the participants with two types of predictions: (1) based on a simple linear model that captures the effect of participating in the program for users who join the online community, and (2) based on a model of social contagion combined with the linear model. The social contagion model enables us to validate the social dynamics in terms of contagion of physical activity behavior and the linear part of the model explains the effects of the program. Mean absolute errors were calculated by comparing empirical data with the combined contagion model and the linear model. Finally, some statistical tests were conducted to show the significance of the difference between the two errors.

The remainder of this paper is structured as follows. First, Section 11.2 discusses the related work with respect to physical activity promotion programs, the role of social influence (social contagion) and the computational model of social contagion. Section 11.3 describes the methods and analyses. Section 11.4 describes the results of the analyses. Finally, Section 11.5 provides a discussion about the choices that we made during the analyses, a reflection on the results, and possible future directions.

### 11.2 Related work

Because a majority of the adults in the Western world does not meet the guidelines for physical activity, public health professionals are aiming at population-wide interventions. Since decades, the area of preventive medicine is investigating how people can be stimulated to be more physically active (Sallis and Owen, 1998). It has become clear that a multi-disciplinary approach is required, combining personal level mediators with concepts and perspectives from other fields, for example sociology and urban-planning (Abby C King et al., 2002). More recently, the smartphone has been discovered as tool for measuring and influencing physical activity (Bort-Roig et al., 2014). Many of these technology-mediated interventions use some kind of social influence. In an earlier review of 57 physical activity apps (Middelweerd et al., 2014), we found that 37 times some social support feature was included. In (Bort-Roig et al., 2014), social support networking was identified as one of the most useful strategies.

A specific appearance of social influence is the phenomenon of social contagion (Christakis and Fowler, 2013). It has been shown that people can influence each other via their social networks up to three degrees of distance. Although these claims have been criticized (Shalizi and Thomas, 2011), one could imagine that people transitively influence each other via social relations. In (Araújo and Treur, 2016; Bosse et al., 2015), a temporal-causal computational model is presented that describes how the mutual absorption of emotions in a social network affects the emotions of the individuals. This model was used as the basis for the study that is described in this paper. Our assumption is that physical activity behavior is influenced by internal states like motivation, attitudes and goals, and that those spread in a similar way as described in the model of emotion contagion. In earlier work, we used a similar approach to predict the contagion of physical activity in a small social network (Araújo, Tran, et al., 2015).

The model (Araújo and Treur, 2016) describes how internal state $q_A$ of person $A$ affects the internal states of other persons $B_i$. This process is determined by the strength by which the state is expressed ($\varepsilon_A$), the openness of the receiver ($\delta_B$) and the strength of the channel
between them ($\alpha_{AB}$). Together, these factors determine the \textit{connection weight} $\omega_{AB}$. Thus, the impact of the state of person $A$ on the state of person $B$ is:

$$\text{impact}_{AB}(t) = \omega_{AB}q_A$$  \hfill (11.1)

The aggregated impact $\text{aggimpact}_B(t)$ at time $t$ of the states $q_{A_i}$ of all connected persons on state $q_B$ is modeled as a scaled sum. From this it follows that $\text{aggimpact}_B(t)$ is calculated as a weighted average of all the impacts of the different connections of a person:

$$\text{aggimpact}_B(t) = \sum_{A_i \neq B} w_{AB}q_{A_i}(t)$$  \hfill (11.2)

with $w_{AB}$ chosen in such a way that it is proportional to $\omega_{AB}$ and the sum of all weights is 1. The new state for each person in the network is calculated by integrating some factor $\eta$ of the aggregated impact:

$$\text{contagion\_effect}(t) = \eta_A[\text{aggimpact}_B(t) - q_B(t)]$$  \hfill (11.3)

$$q_B(t + \Delta t) = q_B(t) + \text{contagion\_effect}(t)\Delta t$$  \hfill (11.4)

For a more detailed description of the model, see (Araújo and Treur, 2016).

For the purpose of this study, we assumed that all people have the same expressiveness and openness, and that all connections were of the same strength. This was done out of necessity, as our data set does not contain specific information about these factors. The model’s parameters for openness, expressiveness and channel strength were thus set to a default value of 0.5.

\section{11.3 Methods}

This section describes how the data was collected and preprocessed, as well as what types of analyses were run.

\subsection{11.3.1 Data collection}

The data originates from a physical activity promotion program in which participants are asked to wear an activity monitor that measures physical activity level (PAL) using an accelerometer. Based on the activity data that is repeatedly uploaded by the participants, the program stimulates them towards a more active lifestyle by gradually increasing the weekly activity targets over a 12-week activity plan. The baseline for this activity plan is established in an initial assessment week. After completing a plan, participants can choose to take another 12-week activity plan or decide to remain at the level of their last completed plan.

After the initial assessment week, participants also get access to a dashboard with information about energy expenditure (calories burnt) and their achievements relative to a weekly goal. The program provides an opt-in online community that allows participants to establish connections and to compare achievements. Each participant in the community will see how their achievements rank compared to other participants with whom they are
Chapter 11. Explaining changes in physical activity through social contagion

connected. Community participants see the ranking within their own network each time they upload data from their activity monitor.

The network structure and some social network analyses are discussed in (Araújo, Klein, et al., 2016), showing the scale-free structure on the distribution of the degrees of the nodes and some homophily characteristics of the edges.

11.3.2 Data preprocessing

The original data set contains data for 52,788 users. Since the aim of this paper is to demonstrate the influence of social contagion on people’s physical activity levels, we are only interested in the 5,041 users who opted in for the online community of the program.

First, any participant that joined the program for testing purposes or users with missing information such as gender or body mass index (BMI) were removed from the data set, as well as participants that didn’t have a start date for their first plan. The resulting data set contains participants for whom valid physical activity data is available. The network was further pruned by removing connections that were initiated by one participant, but never confirmed by the other participant.

As the online community feature was not part of the program until April 28th 2010, all data before that date was disregarded. Community data was available until August 6th 2010, but the PAL data was incomplete for the last couple of days. This can be explained by the fact that some users did not upload their data for those days yet. Therefore, only the data up to July 28th 2010 was considered, resulting in a data selection that spanned a period of 91 days.

Within this period of 91 days, only active and connected participants were considered for the current analysis. In other words, any users who entered the program, but did not join the online community, or users that dropped out of the program before this period started, were removed from the data set. This data cleaning process leaves us with 2,472 relevant nodes in the period between April 28th 2010 and July 28th 2010.

Although the primary unit of physical activity in the data set is the PAL, users see percentages of their goal achieved rather than the PAL itself on their online dashboard. The ranking with connected users is also based on this relative performance. Therefore, our analyses are also based on the ratios of goals achieved, i.e. the current PAL divided over the target PAL.

11.3.3 Model simulations

Previous work has shown that the combination of participation in the program and willingness to join the online community is associated with a small but significant average increase in PAL (Manzoor et al., 2016). The objective of the current work was to demonstrate whether the dynamics of users’ physical activity levels can be (partially) explained by social contagion. Therefore, we compared the predictive performance of two different models: (1) a simple linear model, that describes the effect of the program on (future) members of the community; and (2) a combined model, that captures the social contagion process and incorporates the linear increase as well.

Scenario 1: Simple linear model

The simple linear model describes the effect of the physical activity promotion program on the users’ physical activity levels. Previous analyses have shown that this effect is an average
PAL increase of 0.0005821 per day (Manzoor et al., 2016). These analyses were based on a subset of users from the same data set, with all users being in their first plan and eventually member of the community (either during the period of the collection of the analyzed data or later). The increase in PAL translates to an increase in energy expenditure of 1.05 kCal for an average male with a basal metabolic rate (BMR) of 1800 kCal/day (Mifflin et al., 1990).

To translate this increase in PAL to the unit predicted by the model (i.e., the goal achieved), the simple linear model adds a daily increase of 0.0005821 divided by the current target PAL to the user’s goal achieved, as shown in Equation 11.5 and Equation 11.6.

\[
\text{linear_effect}(t) = \frac{0.0005821}{\text{target_pal}(t)}
\]  (11.5)

\[
\text{goal_achieved}(t + \Delta t) = \text{goal_achieved}(t) + \text{linear_effect}(t)
\]  (11.6)

**Scenario 2: Combined social contagion model**

The combined social contagion model describes the linear increase in PAL as well, but combines it with the model of social contagion that captures the dynamics between the nodes in the network, as summarized in Equation 11.7, where \(\text{contagion_effect}(t)\) denotes the social contagion effect as described in Section 11.2, Equation 11.3. In this case, the state \(q\) represents the percentage of goal achieved. By enriching the social contagion model with the daily increase in PAL (as in the simple linear model), we account for the demonstrated stimulating effect of the program and the community, and thereby nullify a possible disadvantage on the social contagion model.

\[
\text{goal_achieved}(t + \Delta t) = \text{goal_achieved}(t) + \text{contagion_effect}(t) + \text{linear_effect}(t)
\]  (11.7)

As mentioned in Section 11.3.2, the analyses were based on the predictions of the goal achieved, i.e. the proportion of the target PAL achieved by the user, rather than the user’s current PAL. Additionally, the model predictions were done for users in their first plan. Of the 2,472 relevant users identified in Section 11.3.2 1,939 were participating in their first plan for at least part of the time period under consideration. The reason behind this choice is that users in their first plan are most comparable to the general population: they have just entered the program, and therefore have no prior knowledge of or experience with the plans or other parts of the intervention. Also, it is likely that people in their first plan have the highest adherence rates and interact more with the program, which makes them a more interesting population as well. However, users who have not yet started or already completed their first plan can still influence users in their first plan through social contagion. Therefore, they are considered by the social contagion model, but only as input of the contagion process towards the users under consideration (i.e., users in their first plan).

To run the models, the initial values have to be determined. For all users for whom a target PAL is not available (i.e., users who are in their assessment week and have yet to start their first plan), the initial goal achieved value was based on the average PAL of their assessment week and their first target PAL. For all users with a target PAL, the initial goal achieved was calculated by dividing the average PAL for one week before the start date of the simulations (i.e., April 28th 2010) by the current target PAL. If for some reason, no data
was available for that week, the initial goal achieved was based on the average PAL in the month prior to the start date of the simulations. This decision sequence is summarized in Figure 11.1.

- **If** user has not started first plan yet:
  - Average PAL in assessment week and first target PAL of first plan.
- **Else** (i.e., if user has already started and/or completed first plan):
  - **If** data is available in 7 days before start date of simulation:
    * Average PAL in 7 days before simulation start date and target PAL of simulation start date.
  - **Else** (i.e., if no data is available in 7 days before start date of simulation):
    * Average PAL in 30 days before simulation start date and target PAL of simulation start date.

Figure 11.1: Data used to calculate initial goal achieved values for different cases.

In the social contagion model, we used the initial goal achieved values of the simulated nodes as described above, and the empirical data from the surrounding nodes as input to the contagion process. This choice was motivated by the fact that we were only interested in simulating the effect of the behavior of users on users in their first plan, rather than simulating the behavior of those other users as well.

**11.3.4 Analyses**
To evaluate the accuracy of the two models, we first calculated their average predictions for the approximately 1,939 users in their first plan in the data set, as well as the average goal achieved values based on the empirical data. Based on these values, we tested whether there is a significant difference in the magnitude of the errors of the two models with a Mann Whitney U test. In addition, we determined the correlations of both models’ predictions to the empirical data by means of Mann Kendall tests.

**11.4 Results**
As explained in Section 11.3.2, after thorough preprocessing of the data, 2,472 relevant users remained in the period between April 28th 2010 and July 28th 2010. Figure 11.2 shows their empirical data over the 91 days in the data set.

Following the procedures described in Section 11.3.3, the two models were run on the initial data. Figure 11.3 provides an impression of the predicted goal achieved values for the 1,939 users in their first plan by the two models. The simulation of the linear model shows a steady increase in the goal achieved. The combined model shows the effect of the contagion between the users, in combination with the steady increase. Any interruptions of the lines in either plot are caused by users entering the program or community, or by users dropping out of the program.

After averaging the model predictions, as well as the empirical data, for all users in their first plan per day, the graph in Figure 11.4 was obtained. It shows the average predictions
Figure 11.2: Empirical data of 2,472 users between April 28th 2010 and July 28th 2010.

Figure 11.3: Predictions of the simple linear model (left) and the combined model (right). The sharp troughs in the empirical data mark the Sundays, when physical activity levels on average are substantially lower.

Figure 11.4 already gives the impression that the combined model is much closer to the empirical data than the linear model. Indeed, the mean absolute error (MAE) of the linear model is 0.02212, whereas the mean absolute error of the combined model is 0.01321. A Mann-Whitney U test shows that the difference between the errors of the two models is significant, \( p < .001 \).

Besides comparing the size of the errors, we also investigated whether the predicted lines were correlated with the empirical data. A Mann-Kendall test shows that the linear model is significantly correlated with the empirical data, although negatively \( (\tau = -0.46227, \)
Figure 11.4: Average predictions of the two models (green: linear, blue: combined), and the empirical data (red).

\( p < .001 \). The combined model is also significantly correlated, but in this case positively (\( \tau = 0.53895, p < .001 \)).

Table 11.1: Model evaluations.

<table>
<thead>
<tr>
<th></th>
<th>Absolute error</th>
<th>Kendall’s correlation test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>St. dev.</td>
</tr>
<tr>
<td>Linear model</td>
<td>0.02212</td>
<td>0.01378</td>
</tr>
<tr>
<td>Combined model</td>
<td>0.01321</td>
<td>0.00855</td>
</tr>
</tbody>
</table>

11.5 Conclusions

The results described in Section 11.4 show that the combined model, which integrates the social contagion model with a steady linear increase in PAL, is indeed better able to capture the dynamics of the physical activity levels in our data set than the linear model. Its predictions show a significant positive correlation with the empirical data. Additionally, the errors of the combined model’s predictions are significantly smaller than those of the linear model.

One of the main strengths of this work is its foundation on a large set of empirical
data covering several months. Careful and extensive preprocessing of the empirical data was conducted to ensure data that is sensible for the simulated models. For example, we dynamically removed connections to users who practically dropped out of the program (but were still in the system), to prevent their (missing) data from affecting the results.

Another strength of our work is that we compared the performance of the model we were mainly interested in (i.e., the combined model of social contagion and linear increase) to an informed linear model. That way, we do not impose a disadvantage on the baseline model, thus increasing the chances of superiority of our more complex model. However, it is interesting to see that the empirical data shows a development that is actually opposite to the direction of the linear increase model. There are a few possible explanations for this observation. An evident first possible reason is that the current data analysis is based on actual community members in their first plan, whereas the linear model is based on participants who become member of the online community at some point in time, but they might have still been unconnected during their first plan. Therefore, even though the linear model is informed by the data, it is actually not as well informed as it could be. Plans for future research include replicating the analyses in (Manzoor et al., 2016) with actual community members, and investigating whether that leads to a different linear model and thereby possibly different performances of the two models applied in the current analyses. Another possible explanation could be that the linear increase was found after aligning the data by the day in the program rather than the calendar date. Possibly, we would see the same average increase if we aligned our selection of the data set in the same way. The pattern in the current data set is then caused by users in different phases of the first plan entering and leaving the program over time (e.g., because their first plan is finished halfway the period that we selected). A third possible reason for the unexpected observation is that the linear model describes an increase in PAL, whereas it is transformed and applied to the progress towards the target PAL in this work.

One of the limitations of this work is its restricted generalizability. As all analyses were based on data collected in the context of a physical activity promotion program (see also Section 11.3.1), the results cannot directly be transferred to the general population. However, by choosing to focus on people who are exposed to the program for the first time, we have tried to minimize that discrepancy.

Another limitation is that the social contagion model only considers the online community as the network through which the behavior spreads, although contagion also takes place on different levels and in different contexts. Additionally, we did not take into account whose data is actually shown on the user’s dashboard: all connections were treated equally, whereas the performance of friends may not be shown on the dashboard when the difference was too big (e.g., more than 10 position difference). Future work could reveal whether limiting the contagion model to only the connected users who are visible on the dashboard improves the performance of the model. A further limitation is that we used default values of 0.5 for the parameters (for expressiveness, channel strength and openness) in the combined model. In future work, we could investigate whether using calibrated values would yield better results. It is also possible to experiment with models that incorporate the principle of non-linearity in behavior change, e.g. by exploiting thresholds for effects (Giabbanelli et al., 2012).

Up to our knowledge, we present the first analysis of the ability of a computational model of social contagion to capture the pattern of physical activity levels in a community
Chapter 11. Explaining changes in physical activity through social contagion

over time. In order to do so, we compared model predictions of such a contagion model (enriched with an expected linear increase) and of a simple linear model to a dataset of 2,472 interconnected users. The results show that the enriched social contagion model performs better at describing the pattern seen in the empirical data than the linear model, indicating that some of the dynamics of the physical activity levels in the network can be explained by social contagion processes. This is vital information for designers of health interventions with a social component, as such models can then be used to maximize the benefits of social influence processes.
References


King, Abby C, Dan Stokols, Emily Talen, Glenn S Brassington, and Richard Killingsworth (2002). “Theoretical approaches to the promotion of physical activity: forging a trans-
disciplinary paradigm”. In: American journal of preventive medicine 23.2, pages 15–25.
Shalizi, Cosma Rohilla and Andrew C. Thomas (2011). “Homophily and contagion are generically confounded in observational social network studies”. In: Sociological methods & research 40.2, pages 211–239.
12. Effectiveness of upward and downward social comparison of physical activity

Abstract

It has been established that social processes play an important role in achieving and maintaining a healthy lifestyle, but there are still gaps in the knowledge on how to apply such processes in behavior change interventions. One of these mechanisms is social comparison, i.e. the tendency to self-evaluate by comparing oneself to others. Social comparison can be either downward or upward, depending on whether individuals compare themselves to a target that performs worse or better. Depending on personal preferences, the variants can have beneficial or adverse effects. In this paper, we present the results of an experiment where participants (who indicated to prefer either upward comparison or downward comparison) were sequentially shown both directions of social comparison, in order to influence their physical activity levels. The results show that presenting users with the type of social comparison they do not prefer may indeed be counter-effective. Therefore, it is important to take this risk into account when designing physical activity promotion programs with social comparison features.
This chapter appeared as:

12.1 Introduction

Physical inactivity is a major public health issue. Evidence shows that it increases the risk of many health problems, including non-communicable diseases such as type-2 diabetes, cardiovascular diseases, cancer and mental illnesses (Lee et al., 2012). Consequently, physical inactivity is the fourth leading cause of death globally (Kohl et al., 2012). Vice versa, engaging in sufficient levels of physical activity has been associated with effects on both physical and mental health (Conn et al., 2011; Eime et al., 2013; Reiner et al., 2013). However, a large part of the Western population does not meet the global recommendations of being moderately to vigorously active for at least 30 minutes per day on at least five days per week (Hallal et al., 2012). Therefore, healthy lifestyle promotion programs and behavior change interventions are a priority in most Western countries, in order to increase global physical activity levels.

It has been established that social processes play an important role in achieving and maintaining a healthy lifestyle (Zimmerman and Connor, 1989). Several mechanisms that underlie these social influences have been identified, such as priming, social norms, behavior modeling, social facilitation and social support (Bandura, 1998; Cheng et al., 2014; McNeill et al., 2006), but there are still gaps in the knowledge on how to apply such processes in human behavior change interventions. Another mechanism that plays a part in behavior evaluation and behavior change is social comparison (Buunk et al., 2013; Festinger, 1954), which is often applied in physical activity promotion apps as a form of providing feedback to the user (Middelweerd et al., 2014). Social comparison exists in two variants: downward social comparison and upward social comparison (Festinger, 1954), depending on whether the target (with whom one compares oneself) performs worse (i.e., downward) or better (i.e., upward) than the individual. Both variants can be effective and encouraging, for instance by boosting one’s self-view or by motivating improvement, but also counter-effective and discouraging, for instance by advocating inferior standards or by threatening the self-view (Corcoran et al., 2011). Therefore, it is important to carefully design interventions that incorporate a social comparison component.

In this paper, we describe an experiment to test whether social comparison of physical activity via an online intervention leads to a measurable effect on the behavior of participants. More specifically, we investigate whether the direction of the presented social comparison (upward or downward) indeed yields two-sided effects on people’s physical activity levels, depending on the users’ indicated preference for one of these two variants. At the same time, the results of the experiment will indicate whether people are able to assess their own preference truthfully and effectively, according to the effects on their behavior.

The remainder of the paper is organized as follows: Section 12.2 describes some background on social comparison theory. In Section 12.3, we describe the methods used to gather and analyze the data. The results are presented in Section 12.4, and reflected upon in Section 12.5. Finally, Section 12.6 closes the paper with a conclusion.

12.2 Background

Social comparison is defined as the tendency to evaluate oneself through comparison to others, which is an important source of competitive behavior to self-improve (Garcia et al., 2013). In both upward and downward social comparison, people aim to attain or maintain a higher level of performance than others (Festinger, 1954). The desire to achieve or keep
such a superior position is called a ‘comparison concern’ (Garcia et al., 2013). The model presented in (Garcia et al., 2013) shows that two sets of factors can encourage competitive behavior by raising such comparison concerns: individual and situational factors.

Individual factors are those that may vary greatly between individuals, even if they find themselves in comparable situations. The three most important individual factors are relevance, similarity and closeness (Goethals and Darley, 1977). The more relevant a particular dimension of performance (e.g., income, study results or sports achievements) is to an individual, the stronger their comparison concerns will be (Hoffman et al., 1954). Likewise, the more similar a target with whom one compares oneself is, the stronger the effect of the social comparison will be (Kilduff et al., 2010). Finally, the comparison concerns are also stronger when the individual and the target have a close personal relationship than when they don’t know each other (well) (Tesser and Smith, 1980).

Situational factors are those that concern an individual’s perception of the surrounding social environment, by which means they can yield a more general effect on similarly situated individuals (Garcia et al., 2013). Several situational factors can contribute to one’s comparison concerns. For example, incentive structures (i.e., the incentives associated with the comparison) can encourage competitiveness when higher values are expected in case of better (relative) performance. Another factor is the proximity to a standard, i.e. whether an individual is close to the number-one position (or some other meaningful performance metric). The closer to such a standard, the stronger the comparison concerns are. The number of competitors also influences comparison concerns: the fewer competitors, the stronger the competitive behavior of individuals. A final example of a situational factor in social comparison is social category fault lines: when comparing to targets from other social categories (based on gender, nationality, etc.), the comparison concerns are stronger than when comparing within such categories.

As mentioned in the Introduction, social comparison can involve a target (with whom one compares oneself) that performs better (i.e., upward) or worse (i.e., downward) than the individual. The two variants address different underlying motivational processes, implying different benefits and drawbacks. The main positive effect of downward social comparison is that comparing to a lower-performing target can boost the individual’s self-esteem and subjective well-being (Wills, 1981). On the other hand, downward comparison could also result in relatively low goals, since it doesn’t challenge an individual to try harder to minimize the discrepancy with someone else’s performance. The main benefit of upward social comparison is exactly that: it motivates people to self-improve in order to approximate better performing others (Lockwood and Kunda, 1997). In addition, a higher-performing target could provide information and serve as a role model, which allows an individual to learn how to perform better (Maddux, 1995). However, if one deems the performance of superior others to be unattainable, upward social comparison could have a discouraging and deteriorating effect as well (Lockwood and Kunda, 1997).

The preference for either upward or downward social comparison can originate from the motivation behind the social comparison. Although self-evaluation can be achieved with both upward and downward comparison, individuals who strive for self-enhancement (i.e., boosting their self-view) are more inclined towards downward comparison (Suls et al., 2002; Wills, 1981). Similarly, people who pursue self-improvement will tend to opt for upward comparison (Suls et al., 2002). Research has shown that men more often engage in upward social comparison, whereas women tend to compare in a downward direction.
Based on the theory about social comparison described above, an experiment has been set up. The details of the setup and the data collection are described below.

12.3 Methods

The social comparison intervention was implemented as a dynamic web application, developed by (Werkhooven, 2015), in which users could see their own activity data and a comparison with their friend’s data. The web application was accessible through a web browser, and the information displayed in the web application was therefore available to the users at any time and at any location. The activity data was registered using a wireless activity monitor, the Fitbit One (Fitbit Inc., no date), which tracks steps, floors climbed, distance, calories burned and active minutes. For this experiment, the number of steps was taken as main measure for physical activity.

The web application first shows an overview of all participating users, with a name and a profile picture. Then, the users can log in to their personal page by clicking on their own picture or name. The personal pages are secured with a user password, in order to protect the privacy-sensitive information, but also to prevent users from looking at each other’s personal pages.

The personal page in the web application consists of three parts. First, a profile picture and the name of the user are shown at the top of the page. Second, the page contains a graph where the user’s own activity data is visualized. The graph shows the number of steps for the past 24 days in a bar chart. Third, the personal page shows a comparison of the user’s activity data with the activity data of two friends. This comparison is established through a line chart with three lines, a thick solid blue line for the user’s own activity data and thinner solid red and yellow lines for the two friends. The area below the lines was colored in the same color as the lines. Figure 12.1 shows an example of a personal page on a mobile phone.

The choice for which other users’ activity data to show to the user depends on the user’s preference (upward or downward) and the current setting of the system (in line with the user preference or the opposite of the user preference). When the web application employs downward social comparison, the average number of steps of the friends is lower than the user’s average number of steps. Similarly, when the web application shows upward social comparison, the friends have a higher average number of steps than the user. This implies that the physical activity levels of friends were higher or lower for the overall period, but that activity levels on a specific day could be opposite. In other words, the direction of the social comparison is relatively subtle. Another consequence of this setup was that the friends that were shown in the graphs could change per day.

12.3.2 Experimental setup

The effect of the social comparison intervention was tested in a small user study. Twenty members of a Dutch amateur soccer club participated. The participants were between 21 and 26 years old, with an average age of 24.6 years (sd = 1.2 years). The participants were all male, and the vast majority of the group played soccer on the same team. Although the
Figure 12.1: Example of personal page with the user’s own activity data (bar chart) and the comparison to two friends that have a higher average level of physical activity (line chart). The blue line represents the user’s own activity data, and the red and yellow lines represent the activity data of two friends.

Figure 12.2: Example of downward social comparison graph. Again, the blue line represents the user’s own activity data.

Homogeneity of the group evidently limits the possible conclusions that can be drawn from this experiment (see Discussion in Section 12.5), this was a deliberate choice in order to reduce the differences in individual and situational factors (see Background in Section 12.2).
as much as possible.

All participants were asked to use an activity monitor, the Fitbit One™, for a period of four weeks. This four-week experiment period was preceded by a test phase of five days, in which the participants could get used to wearing the device and monitoring their own physical activity. That way, the Hawthorne effect (i.e., the effect of knowing that one is being observed (Parsons, 1974; Sommer, 1968)) and the novelty effect (i.e., the effect of using a new gadget that stimulates physical activity) could hopefully be diminished. The four-week experiment period was divided into two parts: in the first two weeks, ten randomly chosen participants received upward social comparison and the other ten participants received downward social comparison, and vice versa in the last two weeks. This implies that our experiment adopts a partial within-subjects approach, allowing us to test the effect of both directions on each individual participant. During the four weeks that the participants received either one of the forms of social comparison, they were asked to view their personal page on a daily basis. Table 12.1 shows which group received which direction of social comparison in which period.

Table 12.1: Direction of social comparison per group and period.

<table>
<thead>
<tr>
<th></th>
<th>Period 1 (week 1–2)</th>
<th>Period 2 (week 3–4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>Upward</td>
<td>Downward</td>
</tr>
<tr>
<td>Group 2</td>
<td>Downward</td>
<td>Upward</td>
</tr>
</tbody>
</table>

At the end of the four weeks, the participants were asked to fill out a short questionnaire regarding their social comparison preferences. The questionnaire was distributed to the participants after the experimental period, in order to limit their knowledge about the aim of the experiment. The questionnaire consisted of five questions, as shown in Figure 12.4.
Chapter 12. Effectiveness of social comparison on physical activity

The participant’s indicated preference was determined based on question Q2. Using the information about the preference and the setup shown in Table 12.1, it could be determined when the participants were shown social comparison according to their own preference (period 1 or period 2), and when they were shown social comparison opposed to their preference.

Q1. Did you enjoy to see your physical activity compared to other users? (Five-point Likert scale from ‘not at all’ to ‘very much’.)
Q2. Do you prefer to compare your physical activity to individuals that perform worse or individuals that perform better? (Dichotomous: ‘individuals that perform worse’ or ‘individuals that perform better’.)
Q3. How often do you try to figure out what people think who encounter the same problems as you? (Five-point Likert scale from ‘almost never’ to ‘almost always’.)
Q4. How often does it make you feel good when coworkers or classmates perform worse than you? (Five-point Likert scale from ‘almost never’ to ‘almost always’.)
Q5. How often does it make you feel challenged when coworkers or classmates perform better than you? (Five-point Likert scale from ‘almost never’ to ‘almost always’.)

Figure 12.4: Questionnaire about social comparison preferences.

12.3.3 Analysis

The experimental setup described in the previous two subsections combines three different dimension. First, the participants are exposed to either upward or downward social comparison on their personal page. Second, the participants are exposed to the direction of social comparison that is either in line with their preference or opposed to their preference. Third, the participants receive a certain direction of social comparison in the first two weeks of the experiment (period 1) or in the last two weeks (period 2). This setup allows for analysis of the data from different angles.

In accordance with the social comparison theory, we expect that participants who receive social comparison in line with their preference will show an increase in physical activity, whereas showing social comparison opposed to their preference will have adverse effects. Therefore, the following hypotheses were formulated:

H1. Participants who prefer upward social comparison and receive upward social comparison will show an increase in physical activity.
H2. Participants who prefer upward social comparison but receive downward social comparison will show a decrease in physical activity.
H3. Participants who prefer downward social comparison and receive downward social comparison will show an increase in physical activity.
H4. Participants who prefer downward social comparison but receive upward social comparison will show a decrease in physical activity.
H5. Participants who receive social comparison in line with their preference will show an increase in physical activity, and participants who receive
social comparison opposite to their preference will show a decrease in physical activity.

The effect of each social comparison condition is investigated with a Mann-Kendall test, a non-parametric test for statistical dependence. Kendall’s $\tau$ (tau) indicates to what extent the orderings of the data are similar when ranked by each of the variables.

12.4 Results

In this section, the results of the experiment are presented. In Section 12.4.1, the participants’ preferred directions of social comparison are summarized. Then, Section 12.4.2 shows a first glance at the results. The following subsections present the data in more detail: in Section 12.4.3 to Section 12.4.7, we show the results associated with hypothesis H1 to hypothesis H5, respectively.

12.4.1 Social comparison preferences

The experiment described in the previous section was performed with 20 participants. Based on the questionnaire depicted in Figure 12.4, in particular question Q2, it was determined that fifteen participants preferred upward comparison and five participants preferred downward comparison.

12.4.2 Exploratory data analysis

The first question that comes to mind when designing and executing such a relatively small-scale experiment, is whether the sample size and duration are substantial enough to see effects of the social comparison intervention on the physical activity behavior. Figure 12.5 shows the average number of steps of the nine participants who received the direction of social comparison that they did not prefer in the first two weeks and the direction that they preferred in the last two weeks. It is clear to see that the daily step counts decrease substantially in the first two weeks, and after switching to their preferred direction of social comparison, the daily step counts start to increase again. This suggests that even though the experiment is relatively small, the results provide sufficient basis for further analysis.

In the next subsections, we present the data gathered in the experiment in more detail, according to the hypotheses introduced in Section 12.3.3.

12.4.3 Hypothesis H1: Preferred upward, presented upward social comparison

This section presents the data of the participants who preferred and received upward social comparison. Figure 12.6 shows the average number of steps of these 15 participants over a period of two weeks.

The Mann-Kendall test on this data yields that $\tau = .385$, which indicates a moderate positive rank correlation. However, the rank correlation is just under significant, at a $p$-value of $p = .0617$. The results are summarized in Table 12.2.

12.4.4 Hypothesis H2: Preferred upward, presented downward social comparison

This section shows the data of the participants who preferred upward social comparison but received downward social comparison. Figure 12.7 shows the average number of steps of these 15 participants over a period of two weeks.
Chapter 12. Effectiveness of social comparison on physical activity

Figure 12.5: Average number of steps in four-week period (with linear trend line), for participants who received social comparison opposite to their preference in the first two weeks and social comparison in line with their preference in the last two weeks.

Figure 12.6: Average number of steps in two-week period, for participants who preferred and received upward comparison.

The Mann-Kendall test on this data yields that $\tau = -0.538$, which indicates a considerable negative rank correlation. The rank correlation is also significant at a $p$-value of $p = 0.0067$. The results are again summarized in Table 12.2.

12.4.5 Hypothesis H3: Preferred downward, presented downward social comparison

This section shows the effect of downward social comparison for participants who also preferred this direction. Figure 12.8 shows the average number of steps of these 5 participants over a period of two weeks.

The Mann-Kendall test on this data yields that $\tau = -0.297$, which indicates a fair negative rank correlation. However, the rank correlation is not significant at a $p$-value of $p = 0.157$. Again, the results are summarized in Table 12.2.
Figure 12.7: Average number of steps in two-week period, for participants who preferred upward comparison and received downward comparison.

Figure 12.8: Average number of steps in two-week period, for participants who preferred and received downward comparison.

12.4.6 Hypothesis H4: Preferred downward, presented upward social comparison
This section presents the effect of showing downward social comparison for participants who preferred to receive upward social comparison. Figure 12.9 shows the average number of steps of these 5 participants over a period of two weeks.

The Mann-Kendall test on this data yields that $\tau = -0.560$, which indicates a considerable negative rank correlation. The rank correlation is also significant at a $p$-value of $p = 0.0046$. The results are again summarized in Table 12.2.

12.4.7 Hypothesis H5: Preferred vs. non-preferred social comparison
In this section, the data is reorganized to two groups: participants who receive social comparison in line with their preference ("preferred") and participants who receive social comparison opposite to their preference ("non-preferred"). Note that all participants are
represented in both groups, as everyone received both directions of the social comparison in either the first two weeks or the last two weeks.

Figure 12.10 shows the average number of steps of the two weeks when the participants received their preferred social comparison, and Figure 12.11 shows the period when they received the opposite direction. Figure 12.10 shows a fair positive rank correlation ($\tau = .253$), which is not significant ($p = .233$), and Figure 12.11 shows a strong negative rank correlation ($\tau = .648$) that is also significant ($p < .001$).
The results presented in Section 12.4 show that even though the results are not always significant, several interesting trends are visible. Except for the findings associated with hypothesis H3, the direction of Kendall’s \( \tau \) is always in line with the hypothesized results. Moreover, on all occasions when the participants are presented the version of social comparison they do not prefer, there is a moderate to strong negative rank correlation, which is also strongly significant. This indicates that user preferences are indeed important: if not to enhance the motivational effects of social comparison, then at least to avoid the adverse effects of showing users social comparison that discourages them. As social comparison features implemented in physical activity promotion programs frequently show some form of upward comparison (e.g., leaderboards, high scores), it is important to acknowledge the risk of such a representation on the performance of people who prefer the opposite direction of comparison. Often, when physical activity platforms incorporate a ranking among friends, the ranking shows other users with similar activity levels (e.g. three places ahead and three places behind), such as on the Fitbit dashboard webpage (Inc., 2016). Sometimes, the top performers are shown as well (i.e., upward comparison), as for example in the Human app (Human.co, 2016). When showing a certain type of social comparison by default that is
opposite to a user’s preference, such features could inadvertently lead to counter-effective outcomes.

When looking at the results in more detail, it can be seen that the findings show potential for stronger conclusions in bigger follow-up experiments. For example, after discarding the last day of the data associated with hypothesis H1 (see Figure 12.6 in Section 12.4.3), which was an usually “lazy” Sunday (note that Sundays usually have low average activity levels in the graphs shown above), the trend is immediately strong ($\tau = .590$) and significant ($p = .0043$). Similarly, when discarding the last day of the data associated with hypothesis H5a (see Figure 12.10 in Section 12.4.7), the trend becomes more considerable ($\tau = .424$) and closer to significance ($p = .063$).

When considering the significance of the results, the relative small sample size should be taken into account. After all, the results of hypotheses H1 and H2 were based on the data of only fifteen participants, and the results of hypotheses H3 and H4 even on only five participants. It is striking that still significant effects could be found based on such a small sample size. The small sample size could also explain some of the unexpected results. For example, this group of participants who preferred downward social comparison included a participant with highly varying step counts (e.g., $< 1,000$ and $> 34,000$), which could have distorted the group averages.

Indeed, the findings ask for a larger experiment to confirm the results and conclusions of the current study. Also, it would be interesting to see whether the results extend beyond the homogeneous participant population of this study, such as to women, non-exercisers, other age groups and users who do not know each other. Since the participants in this study were members of an amateur soccer club, it can be expected that they are relatively active and healthy. Therefore, it would be especially interesting to see whether these results transfer to a population of inactive people, as they should be the target group of physical activity promotion programs, and because they might experience social comparison differently.

In addition, the current study only investigated the effect of one particular implementation of social comparison. It would be interesting to explore other visualizations of social comparison (e.g., rankings, sorted charts), to see whether the specific representation influences the effect on the users. Another limitation of this current experiment is that the participants’ preference is based on one dichotomous question only. One can imagine that some people do not appreciate social comparison at all, but this was not an option in the current setup. Although the results indicate that people are reasonably able to assess their own preference, further research could investigate other (possibly implicit) ways to determine the user’s preferences. For example, instead of querying it directly, one could try to derive the preferred direction from a (personality) questionnaire, or from the user’s behavior when exposed to the two different variants. This could be helpful in particular for people who do not have a strong personal preference for one of the two options.

12.6 Conclusion

Social comparison features implemented in most physical activity promotion programs often show some form of upward comparison: e.g., leaderboards and high scores. However, the results of this experiment show that presenting users with such types of social comparison may have adverse effects for people who indicate that they prefer downward social comparison. Even though based on a small sample size, our results indicate that the adverse
effects are stronger and more significant than the positive effects. Therefore, it is important that intervention designers take this risk in account when developing a physical activity promotion program with social comparison features.

Acknowledgments
The authors would like to thank Kevin van Werkhooven for the development of the social comparison intervention and his contribution to the setup and execution of the experiment described in this paper.
References


# Part Five

## 13 Description of the development and content of Active2Gether

13.1 Introduction  
13.2 Methods  
13.3 Development of the Active2Gether intervention  
13.4 Summary of the Active2Gether intervention  
13.5 Discussion

## 14 Description of the technical design of Active2Gether

14.1 Introduction  
14.2 System description  
14.3 Discussion  
14.4 Conclusions and future work

## 15 Exploring use and effects of Active2Gether

15.1 Background  
15.2 Methods  
15.3 Results  
15.4 Discussion  
15.5 Conclusion

## 16 Evaluation of user appreciation and adherence of Active2Gether

16.1 Introduction  
16.2 Active2Gether system  
16.3 Methods  
16.4 Results  
16.5 Discussion and conclusions
Abstract

The Active2Gether intervention is an app-based intervention designed to help and encourage young adults to become and remain physically active by means of personalized, real-time activity tracking and context-specific feedback. The aim of this paper is to describe the development and the content of the Active2Gether intervention for physical activity (PA) promotion.

A systematic and stepwise approach was used to develop the Active2Gether intervention. This included formulating objectives and a theoretical framework, selecting behavior change techniques, specifying the tailoring, pilot testing and describing an evaluation protocol. Five steps were undertaken to develop the Active2Gether intervention: 1) definition of the intervention objectives, 2) definition of the theoretical framework, 3) development of the intervention content and communication channel (i.e., selection of behavior change techniques, writing messages, assessment for tailoring), 4) pilot testing, and 5) testing and evaluating the intervention. The primary objective of the intervention is to increase total time spent in moderate-vigorous PA (MVPA) for those who do not meet the Dutch guideline, to maintain PA levels of those who meet the guideline, or to further increase that if they indicate they want to improve further. The theoretical framework is informed by the social cognitive theory, and insights from other theories and evidence were added for specific topics. Development of the intervention content and communication channel resulted in the development of an app that provides highly tailored coaching messages that are framed in an autonomy-supportive style. These coaching messages include behavior change techniques aiming to address relevant behavioral determinants (e.g., self-efficacy, outcome expectations) and are partly context-specific. A model-based reasoning engine has been developed to tailor the intervention with respect to the type of support provided by the app, to send relevant and context-specific messages to the user and to tailor the graphs displayed in the app. For the input of the tailoring, different instruments and sensors are used, such as an activity monitor (Fitbit One), online and mobile questionnaires, and the location services on the user’s mobile phone.

The systematic and stepwise approach resulted in an intervention that is based on theory and input from end-users. The use of a model-based reasoning system to provide context-specific coaching messages goes beyond many existing eHealth and mHealth interventions.
This chapter is based on the article that is conditionally accepted as:

Anouk Middelweerd, Saskia J. te Velde, Julia S. Mollee, Michel C.A. Klein, and Johannes Brug (2018). “Development of Active2Gether: An app-based intervention combining evidence-based behavior change techniques with a model-based reasoning system to promote physical activity among young adults”. In: Journal of Medical Internet Research
13.1 Introduction

Insufficient physical activity (PA) is a risk factor for avoidable burden of disease (World Health Organization, 2010, 2014). About 25 percent of the adult population worldwide does not meet the recommended guidelines for PA (World Health Organization, 2014), while this is around 50 percent in many western countries like the US and the Netherlands (Volksgezondheidenzorg.info, 2016). Moreover, engagement in moderate to vigorous PA decreases with age, in particular when transitioning from adolescence into (young) adulthood (Bell and Lee, 2005; Kwan et al., 2012).

Research examining the determinants of PA mainly focuses on social cognitive and social-ecological factors (Anderson et al., 2006; Giles-Corti and Donovan, 2002; Luszczynska and R. Schwarzer, 2005; Plotnikoff et al., 2013; Ralf Schwarzer et al., 2008; Sniehotta et al., 2005; Young et al., 2014). Social cognitive theories and models, such as the health belief model (Abraham and Sheeran, 2005), the theory of planned behavior (Conner and Sparks, 2005), and the social cognitive model (Luszczynska and R. Schwarzer, 2005), have been developed to explain health behaviors and to guide research on health behavior and behavior change (Bauman et al., 2012; Conner and Norman, 2005). Whereas these models mainly focus on intrapersonal and interpersonal factors, social-ecological models more explicitly recognize that behavior may also be strongly influenced by contextual factors, such as the sociocultural and physical environments people live in (Bauman et al., 2012; Sallis, Cervero, et al., 2006; Sallis, Saelens, et al., 2009). For example, Sallis, Cervero, et al. (2006) propose a framework that recognizes that individuals are physically active within different domains (e.g., recreation, transport, household and occupation), where different factors on multiple levels influence their overall PA behavior. Thus, interventions that aim to increase levels of PA should not only target intra- and interpersonal factors, but should also take their physical and social environments into account.

Health promotion interventions targeting healthy lifestyle behaviors, including those aiming to promote PA (e.g., aiming to change PA behaviors), have been heavily influenced by information and communications technology (ICT) development. Where most health promotion interventions used to be face-to-face activities and/or printed materials, nowadays we rely much more on web-based and mobile (app-based) interventions—eHealth and mHealth interventions— that support and enable the personalized tailoring of face-to-face interventions and the large reach of print-based materials (Crutzen et al., 2013; Vandelanotte et al., 2016). Several reviews and meta-analyses of eHealth interventions targeting PA found small effects on levels of PA in favor of the intervention groups (Davies et al., 2012; Krebs et al., 2010; Webb et al., 2010). The majority of the studies included mainly respondents of Caucasian ethnicity (Davies et al., 2012; Krebs et al., 2010), the majority of the participants was female in most studies (Davies et al., 2012; Krebs et al., 2010), and in most studies higher-educated participants were over-represented (Davies et al., 2012), and thus the results are not generalizable to the populations at large.

A recently published systematic review reviewed studies that used apps in interventions to influence health behavior, including PA (Schoeppe et al., 2016). The majority of those studies that targeted adults reported significant intervention effects (Schoeppe et al., 2016). Furthermore, the majority of the interventions that reported significant changes in behaviors and health-related outcomes included behavior change techniques as goal setting, self-monitoring and feedback on the performance (Schoeppe et al., 2016). Furthermore, several content analyses were conducted to identify if and how constructs
of behavior change theories and behavior change techniques (BCTs) are incorporated in PA promotion apps. Generally, the apps analyzed were lacking applications of behavior change theories and the use of evidence-based behavior change techniques (Conroy et al., 2014; Cowan et al., 2013; Direito et al., 2014; Middelweerd, Mollee, et al., 2014; West et al., 2012). Moreover, apps mostly provide generic advice or tips about PA, and gamification, punishment and context-aware feedback are rare among PA apps (Mollee, Middelweerd, et al., 2017). Only a few apps incorporate some form of adaption to the user (Mollee, Middelweerd, et al., 2017). Lastly, existing apps fail to meet guidelines for PA (Knight et al., 2015; Modave et al., 2015). Despite the fact that health and fitness apps are popular among smartphone users (Intelligence, 2016; Statista, 2016), recent research indicates that most presently available apps lack the necessary empirical basis to make a meaningful difference in PA promotion (Vandelanotte et al., 2016).

In general, it has been found that health promotion interventions informed by established health behavior theory are associated with higher effect sizes than interventions not based on theory (Michie, Abraham, et al., 2009; Vandelanotte et al., 2016; Webb et al., 2010). Likewise, when established BCTs are incorporated, effectiveness is more likely (Michie, Abraham, et al., 2009; Olander et al., 2013; Webb et al., 2010). More specifically, interventions that included a self-monitoring feature in combination with features as prompting intention formation, specific goal setting, providing feedback on performance, or reviewing behavioral goals, were significantly more effective than interventions that did not include these BCTs (Michie, Abraham, et al., 2009).

Systematic reviews further showed that ICT-supported individually-tailored interventions are superior to generic interventions in promoting PA; in effects as well as user engagement and appreciation (Broekhuizen et al., 2012; Brouwer et al., 2011; Krebs et al., 2010; Webb et al., 2010). Moreover, Krebs et al. (Krebs et al., 2010) demonstrated that dynamic tailoring (i.e., iteratively assessing and providing feedback) was associated with larger effect sizes compared to static tailoring (i.e., all feedback is based on one baseline assessment). Modern technology, such as smartphones, smartphone applications (apps) and activity trackers, offers new possibilities in health promotion, especially for young adults of whom the majority owns a smartphone (Center, 2016; TelecomNieuwsNet, 2016). Furthermore, the rapid growth of the popularity and variety of health and fitness apps and activity trackers suggests that young adults will appreciate and adopt an app-based PA intervention.

In summary, innovative mobile technology-based approaches that are evidence-based and include dynamic tailoring may help to effectively support achievement and maintenance of behavior change in the PA domain. However, both the empirical basis and dynamic tailoring are lacking in current apps. Thus, there is a need for PA apps that incorporate constructs of behavior change theories and BCTs, and that provide dynamically tailored feedback.

Therefore, we developed the Active2Gether intervention that combines mobile (app-based) technology with dynamically tailored feedback and aims to go beyond existing (mobile) PA interventions. The aim of the present paper is to describe the systematic development and content of this Active2Gether PA promotion intervention. Section 13.2 provides an overview of the methodology that was used to develop the intervention and a brief description of the target population. Section 13.3 provides detailed information on the systematic development and content of the intervention. A summary of the resulting intervention is presented in Section 4. Finally, Section 5 reflects on the process and looks
Methods

This section provides an overview of the methodology that was used to develop the Active2Gether intervention.

13.2.1 Intervention development

We used a 5-step systematic approach to develop and evaluate the intervention (see Table 13.1). In order to ensure that the app was informed by relevant health behavior and health behavior change theory and evidence, the development was guided by the program-planning model developed by Kreuter et al. (2013), the intervention mapping protocol (Bartholomew et al., 1998), and the Medical Research Council (MRC) framework for the development and evaluation of complex interventions (Goyal et al., 2016). Table 13.1 provides a detailed overview of the stepwise process of the development of Active2Gether. The five steps are further described in Section 13.3.

Table 13.1: Description of stepwise process for the development of Active2Gether.

<table>
<thead>
<tr>
<th>Step</th>
<th>Step description</th>
<th>Overlap with intervention mapping protocol</th>
<th>Overlap with MRC framework</th>
<th>Overlap with program-planning model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Defining the intervention’s primary and secondary objectives</td>
<td>Identifying relevant physical activity behaviors to increase MVPA</td>
<td>Step 1: State program goals</td>
<td>Not applicable</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Defining the main and sub objectives of the intervention</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Defining the theoretical framework</td>
<td>Describing a theoretical framework on how to promote MVPA in general and stair use, active transport and sport in particular</td>
<td>Step 2: Select determinants for behavioral and environmental outcomes</td>
<td>Development: Identify and developing theory</td>
</tr>
</tbody>
</table>
### Chapter 13. Description of the development and content of Active2Gether

<table>
<thead>
<tr>
<th>Step</th>
<th>Description of the Development and Content of Active2Gether</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Selecting behavior change techniques based on theory and evidence in order to address determinants of behavior, based on existing studies and review. Assessing existing applications, what is available? Exploring preferences of end-users. Writing tailored messages. Designing tailoring algorithms for in the reasoning system. Channel of communication: building web-based app and system to combine, interpret data and to send messages.</td>
</tr>
<tr>
<td>4</td>
<td>Pilot testing the intervention to detect errors and impracticalities in order to improve the intervention prior to its implementation.</td>
</tr>
<tr>
<td>5</td>
<td>The intervention will be used by a larger group of participants and then will be analyzed and evaluated with respect to effect, process and impact.</td>
</tr>
</tbody>
</table>

#### 13.2.2 Target population

The Active2Gether intervention focuses on healthy, young adults. Therefore, participants in user studies that contributed to the development of the intervention, and participants in the quasi-experimental trial to assess the intervention’s effectiveness, were eligible for this study if they met the following criteria: (a) aged 18–30 years, (b) being (apparently) healthy, (c) Dutch speaking, (d) signed the informed consent form, and (e) in possession of a suitable smartphone running on iOS or Android.
13.3 Development of the Active2Gether intervention

This section provides more detail on the systematic development of and content of the Active2Gether intervention.

13.3.1 Step 1: Defining the intervention's primary and secondary objectives

Step 1 resulted in the decision that the primary objective of the Active2Gether intervention is to increase total time spent in MVPA for those who do not meet the Dutch guideline, to maintain PA levels of those who meet the guideline, or to further increase that if they indicate they want to improve further. The secondary aims defined are: a) to increase the underlying specific categories of MVPA, i.e. minutes of weekly sports participation, weekly numbers of stairs climbed, and/or weekly minutes of active transport, and b) to enhance the underlying determinants of the PA behaviors.

13.3.2 Step 2: Defining the theoretical framework

As a result of Step 2, a theoretical framework was built based on the relevant scientific literature (please see further details below), since behavior change was to be established by changing the underlying behavioral determinants (Brug et al., 2005) and evidence shows that interventions grounded in evidence-based behavior change theories are more likely to be effective (Davies et al., 2012; Webb et al., 2010). The theoretical framework was subsequently used to develop the content of the intervention, and to explain and predict the PA behaviors of the users of the intervention so that the intervention content could be tailored to each individual user.

The social cognitive theory (SCT) was adopted as a basis for the theoretical framework, as it is one of the most prominent behavior change theories used to inform interventions targeting health behavior change (Greaves et al., 2011; Tougas et al., 2015; Young et al., 2014), and a recent meta-analysis reported that SCT concepts may explain 31 percent of variance in PA (Young et al., 2014). SCT addresses both individual and social factors and recognizes the reciprocal relation between the individual and her or his context or environment. For these reasons, SCT thus guided and informed the intervention’s theoretical framework; insights from other theories and evidence were added for specific topics. Figure 13.1a shows the structural pathways of SCT.

Self-efficacy as a key construct within SCT (as well as in other health behavior theories) (Bandura, 2004; Conner and Norman, 2005) was adopted as a key construct in Active2Gether. Self-efficacy is defined as someone’s beliefs in his or her own capabilities to perform certain actions needed to achieve a desired outcome. Self-efficacy affects PA both directly and indirectly (Figure 13.1). Self-efficacy may influence outcome expectations – one’s beliefs about the positive and negative consequences of one’s behavior, such as participating in physical activities (Bandura, 2004; Conner and Norman, 2005). In other words, people who are more efficacious about being physically active will also be more likely to expect the favorable outcomes of participating in physical activities (Bandura, 2004). Moreover, self-efficacy may also influence how people perceive potential obstacles and impediments (Bandura, 1989, 2004). Goal setting was adopted as a second important basis for change, where goals can be either proximal (i.e., shorter-term intentions to act) or distal (i.e., longer-term goals to achieve something) (Bandura, 2004). Proximal goals are goals set for the shorter term and tend to promote more detailed planning that help people to make action...
plans (Bandura, 2004; Grant, 2012). Distal goals are goals set for the longer term and set the course of personal change (Grant, 2012). According to Bandura (Bandura, 2004), distal or long-term goals can initiate behavior change, but are not sufficient to change PA directly (Figure 13.1). Goal setting is dependent on levels of self-efficacy and perceived barriers and opportunities. In line with this notion, a meta-analysis inspired by the action-control framework indicated that 48% of the participants who intended to be physically active failed to do so. Therefore, forming intentions is often not sufficient to realize behavior change and self-regulatory and action-control techniques are needed to support behavioral enactment (Rhodes and Bruijn, 2013). A further meta-analysis on effective techniques in healthy eating and PA interventions concluded that interventions that offered self-monitoring and addressed self-regulation were more successful in increasing PA than interventions not including those techniques (Michie, Abraham, et al., 2009).

SCT posits that when individuals adapt and revise their behavior, they may adjust their beliefs and goals regarding this behavior (Bandura, 2004). In our theoretical framework, we therefore included ‘satisfaction’, defined as an evaluation of the PA behavior.

In line with SCT, we also recognized that the social environment influences behavior through social norms and that performing certain behavior can evoke social reactions, both positive and negative (Bandura, 2004). In the Active2Gether intervention, we do not only address intrapersonal (e.g., lack of motivation, tiredness) and social barriers (e.g., lack of support), but also contextual impediments (e.g., lack of time, weather, travel distance) (see Figure 13.1).

Lastly, it was decided that users will be categorized based on their awareness of their personal PA levels before they will be coached; people who are overly optimistic about their PA levels – i.e. who believe they engage in adequate amounts of PA while their data show insufficient levels – will be much less likely to be motivated to increase their PA levels (Lechner, 2006).

13.3.3 Step 3: Developing the Active2Gether intervention

In Step 3, the evidence-based behavior change techniques (BCTs) were identified and linked with the behavioral determinants of the theoretical framework and translated into tailored messages. These messages address specific behavioral determinants to create a ‘message library’ of actual feedback and advice messages tailored to all possible levels of the relevant behavioral determinants as recognized in the underlying theoretical framework. In order to actually tailor the messages to the individual users, the tailoring variables – i.e., the PA behaviors and their underlying determinants – are assessed for each individual user. The web-based Active2Gether app and accompanying same-content website are the communication channel to deliver the tailored intervention content. These elements, i.e. the BCTs, the development of the feedback messages and the assessment and tailoring methods are described hereafter, where we elaborate on the content. A more detailed description of the technical development of the Active2Gether intervention is published elsewhere (Klein et al., 2017).

I. Selection of behavior change techniques

We identified the relevant BCTs and linked these with the behavioral determinants of the theoretical framework described in Step 2 by means of a review of the relevant literature, based on an existing taxonomy of BCTs (see Table 13.2) (Abraham and Michie, 2008;
13.3 Development of the Active2Gether intervention

(a) Structural pathways of Bandura’s social cognitive theory.

(b) Specific theoretical framework used for the Active2Gether intervention. (The bold lines and boxes represent the elements that are based on the Social Cognitive Theory and the dotted lines and oval boxes represent behavioral determinants added to the theoretical framework.)

Figure 13.1: Theoretical framework in Active2Gether.
Michie, Richardson, et al., 2013). To explore which BCTs were used in already existing PA promotion apps, a systematic content analysis of such apps available in iTunes and Google Play was conducted (Middelweerd, Mollee, et al., 2014). This content analysis showed that the apps available to date generally lack sufficient incorporation of evidence-based BCTs (Middelweerd, Mollee, et al., 2014). BCTs that were applied most often were: providing feedback on performance, prompting self-monitoring of behavior, prompting specific goal setting, and planning social support or social change (Middelweerd, Mollee, et al., 2014). Additionally, focus group discussions with the target population indicated that participants preferred self-monitoring, goal setting and a ranking feature, but were not willing to share their accomplishments on social media for social comparison and to initiate social support (Middelweerd, van der Laan, et al., 2015). The focus groups further suggested that the Active2Gether app should be highly personalized, have an easy-to-use design and format, include a coaching feature that provided tailored feedback to self-set goals, enable competition with friends by ranking or earning rewards, and include the option to personally customize the application (Middelweerd, van der Laan, et al., 2015). The methods and results of these focus groups have been published in more detail elsewhere (Middelweerd, van der Laan, et al., 2015). Finally, an online cross-sectional survey among 179 young adults to assess their ratings with respect to the importance of specific BCTs applied in apps and their preferences for personalized tailoring (Belmon et al., 2015) confirmed the need for a personal coaching feature and showed that BCTs addressing goal setting, goal reviewing, feedback and self-monitoring were rated as important to be incorporated in an app, whereas social support and social comparison were considered less important (Belmon et al., 2015). The combined results of the review of the literature, the focus group discussions and the survey guided the selection of the BCTs to be included in Active2Gether (see Table 13.2).

Table 13.2: Overview of the behavior change techniques that were selected to target the behavioral determinants of the theoretical framework and how they were applied within the intervention.

<table>
<thead>
<tr>
<th>Determinant</th>
<th>BCT</th>
<th>How applied</th>
<th>Theory BCT</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome expectations</td>
<td>Provide general information on</td>
<td>Messages in general and tailored to aspects of the intake questionnaire</td>
<td>Information-motivation-behavioral skills model</td>
<td>(Abraham and Michie, 2008)</td>
</tr>
<tr>
<td></td>
<td>consequences of behavior in</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>general</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-efficacy</td>
<td>Action planning/time management</td>
<td>Messages prompting the planning of physical activity, e.g. the suggestion</td>
<td>Goal setting theory</td>
<td>(Olander et al., 2013; Sniehotta, 2009;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>to mark time and day in the calendar</td>
<td></td>
<td>Williams and French, 2011)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social comparison</td>
<td>Graph tailored to preference</td>
<td></td>
<td>Social comparison theory</td>
<td>(Williams and French, 2011)</td>
</tr>
<tr>
<td></td>
<td>social comparison (up-/downward)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Component</td>
<td>Description</td>
<td>Theoretical Framework</td>
<td>References</td>
<td></td>
</tr>
<tr>
<td>--------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------</td>
<td>------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Persuasion</td>
<td>Persuasive messages on how to overcome barriers</td>
<td>Social cognitive theory</td>
<td>(Williams and French, 2011)</td>
<td></td>
</tr>
<tr>
<td>Prompt self-monitoring</td>
<td>Messages that prompt to look at the monitoring graphs &amp; display of graph</td>
<td>Self-regulation, social cognitive theory, control theory</td>
<td>(Bandura, 1991; Olander et al., 2013)</td>
<td></td>
</tr>
<tr>
<td>Plan social support</td>
<td>Messages with suggestions to tell friends and ask for support</td>
<td>Social support theories</td>
<td>(Olander et al., 2013; Williams and French, 2011)</td>
<td></td>
</tr>
<tr>
<td>Imaginary reward</td>
<td>Messages that tell the user to be proud if they did well</td>
<td>Self-regulation, social cognitive theory, self-determination theory</td>
<td>(Bandura, 1991)</td>
<td></td>
</tr>
<tr>
<td>Intentions progress towards goal</td>
<td>Messages that tell the user how much he/she has already achieved &amp; display of graph</td>
<td>Self-regulation, social cognitive theory</td>
<td>(Olander et al., 2013; Williams and French, 2011)</td>
<td></td>
</tr>
<tr>
<td>Motivational messages</td>
<td>Messages telling the user how well he/she is doing and to keep up the good work or telling the user some advantage of being physically active</td>
<td>Social cognitive theory</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Modeling</td>
<td>Messages stating how well others are doing &amp; display of graph</td>
<td>Social cognitive theory, theory of planned behavior</td>
<td>(Abraham and Michie, 2008)</td>
<td></td>
</tr>
<tr>
<td>Provide instruction</td>
<td>Messages that prompt the user to prepare the sports bag the night before</td>
<td>Social cognitive theory</td>
<td>(Abraham and Michie, 2008)</td>
<td></td>
</tr>
<tr>
<td>Prompt goal setting</td>
<td>Messages prompting the user to set a goal and providing a suggestion</td>
<td>Self-regulation, social cognitive theory</td>
<td>(Abraham and Michie, 2008)</td>
<td></td>
</tr>
<tr>
<td>Impediments</td>
<td>Messages that provide information on how to deal with a specific barrier</td>
<td>Social cognitive theory</td>
<td>(Abraham and Michie, 2008)</td>
<td></td>
</tr>
</tbody>
</table>
### Chapter 13. Description of the development and content of Active2Gether

<table>
<thead>
<tr>
<th>Social norm (descriptive and inductive)</th>
<th>Prompt barrier identification</th>
<th>Messages that provide information on how to deal with a specific barrier</th>
<th>Social comparison theory</th>
<th>(Abraham and Michie, 2008)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prompt barrier identification</td>
<td>Messages that provide information on how to deal with a specific barrier</td>
<td>Social comparison theory</td>
<td>(Abraham and Michie, 2008)</td>
<td></td>
</tr>
<tr>
<td>Self-regulation</td>
<td>Self-monitoring</td>
<td>Messages that prompt to look at the monitoring graphs &amp; display of graph</td>
<td>Self-regulation, social cognitive theory</td>
<td>(Abraham and Michie, 2008)</td>
</tr>
<tr>
<td>Goal setting</td>
<td>Messages that prompt the user to set a weekly goal</td>
<td>Control theory</td>
<td>(Abraham and Michie, 2008)</td>
<td></td>
</tr>
<tr>
<td>Progress towards goal</td>
<td>Messages that tell the user how much he/she has already achieved &amp; display of graph</td>
<td>Control theory</td>
<td>(Abraham and Michie, 2008)</td>
<td></td>
</tr>
<tr>
<td>Self-evaluation</td>
<td>Messages prompting the user to evaluate how he/she is feeling about failing or achieving the self-set goal</td>
<td>Control theory, integrated theory of health behavior change</td>
<td>(Abraham and Michie, 2008)</td>
<td></td>
</tr>
<tr>
<td>Imaginary reward</td>
<td>Messages that tell the user to be proud if they did well</td>
<td>Self-regulation, social cognitive theory, self-determination theory</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfaction</td>
<td>Self-evaluation</td>
<td>Messages prompting the user to evaluate how he/she is feeling about failing or achieving the self-set goal</td>
<td>Control theory, integrated theory of health behavior change</td>
<td>(Abraham and Michie, 2008)</td>
</tr>
<tr>
<td>Long-term goals</td>
<td>Provide general information on consequences of behavior in general</td>
<td>Messages providing general information on consequences of behavior in general</td>
<td>Information-motivation-behavioral skills model</td>
<td>(Abraham and Michie, 2008)</td>
</tr>
</tbody>
</table>

### II. Writing tailored messages

Next, the BCTs were translated into actual tailored feedback and advice messages. In line with the self-determination theory (Gagné and Deci, 2005), these messages were written in an autonomy-supportive style. Messages were also written in a way to support relatedness and individualization, e.g. by addressing the users personally by their names. By respecting their autonomy and by making them feel related to the Active2Gether intervention, we
Development of the Active2Gether intervention

aimed to increase the user’s willingness to follow up on the coaching messages. Moreover, the messages were written in a positive gain-framed style, i.e., a style that describes the potential gains (e.g., in health, fitness, relaxation) when participating in PA rather focusing on loss (e.g., ill health, lack of fitness, stress) when not engaging in PA (Riet et al., 2009). The majority of the messages were tailored to determinants in the theoretical framework, the weather and occupational status. Creating the message library was an iterative process of brainstorming, writing a set of messages (AM), and providing feedback and suggestions (JM, StV).

A pilot test of a subset of messages among seven female bachelor and master students to test whether the tone-of-voice and the content appealed to the target population indicated that the messages were friendly, motivational, empathic and some were perceived as autocratic whereas some were not. Some minor changes were made in the messages.

III. Assessment for tailoring

Further decisions were made on how to measure the characteristics for tailoring messages. The following subsections describe the assessment of physical activity, the assessment of behavioral determinants, the assessment of the users’ locations and the assessment of friendship connections.

(A) Assessment of physical activity

First of all, after considering functionalities, validity and costs of a range of available activity trackers, it was decided to monitor the user’s activity using the Fitbit One, which includes monitoring of steps and stairs climbed on a minute basis. The Fitbit One was chosen because of its functionalities and the small size. The Fitbit One is a lightweight tri-axial accelerometer with a built-in altitude monitor that (among others) assesses the user’s step activity and number of floors climbed. The activity monitor communicates with the Fitbit app and website that display the collected data, for example by showing a color-coded chart indicating the proximity to the step goal, which is set to a default of 10,000 steps per day. Our test of the validity of the Fitbit One indicated that Fitbit can be considered a valid device to assess step activity for real-time minute-by-minute self-monitoring, although an overestimation of 677 steps per day by the Fitbit was seen (Middelweerd, Ploeg, et al., 2017). However, the validation study indicated that the Fitbit is less suitable for providing instant real-time feedback and for daily feedback on PA intensity levels (i.e., minutes of moderate, vigorous or MVPA) as it substantially and systematically overestimates time spent per intensity level per hour (Middelweerd, Ploeg, et al., 2017). For that reason, the Fitbit is only used to assess step activity.

Fitbit allows developers and researchers to access Fitbit data and thus to link and integrate the Fitbit data into health behavior interventions such as Active2Gether. In order to access the Fitbit data, Fitbit offers an application programming interface (API). The participants need to give permission once for the developers to access their activity data, this then can be collected regularly and a summarized version of the data is stored in the Active2Gether database. This data is utilized in several ways: for presenting the activity level to the user, for determining the type of coaching and to tailor coaching messages.

(B) Assessment of behavioral determinants

It was decided that behavioral determinants are assessed by means of a questionnaire, with a long and a short version, which were selected based on validations of such questionnaires. The long version is part of an ‘intake’
questionnaire before the actual intervention and as a point of departure for the tailored intervention, while the short version is used repeatedly throughout the intervention period to dynamically tailor the intervention content to the user. The literature was reviewed for relevant, existing and validated questionnaires. The long version is based on existing questionnaires that have previously been validated (i.e., Neighborhood Quality of Life Survey, Self-efficacy scales) or questions used in previous studies and were translated and adapted where necessary (Sallis, Cervero, et al., 2006; Sallis, Saelens, et al., 2009; Van Sluijs et al., 2004). In the short questionnaire, we decided to use single item questions to assess each of the behavioral determinants that are part of the framework and the system. In the short version of the questionnaire, all determinants are specified for each coaching domain (i.e., sports participation, stairs use and active transport). These items were not pretested as such, but were based on the long questionnaire. The Appendix provides an overview of the questions asked in the long and short version of the questionnaire, including the answer options.

(c) Assessment of location We also included questions about the participants’ significant places (e.g., home address, parental home, sports location, university, work location) in the intake questionnaire. These questions focus on travel options from home to the significant locations, thus information about the active and non-active transportation options. Additionally, information about the number of stairs available at each location and the maximal number of stairs that the participant is willing to climb in one go is assessed as well.

The user’s location (GPS coordinates) is collected using Google’s location services that can be linked with the Active2Gether app. The location data is used to determine whether the user visited his/her significant locations (e.g., home, study/work place, sports club) and to trigger user input about transport and travels that have been made. In addition, information about the characteristics of locations is used for personalized coaching messages to the user. For instance, if a person is being coached to using the stairs more often at work or at the university, it is only useful to suggest this when the option to climb the stairs is indeed present at the worksite/university.

(d) Assessment of friends Information regarding the participants’ friends is collected using the Facebook API. Users are asked to provide access to their Facebook ID and their connections by logging into Facebook once and giving permission for this. It is important to note that Facebook does not provide personal information about someone’s Facebook connections, but only a list of Facebook IDs of their connections. This information can be used to see whether any Active2Gether users are connected on Facebook. If two participants of the current intervention are connected on Facebook, they see a ranking within the app that shows both users’ achievements. In this way, the users only share their achievements with a closed group and not with ‘everybody’, according to the preferences stated in the focus group discussions.

IV. Tailoring and personalization
In order to realize high and dynamic tailoring, we combined evidence-based BCTs with predictive modeling (i.e., reasoning based on a computational model). Tailoring and personalization of the intervention content is realized in six ways: (A) determining the personally
appropriate type of support (i.e., education, coaching, or feedback), (B) selecting the personally preferred domain of PA for coaching (i.e., sports participation, stair use, active transport), (C) suggesting a weekly goal, (D) selecting the personally appropriate behavioral determinants for coaching, (E) sending relevant coaching messages, (F) tailoring and personalization of the app content. To realize such tailored coaching, we developed a system that combines detailed behavior monitoring with intelligent data interpretation and model-based predictions. Thus, combining the data from the different sources, the system enables personalization of the coaching strategies to try to achieve the most positive effect on behavior change. Detailed information on the system and the development of the system can be found elsewhere (Klein et al., 2017). In the following subsections, each of the tailoring levels is explained in more detail, after which the actual communication channel for the tailored feedback and advice is explained.

(A) Determining type of support  First, the type of support (i.e., education, coaching, feedback) is determined based on the user’s actual activity level (i.e., assessed with the Fitbit One) and the user’s perception of his/her activity level. If users do not meet the Dutch PA guidelines, but think they are sufficiently active, they will be educated to create more awareness (education). Users will be coached towards more physically active behavior (coaching) if they are aware that they are insufficiently active or if they are sufficiently active but still motivated to be more physically active. Users will receive positive feedback to maintain their activity level (feedback) if activity levels are already according to recommendations (and they don’t have the desire to increase their PA level). Users for whom the educational option was deemed most suitable will receive educational messages for a week and then they will automatically receive coaching messages.

(B) Selecting a coaching domain  As mentioned in Step 1, Active2Gether targets sports participation, stair use and active transport. Users that are being coached towards more PA can select one of the three coaching domains. Based on the user’s activity level and the opportunities for increasing PA in the three domains for a specific user in his own context (based on information about the user’s context from the intake questionnaire), the user receives a recommendation. However, the final decision about the coaching domain is up to the user, the system only provides an informed suggestion.

(C) Suggesting weekly individually tailored goals  Based on the user’s behavior, a suitable weekly goal is estimated for increasing or maintaining PA in the specific domain tailored to the specific. The system suggests individual domain-specific goals that are tailored towards the actual activity levels of the user. Users who did not meet their previous goal will be encouraged to stick to that goal, whereas users who met their goal will be encouraged to make further steps by raising their goal by 10%. However, the final decision about the coaching domain and actual goals set is again up to the user, as the system only provides tailored suggestions. Additional to the domain-specific goal, the app shows a chart that indicates the proximity to the 70,000 steps per week goal (i.e., averaging 10,000 steps per day) that is set as default in the Fitbit app as well.
(d) Selecting potentially effective behavioral determinants for coaching  In order to surpass existing app-based interventions, we used advanced artificial intelligence based techniques. The system makes use of a computational model of behavior change, based on the theoretical framework, that is used to predict behaviors (e.g., sports, stair use, active transport) for each participant. The relevant behavioral determinants are assessed with the short questionnaire on a weekly basis based on the single items described in the Appendix. The computational model is used to simulate how changes in scores on determinants can lead to changes in the behaviors. Different changes, in terms of which determinant is selected to change, are simulated and the results are ranked by the size of the effect on the behavior. This rank order is then used to probabilistically select the most relevant determinant to be coached on. This simulation process is repeated weekly based on the most recent answers on the short questionnaire to keep the dynamic tailoring up-to-date.

(e) Compiling and sending individually tailored coaching messages  Users receive feedback and advice messages from the message library that are tailored to their personal activity and behavioral personal determinants and contextual factors. During the day, the system checks three times whether a relevant message can be sent to the user. To be specific, the system creates a new personal message library for each user three times per day that eliminates irrelevant coaching messages. Detailed information on how the system creates this new personal messages library is described elsewhere (Klein et al., 2017). For example, messages that should only be sent on a specific day of the week or within a specific time window will be eliminated when necessary. Thus, the user receives up to three coaching messages per day. By only including such personally tailored messages, irrelevant feedback and advice is reduced, increasing personal relevance. Some messages are further personalized by explicitly referring to data the users provided themselves, such as the goals they set, the number of steps taken, or the number of stairs climbed. Educational messages are sent in the same way. The messages are not individually tailored, but contain evidence-based information on the importance and benefits of PA. The messages pop up on the smartphone with a push notification, and are presented as overlay on top of the dashboard (i.e., the collection of graphs in the app that depict the user’s data). As long as the app is not opened to read the message, the user will receive a notification every 15 minutes. In addition, the dashboard displays the five most recent messages.

(f) Further personalization of the app content  Finally, the messages and graphs displayed in the app are further personalized by mentioning the name of the user, presenting the user’s own activity data (a graph with step activity data and a graph with data from the coaching domain), and displaying the comparison with others in the user’s preferred way (i.e., up- or downward comparison) (Mollee and Klein, 2016). The other users are either Facebook friends or other participants (i.e., who are about equally active as the user and who will not be mentioned by name). Detailed information on the selection of users for the ranking feature can be found elsewhere (Klein et al., 2017). Figure 13.2 shows a screenshot of the app. This way, the users only share their achievements with a closed group and not with ‘everybody’, according to the preferences stated in the focus group discussions.
13.3 Development of the Active2Gether intervention

The Active2Gether app is a web-based application that is suitable for Android phones running on version 4.0 or higher. The app shows the website in a format that is viewable for smaller screens. Thus, the intervention content was accessible through the app or through the website.

The app shows a non-personalized, generic avatar with a welcome message that mentions the user’s current weekly goal. The app displays the current number of daily steps and stairs climbed. In addition, the app shows 4 graphs: 1) a bar chart with the step progress towards 70,000 steps per week, 2) a ranking with six other Active2Gether users – where possible Facebook friends – based on the step activity over the last seven days, 3) the activity data for each week day for the current coaching domain (i.e., minutes of sport activity, numbers of stair climbed or minutes of active transport), 4) the step activity for each week day. The third and fourth graph display the user’s own data as well as the average data assessed within Active2Gether. Moreover, these graphs can be adjusted according to the user’s preferences: they can show data for the last week, last month or from the first use.

Tailored messages and short questions are sent via push messages through the app. After the user reads the messages, they are displayed at the bottom of the app. Only the five messages sent most recently are displayed in the app. Figure 13.2 shows a screenshot of the app.

(6) Communication channel and display The Active2Gether app is a web-based application that is suitable for Android phones running on version 4.0 or higher. The app shows the website in a format that is viewable for smaller screens. Thus, the intervention content was accessible through the app or through the website.

The app shows a non-personalized, generic avatar with a welcome message that mentions the user’s current weekly goal. The app displays the current number of daily steps and stairs climbed. In addition, the app shows 4 graphs: 1) a bar chart with the step progress towards 70,000 steps per week, 2) a ranking with six other Active2Gether users – where possible Facebook friends – based on the step activity over the last seven days, 3) the activity data for each week day for the current coaching domain (i.e., minutes of sport activity, numbers of stair climbed or minutes of active transport), 4) the step activity for each week day. The third and fourth graph display the user’s own data as well as the average data assessed within Active2Gether. Moreover, these graphs can be adjusted according to the user’s preferences: they can show data for the last week, last month or from the first use.

Tailored messages and short questions are sent via push messages through the app. After the user reads the messages, they are displayed at the bottom of the app. Only the five messages sent most recently are displayed in the app. Figure 13.2 shows a screenshot of the app.
13.3.4 **Step 4: Pilot testing**

In order to detect possible bugs in the system, and to assess user friendliness and appreciation, the app was pilot-tested in two steps. First, the Active2Gether team (AM, JM, AMR, StV, MK) used the initial version of Active2Gether. Bugs and nuisances et cetera were monitored, listed and fixed accordingly when and where possible. Second, seven people from the target population (5 women, 21–28 years old) were recruited to use the adjusted version of the app, monitor bugs, nuisances and provide feedback in person and answered a questionnaire regarding use, user friendliness and appreciation. The app was further adjusted based on that information. For example, the timing of the different steps in the tailoring process (i.e., determining the type of feedback, the coaching domain, the weekly goal and the most promising behavioral determinants) did originally not account for exceptional cases, in which a user takes very long to complete a step, which caused a next step to be skipped. In the adjusted version, multiple checks and safety mechanisms were implemented to make sure that the tailoring process could still be finished correctly in such conditions. Also, automated messages to remind users to charge their Fitbit and to synchronize their data were added to the system, because of the observation that participants in the pilot study sometimes did not notice when it was necessary to do so.

13.3.5 **Step 5: Testing and evaluating the intervention**

After developing the intervention, the intervention was evaluated on effectiveness and user appreciation in March 2016. A three-arm quasi-experimental trial — with an active control group — with a baseline and two follow-up assessments at 6 and 12 weeks was conducted to examine the effectiveness of the Active2Gether intervention. This trial is registered in the Dutch trial registry, No. NTR5630.

13.4 **Summary of the Active2Gether intervention**

The Active2Gether intervention is an app-based intervention designed to help and encourage young adults to become and remain physically active by focusing on the domains of active transport, stair climbing and sports participation. To do so, participants are categorized into one of three awareness categories (education, coaching and feedback). Participants in the education category receive educational messages on the benefits of PA, and participants in the feedback category receive motivational messages to maintain their active lifestyle. Participants in the coaching category are coached on sports participation, taking the stairs or active transport. Every week, the participants are asked to choose one of these three coaching domains and to set a weekly goal. Participants receive a suggestion for a coaching domain and a weekly goal based on their previous behavior, but the final decision is up to the user. The participants receive a Fitbit One activity tracker that can be synchronized with the app and that allows the participants to monitor their PA behavior. Lastly, the app sends (daily) coaching messages addressing relevant behavioral determinants. The content of the messages is tailored to the user’s behavioral determinants, occupational status and the local weather conditions. Lastly, the app displays the activity data of the participant, including a graph displaying the activity data of six other participants, preferably friends. The graph with the activity data of others ranks the participants based on their weekly step activity and based on the user’s preferences for social comparison, i.e. upward or downward comparison. Detailed information on the assessment and tailoring are provided in Step 3 in Section 13.3.
13.5 Discussion

The current article describes the development and the content of Active2Gether, an app-based intervention, which was developed using a systematic and stepwise approach. The aim of the Active2Gether intervention is to empower young adults to become and remain physically active by providing them with app-based tailored coaching and feedback. Active2Gether makes use of an activity tracker and personalized, context-specific feedback. It focuses on three PA domains, builds on established behavior theory, and applies evidence-based BCTs and a model-based reasoning system in order to provide individually tailored coaching messages based on current scores on behavioral determinants. The development and the content of Active2Gether was a stepwise and time-consuming process. A strength of this approach is the involvement of the target population. However, as the possibilities of modern technology are constantly and rapidly changing and evolving, possibilities and preferences that were assessed at the beginning of the development might be outdated and/or no longer preferred today. The development and the content of Active2Gether were guided by relevant health behavior theories and scientific evidence, thereby aiming to develop an intervention that provides highly tailored feedback. Consequently, less attention was paid to app design and aesthetics, which might have resulted in a less appealing app compared to commercial apps. Furthermore, the app is only available for Android devices running on a 4.0 version and therefore is not available for older Android devices and iPhones or smartphones running on other operating systems.

Active2Gether incorporates a number of conditions to secure high levels of engagement. First, our approach, integrating a model-based reasoning system, allows us to provide the user a dynamically tailored intervention that adjusts to the changes in the user. Second, by applying multiple levels of tailoring in the app and the content of messages (i.e., type of support, coaching domain, coaching messages, weekly goals), the app is likely to be regarded as personally relevant and to increase feelings of relatedness. Third, by comparing the user’s PA with other Active2Gether users (preferably with their Facebook friends), we expect to further increase the personal relevance and relatedness. Lastly, by giving the user the option to select from three PA domains and set their own goals with guidance and suggestions based on their own input, we expect higher levels of autonomy, resulting in higher motivation to follow up on the coaching messages. However, in order to implement these different levels of tailoring, detailed user information is needed repeatedly and thus frequent user input is needed, which increases user burden.

To date, mobile phones and PDAs have been used to monitor and PA with either smartphone apps or external devices, to deliver feedback, to provide information and to offer a support system to the participants (Vandelanotte et al., 2016). Active2Gether makes use of an external device, the Fitbit One, to monitor PA and to provide feedback through the app based on the user’s behavior. However, Active2Gether goes beyond existing interventions by combining data from multiple sources in order to send context-specific messages. Furthermore, the majority of the published interventions focuses on step activity (Kirwan et al., 2012; Poirier et al., 2016), whereas Active2Gether focuses on sports activity, climbing the stairs and active transport as well. Therefore, the app may be more appealing to participants who do not like to participate in sports, especially because the user can adapt his/her coaching domain every week. However, Active2Gether does not yet incorporate geo-fencing (i.e., sending location-triggered messages), which would further improve the possibilities for context specificity and real-time feedback and advice by for example sending
a reminder to climb the stairs at work when users are close to their work location.

So far, the majority of app-based interventions to promote PA showed positive effects (Schoeppe et al., 2016). In line with other app-based PA interventions, Active2Gether makes use of self-monitoring, goal setting and providing feedback. However, Active2Gether provides dynamically tailored feedback using artificial intelligence based techniques and including conditional factors (i.e., weather), whereas other interventions use logic statements and decision rules to specify which messages should be sent to the user. For example, Active2Gether uniquely assesses the behavioral determinants every week to provide tailored advice and feedback on the current behavior, whereas most studies mostly provide feedback on the current behavior only (Glynn et al., 2014; Kernot et al., 2013; King et al., 2013, 2016; Stuckey et al., 2011). Current app-based interventions to promote PA focus on step activity or overall MVPA (Fukuoka et al., 2010; Glynn et al., 2014; Kernot et al., 2013; King et al., 2013, 2016; Stuckey et al., 2011), whereas Active2Gether focuses on sports activities, active transport and climbing the stairs as well. As the majority of the app-based interventions reported significant effects (Schoeppe et al., 2016), and Active2Gether goes beyond those apps and includes proven to be effective BCTs, we expect to see significant intervention effects compared to the control groups.

Active2Gether is ambitious and innovative and incorporates certain risks. For example, the intervention highly relies on input from the activity monitor and location sensor and thus on the user to turn on and synchronize the tracker with the server. Furthermore, it relies on responses from the users on repeated questionnaires. If they do not provide input at all or if they do not provide true and honest answers, the coaching messages that are informed by this information may become irrelevant and non-tailored. Moreover, if a participant is not a Facebook user or has no appropriate contacts, the personalization could be limited. Finally, if technical problems are encountered, this may result in errors in synchronization and sending messages late or not at all. To limit the burden for the participants and to minimize their input to reduce potential technical problems, future research could make use of smartphone sensors to assess the participant’s behavior.

The overall effectiveness of Active2Gether thus needs to be – and was – evaluated in a quasi-experimental trial with a 12 week follow-up. However, as app-based interventions offer the possibility to deliver just-in-time interventions that are relevant for the user’s situation for that particular moment, a study is needed to examine the possible effectiveness of specific real-time feedback and advice moments (Klasnja et al., 2015). Ecological momentary assessment (Shiffman et al., 2008) in such a quasi-experimental trial setting may help to assess potential specific effects throughout the intervention period. Evaluation of the efficacy of the intervention and the usability can help to further adapt and improve the intervention for future research. Furthermore, data collected during the trial can provide insights on how to further personalize content to the users. The quasi-experimental trial also includes monitoring of app use as well as a process evaluation of app use and appreciation that will provide information on larger scale dissemination and implementation, as well as information on changes required to improve conditions for wider use of the app.

As the intervention has been developed with an early consideration of the preferences of the target population, it is more likely to meet the expectations of the target population. Consequently, the intervention is more likely to be adopted by them. However, the intervention might be prone to technical errors and a significant input from the user is needed to provide tailored feedback. This might be a burden for the participants, leading to a lower adoption
rate. We conducted a small pilot study to test the Active2Gether app and to detect bugs and technical errors, but ideally the pilot study would have been conducted with a larger sample. The current version of the Active2Gether intervention has been developed for healthy young adults owning a smartphone running on Android version 4.0 or higher. The content needs to be adjusted before offering the intervention to other target populations.

**Abbreviations**

API  Application programming interface  
App  Smartphone application  
BCT  Behavior change technique  
ICT  Information and communications technology  
MRC  Medical Research Council  
MVPA  Moderate-vigorous physical activity  
PA  Physical activity  
PDA  Personal digital assistant  
SCT  Social cognitive theory

**Competing interests**

This research is partly funded by Philips. The authors declare that there is no conflict of interests regarding the publication of this paper.

**Acknowledgments**

The authors want to thank Adnan Manzoor Rajper for his contribution to the project.
Chapter 13. Description of the development and content of Active2Gether

References


Belmon, Laura S., Anouk Middelweerd, Saskia J. te Velde, and Johannes Brug (2015). “Dutch young adults ratings of behavior change techniques applied in mobile phone apps to promote physical activity: a cross-sectional survey”. In: *JMIR mHealth and uHealth* 3.4.


Intelligence, SurveyMonkey (2016). *These fitness app statistics show what’s going right (and wrong) for Fitbit*. [Online].
Chapter 13. Description of the development and content of Active2Gether


Michie, Susan, Michelle Richardson, Marie Johnston, Charles Abraham, Jill Francis, Wendy Hardeman, Martin P. Eccles, James Cane, and Caroline E. Wood (2013). “The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an
international consensus for the reporting of behavior change interventions”. In: *Annals of behavioral medicine* 46.1, pages 81–95.


Regulation as a Framework for Chronic Health Condition Interventions”. In: PLOS ONE 10.8. Edited by Delphine Sophie Courvoisier, e0134977.


Williams, S. L. and D. P. French (2011). “What are the most effective intervention techniques for changing physical activity self-efficacy and physical activity behaviour—and are they the same?” In: Health Education Research 26.2, pages 308–322.


### Appendix – Overview of questions used for the long and short versions of the questionnaires.

<table>
<thead>
<tr>
<th>Behavioral determinant</th>
<th>Example question in long version</th>
<th>No. of items in long version</th>
<th>Question in short version</th>
<th>Answer options</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Outcome expectations</strong> [NQLS¹, (Sallis, 2010)]</td>
<td>If I participate in regular physical activity or sports, then: • I will improve my health • I will feel more attractive • I will lose weight • I will improve my physical fitness • I will feel relaxation • I will feel less tension and stress</td>
<td>6</td>
<td>If I participate in regular sports, then it will have a positive effect, for example on my healthy, appearance, weight or how I will feel. If I regularly take the stairs, then it will have a positive effect, for example on my healthy, appearance, weight or how I will feel. If I regularly bike or walk, then it will have a positive effect, for example on my healthy, appearance, weight or how I will feel.</td>
<td>1 – No reason at all 2 – A slightly important reason 3 – A quite important reason 4 – A very important reason</td>
</tr>
<tr>
<td><strong>Self-efficacy</strong> [Self-efficacy scales for exercise (Sallis, Pinski, et al., 1988), NQLS¹ (Sallis, 2010)]</td>
<td>How confident are you that you could do PA, in each of the following situations? I’m confident that I could: • Do PA even when I’m tired • Do PA even when I’m in a bad mood • Do PA even when I feel I don’t have time • Do PA even when I am on holiday • Do PA even when it is raining</td>
<td>12</td>
<td>How confident are you that you will do sports in the next week even when you’re tired, busy or when it’s bad weather? How confident are you that you will take the stairs in the next week even when you’re tired, you’re in a hurry or you’re with others? How confident are you that you will cycle or walk to work/the university in the next week, even when you’re tired, you’re busy or when it’s bad weather?</td>
<td>1 – Not at all confident 2 – Slightly confident 3 – Moderately confident 4 – Very confident 5 – Extremely confident</td>
</tr>
</tbody>
</table>
### Perceived barriers for sport [NQLS\(^1\) (Sal-lis, 2010)]

<table>
<thead>
<tr>
<th>How often do the following barriers prevent you from doing sports activities?</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Bad weather</td>
</tr>
<tr>
<td>• Lack of time</td>
</tr>
<tr>
<td>• Lack of interest in exercise</td>
</tr>
<tr>
<td>• Other priorities</td>
</tr>
<tr>
<td>• Lack of skills or knowledge</td>
</tr>
<tr>
<td>• Lack of equipment</td>
</tr>
<tr>
<td>• Lack of facilities or space</td>
</tr>
<tr>
<td>• Lack of physical fitness</td>
</tr>
<tr>
<td>• Lack of energy</td>
</tr>
<tr>
<td>• Lack of money</td>
</tr>
<tr>
<td>• Lack of company</td>
</tr>
<tr>
<td>• Self-conscious about my looks when I exercise</td>
</tr>
</tbody>
</table>

### Perceived barriers for active transport

<table>
<thead>
<tr>
<th>How often do barriers prevent you from traveling by bike or by walking instead of traveling by car or public transport?</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Bad weather</td>
</tr>
<tr>
<td>• Lack of time</td>
</tr>
<tr>
<td>• Lack of physical fitness</td>
</tr>
<tr>
<td>• Lack of energy</td>
</tr>
<tr>
<td>• Too many pieces of luggage</td>
</tr>
<tr>
<td>• Travel distance is too far away</td>
</tr>
<tr>
<td>• No suitable bike</td>
</tr>
</tbody>
</table>

### Perceived barriers for stairs climbing

<table>
<thead>
<tr>
<th>How often do barriers prevent you from climbing the stairs?</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Lack of physical fitness</td>
</tr>
<tr>
<td>• Lack of energy</td>
</tr>
<tr>
<td>• Too many pieces of luggage</td>
</tr>
<tr>
<td>• Too many flights of stairs</td>
</tr>
</tbody>
</table>

### How often do barriers prevent you from participating in sports or exercise activities? Think for example of lack of time, lack of energy, costs, lack of company.

| 1 – Never |
| 2 – Rarely |
| 3 – Sometimes |
| 4 – Often |
| 5 – Very often |

### How often do barriers prevent you from cycling or walking to work / the university instead of traveling by public transport or car? Think for example of lack of time, lack of physical fitness, lack of energy or too many pieces of luggage.

| 1 – Never |
| 2 – Rarely |
| 3 – Sometimes |
| 4 – Often |
| 5 – Very often |

### How often do barriers prevent you from climbing the stairs? For example, barriers as being in a hurry, lack of physical fitness, lack of energy or carrying too many pieces of luggage?

| 1 – Never |
| 2 – Rarely |
| 3 – Sometimes |
| 4 – Often |
| 5 – Very often |
### Social norm descriptive

<table>
<thead>
<tr>
<th>Question</th>
<th>Option 1</th>
<th>Option 2</th>
<th>Option 3</th>
<th>Option 4</th>
<th>Option 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>My friends think that I should be sufficient physically active.</td>
<td>1 – I strongly disagree</td>
<td>2 – I somewhat disagree</td>
<td>3 – Neutral</td>
<td>4 – I somewhat agree</td>
<td>5 – I strongly agree</td>
</tr>
<tr>
<td>My fellow students think that I should be sufficient physically active.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>My brother(s) and/or sister(s) think that I should be sufficient</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>physically active.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Social norm injunctive

<table>
<thead>
<tr>
<th>Question</th>
<th>Option 1</th>
<th>Option 2</th>
<th>Option 3</th>
<th>Option 4</th>
<th>Option 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>How often do your friends / roommates / brothers or sisters / parents</td>
<td>1 – Never</td>
<td>2 – Rarely</td>
<td>3 – Sometimes</td>
<td>4 – Often</td>
<td>5 – Very often</td>
</tr>
<tr>
<td>participate in physical activities?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Intentions

<table>
<thead>
<tr>
<th>Question</th>
<th>Option 1</th>
<th>Option 2</th>
<th>Option 3</th>
<th>Option 4</th>
<th>Option 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do you intend to do sports (more often) within the next week / month / 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>months even if you think you’re already sufficiently active?</td>
<td>1 – Most definitely will not</td>
<td>2 – Probably will not</td>
<td>3 – Maybe / maybe not</td>
<td>4 – Probably will</td>
<td>5 – Most definitely will</td>
</tr>
<tr>
<td>I intend to do sports (more often) within the next week.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I intend to climb the stairs (more often) within the next week.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I intend to bike or walk to work/the university (more often) within the</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>next week.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Self-regulation

[Anderson et al, 2006 (Anderson et al., 2006)]

I keep track of how active I am. I check whether I met my goals. In the last three months, I:

- Set aside time for my daily physical activity
- Walked or biked instead of drove or traveled by public transport
- I exercised or did physical activities with someone else
- I wrote it down in my calendar to do sports/physical activity
- I planned do sports/exercise even when the weather was bad

7

I planned and set goals do regular sports/exercise. When I wasn’t able do it I evaluated why I wasn’t able to do it and whether I needed to change something.

1 – Never
2 – Rarely
3 – Sometimes
4 – Often
5 – Very often

### Satisfaction

<table>
<thead>
<tr>
<th>How satisfied are you about how physically active you are?</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>How satisfied are you about how often you did sports in the last week?</td>
<td>0 – Very unhappy</td>
</tr>
<tr>
<td>How satisfied are you about how often you climbed the stairs last week?</td>
<td>10 – Very happy</td>
</tr>
<tr>
<td>How satisfied are you about how often you biked or walked to work/the university last week?</td>
<td></td>
</tr>
</tbody>
</table>

### Long-term goals

<table>
<thead>
<tr>
<th>How motivated are you to be more physically active?</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Scale: 0 – Very unmotivated to 10 – Very motivated)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>How important do you think it is to be more physically active?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 – Very important</td>
</tr>
<tr>
<td>2 – Important</td>
</tr>
<tr>
<td>3 – Neutral</td>
</tr>
<tr>
<td>4 – Not important</td>
</tr>
<tr>
<td>5 – Totally not important</td>
</tr>
</tbody>
</table>

1 NQLS: Neighborhood Quality of Life Study (Sallis, 2010).
2 Questions assessing social norm injunctive were not included in the short questionnaire.
14. Description of the technical design of Active2Gether

Abstract
Lack of physical activity is an increasingly important health risk. Modern mobile technology, such as smartphones and digital measurement devices, provides new opportunities to tackle physical inactivity. This paper describes the design of a system that aims to encourage young adults to be more physically active. The system monitors the user’s behavior, uses social comparison and provides tailored and personalized feedback based on intelligent reasoning mechanisms. As the name suggests, social processes play an important role in the Active2Gether system. The design choices and functioning of the system are described in detail. Based on the experiences with the development and deployment of the system, a number of lessons learnt are provided and suggestions are proposed for improvements in future developments.
Physical inactivity is an increasingly serious health problem: the World Health Organization (WHO) has identified that it is the fourth leading risk factor for global mortality (World Health Organization, 2014). The organization estimates that a lack of physical activity leads to 3.2 million deaths per year globally. Physical inactivity has all to do with modern sedentary lifestyles, which are led by 60% to 85% of people worldwide, according to the WHO. One aspect of a sedentary lifestyle is that people are more inclined to passive modes of transportation. Active traveling modes such as biking and walking can contribute to a healthy level of physical activity (Sahlqvist et al., 2012). Another aspect of a sedentary lifestyle is related to the work environment, where much work is done by people seated in chairs in front of computers. Research suggests that having desk jobs increases health risks up to 50%. Integrating small activities in work routines can help to increase physical activity and lower health risks (Levine, 2014).

At the same time, modern mobile technology, such as smartphones and digital measurement devices, provides new opportunities to tackle physical inactivity. In 2015, 43% of the adults worldwide owned a smartphone, with percentages up to 70% for developed countries (Poushter, 2016). Smartphones allow for continuous and real-time monitoring of activity behavior via built-in sensors such as accelerometers, and provide possibilities for giving contextualized and personalized feedback. This makes the smartphone a potentially powerful device for real-time coaching of people towards a more active lifestyle. To use smartphones and sensors for this aim, they should be integrated into a behavior change support system, which is defined as an information system designed to form, alter, or reinforce attitudes or behaviors (Oinas-Kukkonen, 2010).

In this paper, we describe the design of such a system in detail, together with lessons learnt and suggestions for future developments. The system is developed in context of an interdisciplinary research project and is called Active2Gether. The goal of the project is to combine domain knowledge from experts in physical activity interventions with modern mobile technology to design an intervention that encourages physical activity among healthy young adults. One of the innovative aspects of the system is that it exploits model-based reasoning techniques for tailoring the coaching to the needs of the user. Up to now, this has hardly been applied within existing interventions (Mollee, Middelweerd, Kurvers, et al., 2017).

As the name suggests, social processes play an important role in the Active2Gether system. This is reflected in different ways, such as the implementation of social comparison mechanisms on both an individual and a group level. In addition, the system addresses psychological constructs as social norms and social aspects of outcome expectations in its coaching messages. The aim of the Active2Gether system is to increase or maintain levels of physical activity among young adults in the age group of 18 to 30 years. The system is being evaluated in a trial (see (Middelweerd, te Velde, et al., 2018) for a detailed description) in which over 100 participants, aged between 18 and 30 years old, used either a variant of the Active2Gether system or the standard website that belongs to a commercial activity tracker for approximately three months. The user evaluation of the system by the participants is described in (Mollee, Middelweerd, te Velde, et al., 2017); in this paper, we focus on the architecture and functionality of the system.
14.2 System description

To the user, the Active2Gether system presents itself as an Android-based mobile phone app that continuously monitors the context of a person. One of the distinct features of the system is that it implements evidence-based behavior change techniques, unlike most apps that are currently available in the app stores (Middelweerd, Mollee, et al., 2014). The most promising behavior change techniques are employed in the app, including self-monitoring, performance feedback, goal setting and social comparison.

The app performs four main functions: it communicates with the user about his/her objectives regarding physical activity for the next week, provides timely and personalized feedback, facilitates self-monitoring based on several collected data sources, and supports social comparison with the help of Facebook friendship relations. The system focuses on three types of physical activity: leisure time sports activities, active transport and stair walking. Users can choose to be coached on at most one of these three domains at the same time.

In the following sections, we describe the design of the system in detail. We first describe the data that are collected by the system. We then provide an overview of the architecture of the system and describe the layout choices. After that, we explain the working of the reasoning engine and the selection and filtering of coaching messages. Finally, we describe the implementation of the social comparison functionality.

14.2.1 Data collection

The system uses several mechanisms to collect information from and about the users. Below, we describe how questionnaires as well as sensor measurements are used to understand the user behavior.

Intake questionnaire

The user starts with filling in an online intake questionnaire. Besides demographic information, this questionnaire asks about their significant locations (such as home, work, study, etc.), their travel options between these locations, and psychological factors underlying their physical activity behavior (e.g., skills, barriers, goals, and outcome expectations). The answers to these questions are used for tailoring the messages to the user’s personal situation and are used in the model-based reasoning about the effect of specific coaching strategies on a specific user (see Section 14.2.4).

Activity tracker

The commercial Fitbit One is used as activity tracker that registers the daily number of steps and the number of stairs climbed (Fitbit Inc., no date). Users receive an account on the Fitbit website. The Fitbit device uses Bluetooth LE to automatically synchronize activity data with the Fitbit servers, either via the Fitbit mobile phone app, or via a Bluetooth LE dongle and a pc. After the first login to the Active2Gether system, the user is asked to connect to his/her Fitbit account with the Active2Gether system. A connection is established through an open authentication mechanism. Once the connection is made, Fitbit provides an authentication key for the user. This key is stored in the database, so the Active2Gether system can directly access the Fitbit web service to receive activity data for a user and store it into the Active2Gether database. A script runs every hour to update the database with the most recent activity data. Another script periodically checks whether the battery level is low.
or whether the last synchronization is more than three days ago, and, in that case, a reminder is sent.

GPS location
The Active2Gether app uses the built-in GPS sensor for recording the GPS coordinates (latitude and longitude). As soon as a user logs in, he is asked to authorize the use of location tracking. It is possible for a user to turn off the location detection option, but this will disable certain features. In a separate experiment, we compared different time intervals for collecting GPS data (Manzoor et al., 2017). It turned out that a frequency of five minutes provides a good balance between battery consumption and precision. Every 15 min, the data on the mobile phone are synchronized with the server. In the database, latitude, longitude, speed, accuracy and time stamp are stored for each observation.

A script runs every night to see whether a user has visited one of his important locations by comparing the GPS trace with the coordinates of the important locations. Since the user locations are provided in descriptive form, geocoding is used to transform them into latitude and longitude numbers. The system stores the number of minutes at each of the locations. If the duration at a location is longer than 0 min, we can conclude that the user visited that location.

Daily questions
Every day, a number of questions is posed to the user via the app. Information about the visited locations is used to prompt the user about the travel options that he used to go there. As the users had to list two options (active and inactive) in the intake questionnaire, they are asked to choose between these two options. Since the system is aware of the activity level of the different options for each user, it can derive the types of transport used and the amount of active travel minutes.

In addition, the user is prompted about the sports activities during the day before. When a user regularly answers that he did not participate in sports activities, the frequency of asking about sports is decreased to once per week.

14.2.2 Architecture
The system is comprised of five main components: (1) an app on a mobile phone; (2) a commercial activity tracker; (3) a database with user (activity) data and persuasive messages; (4) a model-based reasoning engine to interpret the data and predict the effect of different coaching strategies; and (5) a communication engine that selects and sends questions and messages to the app. Figure 14.1 shows those main components.

Figure 14.2 provides an overview of the most important data flows in the system. The details are provided in the following sections.

Structure of the app
In order to provide users the possibility to also view their information via a website, the main dashboard of the system is developed as a web page. Within the app, the main component is a GUI element (i.e., a WebView component) that renders this web-based dashboard. Since a responsive web design approach is followed, the website automatically adapts to smaller screen sizes. Although the dashboard is actually a website, users do not notice this. The app behaves like a native app and users do not need to login separately via the WebView: once a user has registered his/her account for the Active2Gether system in the Android system, the
app uses those credentials to automatically log in the user and to show the appropriate page inside the WebView component as if it is a screen in the app itself.

The other functions of the mobile phone app are to facilitate the communication with the
user and to monitor the user’s location. The latter is done with the help of Google location services. Using the built-in Android synchronization system, the app connects every 15 min via a web service to the communication engine of the system. Messages or questions that are prepared for the user by the reasoning engine are collected and answers and read notifications are sent back. Whenever a new message or question is sent to a user, it appears in the status bar and when the user clicks on the message, it is shown as an overlay on the main screen.

### 14.2.3 Layout and visual design

In order to show a consistent look and feel to the user, a professional designer was hired to design and recommend different aspects of the user interface. The designer helped in suggesting layout, fonts and a coloring scheme for the website and consequently the dashboard of the app. There are eight panels (small rectangular windows) on the website, which show different kinds of information to the user depending on the chosen coaching domain for the current week. The first panel shows a picture of the coach and a welcome message corresponding to the current coaching domain. For example, if the current coaching domain is active transportation, the message is: “Hi Adnan! You have chosen to focus on active transportation this week. Your goal is to spend this week at least 36 minutes of active transportation. I will support your efforts.” The activity data are presented in many different views. Two small panels show the most recent steps and floors count for the present day. Whenever a user visits the website or opens the app, a dynamic script runs to show a summary of the most recent data in the Active2Gether database.

A panel with the caption “Progress to weekly step goal” shows a progress bar towards a weekly goal of 70,000 steps. There is another panel that shows the performance of other users (see Section 14.2.6 about social comparison below). Another panel shows the type of physical activity based on the chosen domain for the current week (active transport, stair walking or sports activities). In the first week, when no domain has been chosen yet, this panel shows the number of active minutes based on the reported sports activities and active travel choices. A similar panel shows the user’s activity in terms of the number of steps. The latter two panels provide an option to the user to view historical data per week, per month or from the beginning. This option is useful for those users who want to see their own past performances. They also provide the user an opportunity to compare his/her performance with the average values of all users. The final panel is dedicated to show the most recent messages. Figure 14.3 gives an example of the dashboard.

### 14.2.4 Model-based reasoning

One of the fundamental components of the Active2Gether system is the so-called reasoning engine, which analyzes and interprets the user’s data and determines what type of support the user should receive. A core component of this reasoning engine is a computational model, which is discussed below. The reasoning process can be split up into three parts: assessing the user’s activity and awareness level, suggesting a coaching domain based on hypothesized room for improvement, and predicting the most promising coaching determinants. Figure 14.4 shows a flow chart of the processes taking place in the reasoning engine.
Assessing the user’s awareness phase

In the first step of the reasoning engine, the user’s current activity level and awareness phase are assessed to determine what type of support they need. The assessment is based on two evaluations of the user’s physical activity level, namely an objective evaluation (i.e., whether the user meets the norm) and a subjective evaluation (i.e., whether the user thinks he/she is sufficiently physically active).

Using that information, the user is assigned to one of four categories, each representing an awareness state regarding their physical activity. The resulting categories are summarized in Table 14.1. Category 1, in which users believe that they are sufficiently active but objectively do not meet the norm, is more or less comparable with the precontemplation phase in Prochaska’s Transtheoretical Model of behavior change (Prochaska and DiClemente, 1982). In this phase, people are uninformed or under informed about the consequences of
their behavior and education about the consequences is needed. Category 4 is similar to the *maintenance* phase, while Categories 2 and 3 are comparable to the *action* phase in Prochaska’s model. Because our system can objectively measure whether people meet the norm, we can determine an awareness phase in a simpler and more accurate way than with the questions that are often used for determining the stage according to the transtheoretical model, and without the strong assumption that people always go through all phases.

Based on the categorization of the awareness of users, the system determines which type of support (i.e., *education*, *coaching* or *feedback*) the user needs. This assessment is reflected in the type of motivational messages that the user receives from the app (see also Section 14.2.5). In addition, for users who receive coaching, the system guides them to choose a coaching domain, prompts them to set a specific goal, and predicts the most promising coaching determinants, as further explained below. This evaluation is repeated every three weeks, in order to continuously tailor the system to the user’s current state. Thus, instead of treating all users the same, their specific needs and wishes are taken into account. This should lead to improved user acceptance and adherence, and consequently to increased effectiveness of the intervention (Kroeze et al., 2006).

**Suggesting a coaching domain and goal**

Users that are assigned to the coaching category are guided by the Active2Gether system to focus their behavior change efforts, by advising on the choice of a specific coaching domain and a goal. This cycle is repeated on a weekly basis. The coaching domains are parts of the

Figure 14.4: Process flow chart of reasoning engine.
user’s daily life: (1) stair use at significant locations (e.g., home, work, and university); (2) active transport to significant locations; and (3) leisure-time sports activities.

First, detailed information about the user’s context and behavior is used to identify in which domains the user could be more physically active. The user’s physical activity in each of the three domains is estimated based on a combination of activity data collected through the activity monitor (number of stairs climbed) and daily user input through the app (selected transport options to visited locations, time spent on sports activities). These physical activity values are then evaluated by comparing them to estimated “maximum” or “ideal” values, which are based on information about the user’s context and visits to their important locations. This context information is collected through an intake questionnaire, and includes information about the addresses of their significant locations, (active and non-active) travel options between these locations, relevant floor numbers on these locations, and the availability of stairs. For example, if a user works on the third floor and on average climbs another three floors during the day, a total number of six floors during a work day would be reasonable. For a user that works on the second floor, but on average climbs another eight floors during a work day, a total number of six floors is comparatively low. In addition, the more often the user has gone to work, the higher the expected number of stairs becomes. Similar evaluations are developed for the physical activity level in active transport and sports activities. Using these evaluations, the domain with the largest potential for improvement can be detected, as the evaluation score for that domain will be lowest. This domain is then suggested to the user as focus for the coaching in the upcoming week. However, the user is allowed to overrule this suggestion and opt for another domain.

After selection of a coaching domain, the user is asked to set a specific goal for this coaching domain, i.e., weekly time spent on active transport, weekly time spent on sports or daily number of stairs climbed. If users did meet the previous goal in this coaching domain, the system suggests increasing their goal by 10%. If users did not meet the goal last time, the system advises to keep the goal at the same level. Again, to ensure the user’s autonomy, the final decision on the goal is up to the user.

Table 14.1: User categories based on objective/subjective evaluation.

<table>
<thead>
<tr>
<th>No.</th>
<th>Objective</th>
<th>Subjective</th>
<th>User category</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Insufficient</td>
<td>Sufficient</td>
<td>The user is unaware that he/she is insufficiently physically active, and will be educated to increase this awareness.</td>
</tr>
<tr>
<td>2</td>
<td>Insufficient</td>
<td>Insufficient</td>
<td>The user is aware that he/she is insufficiently physically active, and will be coached to increase his/her physical activity level.</td>
</tr>
<tr>
<td>3</td>
<td>Sufficient</td>
<td>Insufficient</td>
<td>The user is sufficiently physically active, but still wants to be coached to increase his/her physical activity level.</td>
</tr>
<tr>
<td>4</td>
<td>Sufficient</td>
<td>Sufficient</td>
<td>The user is sufficiently physically active, and wants to maintain his/her physical activity level. This user will not be coached to increase his/her physical activity level, but only receive feedback.</td>
</tr>
</tbody>
</table>
This relative evaluation of the user’s behavior and the recommendation of a certain coaching domain respect the individuality of the users more than general physical activity guidelines. It prevents the system from imposing the same expectations on all users, even though their personal situations may be completely different.

Finding the most promising coaching determinants

Once the coaching domain is selected, the system investigates on which personal determinants the coaching messages should focus to yield the most promising effect on the desired behavior. These behavioral determinants are personal psychological concepts that govern the engagement in healthy behavior. The system contains a large collection of coaching messages, with subsets that each target one of the personal determinants. The messages are based on established behavior change techniques, such as prompting barrier identification, providing information on consequences, and prompting goal setting (Middelweerd, te Velde, et al., 2018).

In order to determine what messages are most likely to positively affect the user’s behavior, the effects of improving each one of the personal determinants are estimated based on simulations of a dynamic computational model. This model is a formalization of the dynamics between these personal determinants and the behavior, where each of the concepts is represented by a numerical value in range $[0, 1]$. The model is mainly based on the social cognitive theory, that describes the reasons why people fail or succeed to exhibit some desired (health) behavior from both social and cognitive determinants (Bandura, 1998, 2004). Other theories (e.g., self-regulation theory and health action process approach) and literature were consulted to extend the model to incorporate more relevant aspects. The model that was implemented in the Active2Gether system is an adaptation of the computational model presented in (Mollee and van der Wal, 2013). Revisions between the two versions of the model were motivated by a decrease in conceptual detail and computational complexity of the model, and by suggestions of experts in the domain of behavior change. The resulting model contains determinants such as intentions, self-efficacy and outcome expectations.

The values of all parameters in the model can be adjusted by the modeler. In the current implementation, the parameters were chosen based on correlations between the concepts found in literature (Plotnikoff, Costigan, et al., 2013; Plotnikoff, Lippke, et al., 2008; Rovniak et al., 2002). We performed simulations to find values that keep the ratio between the parameters in accordance with empirical findings (Mollee, Araújo, et al., 2017). A more detailed description and a preliminary validation of this model on empirical data, which showed promising results, can be found in (Mollee and Klein, 2017). Figure 14.5 shows a graphical representation of the computational model.

The simulation process starts with estimating the current states of the personal determinants by means of short questions via the app. The resulting values are used as input for the computational model. To simulate the effect of targeting one of the determinants, one of the values obtained from the app questionnaire is increased according to the hypothesized effect of sending coaching messages about this determinant. Then, the computational model simulates the dynamics between the determinants and estimates the effect on the behavior. By running simulations for each possible targeted determinant, a list of determinants is constructed, ordered by the most promising effect on the behavior variable. This order is taken into account when selecting coaching messages to the user. As with the selection of a coaching domain and goal, the simulation cycle is repeated weekly, in order to tailor to the
user’s strongest psychological needs at all times.

In contrast to the relative evaluation of the user’s behavior for suggesting a coaching domain, this part of the reasoning engine does not tailor the intervention based on information about the user’s environment, but rather on information about his/her motivational state of mind. This way, the users will receive support on the aspects that are relevant to their motivation and behavior.

14.2.5 Selection and tailoring of coaching messages

Once the order of the most promising coaching strategies is established (as described in Section 14.2.4), all necessary elements to send the coaching messages are in place. As explained before, each coaching strategy represents a set of messages targeting one of the concepts from the computational model. The messages are based on established behavior change techniques (e.g., prompting specific goal setting, time management; see (Abraham and Michie, 2008) for a complete taxonomy), while also taking into account user preferences (Middelweerd, van der Laan, et al., 2015). They were written to be motivational, personally relevant and trustworthy, and were annotated with restrictions for the circumstances under which they are relevant (e.g., day and time, the user’s awareness phase and coaching domain, the user’s perceptions reported in the intake questionnaire). A detailed description of the techniques that are applied in the messages and the related determinants is provided in (Middelweerd, te Velde, et al., 2018). The messages are sent up to three times a day. In order to send a message, it has to be selected from the set of available messages and (if applicable)
social comparison

The Active2Gether system uses social comparison, as one of the core ideas behind the system is that a healthy lifestyle can be maintained and achieved in the presence of a social support network. Social comparison is implemented in two ways in the Active2Gether system, namely on a group level and on an individual level. On the group level, social comparison is implemented by showing group averages adjacent to the user’s physical activity data in the graphs on the dashboard. This allows the user to compare his/her daily performance to other Active2Gether users anonymously.

On the individual level, social comparison is implemented as a ranking of the user’s performance within a list of other users, which is shown in one of the panels in the app/website. The ranking automatically updates every time a user visits the website or opens the app. Overall physical activity is used as a basis for the comparison, which is determined by the number of steps taken by an individual in the last seven days, but implementations based on activity data for one of the coaching domains are also conceivable.

In order to increase the relevance of the comparative data, the system tries to show actual friends of the user in the ranking. To do so, it extracts friendship relations from Facebook. Facebook is an obvious choice, since it is one of the most popular social media
Chapter 14. Description of the technical design of Active2Gether

websites globally and also very popular among Dutch young adults. A connection to the Facebook Graph API is established through an open authentication mechanism. As a new user logs in for the first time on the system, it asks the user whether he/she wants to grant access to Active2Gether to check for friendships with other Active2Gether users. If permission is granted and a match is found, the friendship connection is also registered in the Active2Gether system. It is not mandatory for users to grant access, but individuals who do not opt for such explicit social comparison can only observe the activities of other users anonymously, which will probably make the social comparison less effective.

Social comparison can be either upward or downward, depending on whether an individual compares to targets that perform better or worse. Both variants address different underlying motivational processes. Upward comparison can be beneficial if individuals use the target as a role model and motivation to self-improve, but it can have a discouraging effect if the target’s performance seems unattainable (Lockwood and Kunda, 1997; Maddux, 1995). Downward comparison can boost an individual’s self-esteem and thereby lay the groundwork for self-improvement (Wills, 1981). However, downward comparison could also have an adverse effect, since it leads to relatively low goals and since it does not challenge an individual to minimize the discrepancy with a better performing individual. An experiment testing the effects of presenting people with their preferred or opposite direction of social comparison in the domain of physical activity showed that it is important to take personal preferences into account (Mollee and Klein, 2016). Participants who were shown the type of social comparison opposite to their preference showed a decrease in overall physical activity. Showing the preferred direction of social comparison typically resulted in an upward trend, which was not statistically significant however. Even though the preferences were based on a simple question (whether the participants prefer to compare themselves to individuals who perform better or worse) and the sample size was small, the results demonstrated that these preferences matter: if not to enhance the motivational effects of social comparison, then at least to avoid the potential adverse effects of social comparison (Mollee and Klein, 2016).

When creating the ranking, the Active2Gether system takes the social comparison preference (upward or downward) of the users into account. The comparison preference of users was determined in the intake questionnaire with the same question as in the experiment described above (Mollee and Klein, 2016). The system first tries to find up to six friends whose activity level is in line with the preferred comparison direction. If there are more than six friends that match the preferred direction, the six friends that are closest to the current user in terms of step total of the last seven days are selected. If there are less than six friends in the preferred direction, the system selects other users in the preferred direction. If not sufficient people are found to create the ranking, then the system searches for users in the opposite direction, since it is not appropriate to show an empty ranking list to the user. This could happen, for example, if an individual’s preference for social comparison is upward, but his performance is among the best of the users. The data of befriended users are shown with their first name, but data of other users are shown with their initials only, to maintain a level of anonymity. Algorithm 14.1 shows the step-by-step process of selecting the friends or other users to show in the ranking.
Algorithm 14.1: Algorithm for selecting users for the social comparison.

1. **Goal:** show up to 7 users in total: the actual user plus maximum 6 other users.
2. // Select up to 6 friends most suited for the comparison.
3. if preference == downward then
   4. Select up to 6 friends with a lower total step count than the user ordered from high to low.
4. else if preference == upward then
   5. Select up to 6 friends with a higher total step count than the user ordered from low to high.
6. RESULTING #FRIENDS = number of selected friends
7. if RESULTING #FRIENDS < 6 users then
   8. // Add other users until there are 6 other users shown in total.
   9. if preference == downward then
      10. Select all users with a lower step total than the user, and then select the top (6 - RESULTING #FRIENDS) of that list.
   11. else if preference == upward then
      12. Select all users with a higher step total than the user, and then select the bottom (6 - RESULTING #FRIENDS) of that list.
   13. Merge resulting list with list of friends.
   14. RESULTING #USERS = number of selected users in list
   15. if RESULTING #USERS < 6 users then
      16. if preference == downward then
         17. Select all users with a higher step total than the user, and then select the top (6 - RESULTING #USERS) of that list.
      18. else if preference == upward then
         19. Select all users with a lower step total than the user, and then select the bottom (6 - RESULTING #USERS) of that list.
   20. Order the resulting list from highest to lowest step count.
   21. Show the full name of the friends.
   22. Show the other users anonymized (only initials).

14.3 Discussion

In this section, we offer suggestions for improvements based on our experiences with the development and use of the current Active2Gether system.

14.3.1 Data collection

Advances in (mobile) technology open up ways to improve the location and travel monitoring in the Active2Gether system.

First, for determining modes of transportation, we currently use daily questions in combination with location data (i.e., prompted user input). At the time of the design, this was a reasonable choice, considering the state-of-the-art and the consequences on battery consumption during full-time use of accelerometers in the mobile phone. Nowadays, the Google Activity Recognition API and the iOS CM Motion Activity class would be logical candidates, as they are the de-facto standards for providing location information (Apple Inc.,
Chapter 14. Description of the technical design of Active2Gether

2017; Google LLC, 2017). Power consumption remains an important issue, however.

Second, we used a complicated questionnaire for reporting significant locations and their characteristics (such as relevant floors). It is difficult and cumbersome to answer, and also difficult to update during the intervention. Therefore, we recommend to automatically detect significant locations (Manzoor et al., 2017), which can also be done via Google location services. A remaining drawback, however, is that the users are required to provide this privacy-sensitive information.

14.3.2 Architecture

The decision to use a combination of a web-based approach and a native app, which requires more-or-less permanent Internet connection, turned out well. Most users in the Netherlands apparently have good Internet connections on their phones. A drawback of our current choice was that integration with third-party APIs (i.e., Fitbit and Facebook authentication) was difficult, since Fitbit does not allow using their authentication API through the WebView. This can be solved with the newer Chrome tabs approach.

We decided to copy the data from Fitbit servers to our own database: a Cron job runs periodically to fetch the data through a web service. The advantage is that we could do our analyses more easily (e.g., summarizing the data for different time periods every hour) and have a good performance when we query the data. A disadvantage is that the information sometimes lags behind. If the performance is sufficient, we would recommend to dynamically invoke a web service at runtime. Another possible solution is the use of the use of more advanced services, such as the Fitbit subscription API, which allows sending notification to our system when new data are available.

Related to the point above, we let participants use the Fitbit app to synchronize their data with the activity tracker. This was necessary because it is not possible to read out the Fitbit device directly. As a consequence, the Fitbit app or a computer was needed in addition to our own system. It also required an additional step in the initialization, as users had to create an account on the Fitbit website. For future applications, direct communication between the coaching system and activity trackers is preferred; however, this is likely not easy with commercially available trackers.

We have decided to partly develop a native app for Android. A native app was necessary for implementing the location detection. New developments in standardization of location detection APIs and more advanced techniques for platform-independent development might result in a different choice, which could enlarge the potential user base. Another option would be to use the Google Fit API for getting information about the activity of users. Since its introduction, many systems and wearables directly integrate with this service.

14.3.3 Layout and visual design

Although the user interface of the app was designed in collaboration with a graphic designer, we put only minimal effort in its design. We did not receive any signs that this hampered the use of the app, but we imagine that following the design guidelines of the respective platform (Android and iOS) would improve the users’ perception.
14.3 Discussion

14.3.4 Model-based reasoning

Some of the design choices in the personalization process described in Section 14.2.4 seem successful, but we identified opportunities for improvement in others. Our findings on these elements of the Active2Gether system are described below.

Based on anecdotal feedback, we conclude that determining the user’s awareness phase by comparing their actual behavior to their perception works well. Acknowledging the users’ awareness of the need to change is a useful way of tailoring the coaching messages.

In contrast, the suggestion of a coaching domain could be improved. The current approach is not very flexible, as the scores for active transport and stair walking are based on the characteristics of the significant locations that were identified via the intake questionnaire. Any physical activity related to these domains on other locations is ignored during the evaluation of the user’s behavior. Since that activity is not taken into account for either the actual behavior or the “ideal” behavior, this simplification should not distort the behavior scores. However, it is recommended to also take behavior on (or during transit to) other locations into account, to get a more complete picture of the user’s behavior. This could be achieved by using more adaptive behavior evaluation algorithms, which learn the user’s potential or ideal from past behavior, possibly in combination with other (web) sources.

Part of the selection of coaching messages is based on the simulations of a computational model. Although a preliminary validation of the model showed very promising results (Mollee and Klein, 2017), the added value of the model in predicting the most effective coaching determinants still has to be evaluated. In theory, an adaptive approach can be used to learn the effect of specific (sets of) messages on a person’s behavior, which might lead to better suggestions for coaching determinants. The outcome of evaluating the model could for example lead to the decision to use personal and adaptive parameters in the computational model, or to take an entirely different approach (e.g., machine learning techniques).

14.3.5 Selection and tailoring of coaching messages

The messages that people receive are very diverse, but sometimes still give the impression that they are redundant or not on topic. For a more detailed investigation of the user experience of the messages, see (Mollee, Middelweerd, te Velde, et al., 2017). We have a number of suggestions for improvement.

First, the personal relevance of the messages could be improved. For example, the messages are only sent at specific times during the day, but the users’ physical location could be used to trigger messages as well. Furthermore, the selection of messages to be sent could be based on more complex combinations of information. For instance, combining the current location with the relevant floors on that location and the availability of stairs and the maximum number of stairs that a user is willing to walk. Incorporating these ideas would increase the context-awareness of the system. In addition, the messages should contain less trivial content in order to better fit the expectations of the target group.

Next, we implemented the selection of the message to be sent in such a way that the system sends the message that has been sent the longest time ago. However, if only a few messages are relevant, the users will still receive the same messages in a short time period. Therefore, it is important to adhere to a minimum amount of time between resending the same message. In addition, the interdependence or similarity between messages should be taken into account: if messages are different formulations of the same content, the system should be aware of that. If such improvements imply that a user does not receive
any coaching messages for some period of time, the system could observe this and send a warning to the developers to make sure that this lack of relevant messages is noticed and possibly remedied.

Finally, the coaching messages could be improved by implementing a feedback mechanism. Instead of only being able to click “OK” to close a message, the user could rate the message, and this feedback could be used to further tailor the system.

14.3.6 Social comparison

As explained in Section 14.2.6, the position of users in the ranking is based on their indicated preference. If people prefer upward comparison, they are shown users who perform better, and vice versa. This implies that users are always at the top or bottom position in the overview, irrespective of their performance. Although a study has shown that it matters to take the preferred direction of social comparison into account (Mollee and Klein, 2016), the specific implementation might still allow room for improvement. A negative consequence of the current design decision is that users may become demotivated if they do not see any acknowledgment of their efforts. Therefore, a less strict selection of other users to show in the ranking might work better to motivate users through social comparison.

In addition, social comparison is more effective if you know the people you are comparing with. If users only have a few Facebook connections that are also using the system, it is likely that they mostly see anonymized other users that they do not know. Therefore, it might be better to allow adding connections via the system directly, or to invite friends to start using the system as well. Another option could be to select similar users (in terms of occupational status, home town, gender, age, etc.) to show in the ranking, and to show and emphasize these similarities in the design, in order to strengthen the perceived closeness to the other users.

Related to this is the issue that social comparison might not be equally beneficial for all types of people. For example, it is expected that patients and individuals managing chronic conditions are not so much interested in social comparison, but could benefit from social support. Although our system targets healthy individuals, in general it is important to take such personal characteristics into account when reasoning about the specific behavior change techniques that are applied by the system to the users.

A final consideration is of ethical nature. In the current implementation, it is easily possible that user A is shown the data of user B, but not vice versa. This means that individual reciprocity of information sharing is not ensured, which could cause objections from potential users. In that case, a more sophisticated selection mechanism should be developed, in which such reciprocity is maintained.

14.4 Conclusions and future work

In this paper, we have described the design of the Active2Gether coaching system in detail. The coaching system aims to encourage physical activity among young adults by combining evidence-based behavior change techniques with elements from modern (mobile) technology, such as location monitoring and model-based reasoning.

The effectiveness of the system is currently being evaluated in a three-month trial with more than 100 participants between 18 and 30 years old. To determine the added value of the tailored messaging, a three-armed design has been chosen. One group uses the full
version of the Active2Gether system, a second group uses the Active2Gether system but does not receive tailored messages, and the third group uses the standard website and app that belong to the Fitbit tracker. The participants start with an intake questionnaire that contains questions about their personal situation, their current exercise behavior and perceptions about physical activity. After three months, a similar questionnaire is sent out. The participants are also asked to wear a validated activity monitor in the first week of the intervention and after three months. This allows us to conclude whether using the system leads to a significant increase in physical activity. However, because several behavior change techniques have been employed in the system, it is difficult to identify which technique actually influences behavior. In the case a positive overall effect is found, we will further analyze which messages were sent to users to identify the contribution of specific techniques (focusing on specific determinants) on the behavior change. In addition, future research is needed with variants of the system in which only specific components (i.e., social comparison, self-monitoring and goal setting) are functional.

In the current paper, we have discussed the architecture and the implementation of the Active2Gether system. In addition, we have shared lessons learnt during the design, implementation and evaluation of the system, as well as recommendations for further development and improvement. We believe that these insights and the detailed description of the technological choices will prove helpful to designers and developers of healthy lifestyle interventions to produce effective and appealing coaching systems.

Acknowledgments

The authors would like to thank Anouk Middelweerd, Saskia J. te Velde and Aart T. van Halteren for their contribution to the conceptual design of the Active2Gether system.
References


Plotnikoff, Ronald C., Sonia Lippke, Kerry S. Courneya, Nick Birkett, and Ronald J. Sigal (2008). “Physical Activity and Social Cognitive Theory: A Test in a Population Sample of Adults with Type 1 or Type 2 Diabetes”. In: Applied Psychology 57.4, pages 628–643.


Abstract

Insufficient physical activity (PA) is highly prevalent and associated with adverse health conditions and risk for non-communicable diseases. In order to increase levels of PA, effective interventions to promote PA are needed. Nowadays, modern technologies such as smartphones, smartphone applications (apps) and activity trackers offer new possibilities in health promotion. This study explored the use and short-term effects of an app-based intervention (Active2Gether (A2G)) to increase levels of PA in young adults.

Young adults aged 18–30 years were recruited ($N = 104$) using diverse recruitment strategies. The participants were allocated to the A2G-Full condition (tailored coaching messages, self-monitoring, social comparison), A2G-Light condition (self-monitoring, social comparison) and the Fitbit-only control condition (self-monitoring). All participants received a Fitbit One activity tracker – that could be synchronized with the intervention apps – to monitor PA behavior. A 12-week quasi-experimental trial was conducted to explore intervention effects on weekly moderate-vigorous PA (MVPA) and relevant behavioral determinants (self-efficacy, outcome expectations, social norm, intentions, satisfaction, perceived barriers, long-term goals). The ActiGraph wGT3XBT and GT3X+ were used to assess baseline and post-intervention follow-up PA.

Compared to the Fitbit condition, the A2G-Light condition showed the largest effect sizes for minutes of MVPA per day ($B = 3.1$, 95%CI = [-6.7, 12.9]), and smaller effect sizes were seen for the A2G-Full condition ($B = 1.2$, 95%CI = [-8.7, 11.1]). Linear and logistic regression analyses for the intervention effects on the behavioral determinants at post-intervention follow-up showed no significant intervention effects of the A2G-Full and A2G-Light condition. The overall engagement with the Fitbit activity tracker was high (median = 88 percent of the days), but this was lower in the Fitbit condition. Participants in the A2G conditions reported more technical problems than in the Fitbit condition.

The current study showed no statistically significant differences in MVPA or determinants of MVPA after exposure to the A2G-Full condition as compared to A2G-Light or Fitbit condition. This might partly be explained by the small sample size and the low rates of satisfaction in the participants in the two A2G conditions that might be due to the high rates of technical problems.
This chapter is based on the manuscript that will be submitted as:

15.1 Background

Insufficient physical activity (PA) is associated with adverse health conditions and noncommunicable diseases such as cardiovascular diseases, cancer and diabetes (I.-M. Lee et al., 2012; World Health Organization, 2010). Worldwide, about 25% of the adult populations does not meet the recommended guidelines for PA (World Health Organization, 2014). In western countries like the US and the Netherlands, about 50% of the population does not meet the guidelines (Volksgezondheidenzorg.info, 2016). Moreover, engagement in moderate-vigorous PA (MVPA) decreases with age, in particular when transitioning from adolescence into (young) adulthood (Bell and C. Lee, 2005; Kwan et al., 2012).

In order to increase levels of PA, effective interventions to promote PA are needed. Available research showed that interventions informed by established health behavior theories were associated with higher effect sizes than interventions not based on theory (Michie et al., 2009; Webb et al., 2010). Furthermore, interventions are more likely to be effective when established behavior change techniques, such as self-monitoring, goal setting and providing feedback on performance, are incorporated (Michie et al., 2009). Systematic reviews further show that individually-tailored interventions are superior to generic interventions in promoting PA, in effects as well as user engagement and appreciation (Broekhuizen et al., 2012; Brouwer et al., 2011; Cruzen et al., 2013; Krebs et al., 2010; Short et al., 2014; Vandelanotte et al., 2016; Webb et al., 2010). Moreover, Krebs et al. (2010) demonstrated that dynamic tailoring (i.e., iteratively assessing and providing feedback) was associated with larger effect sizes compared to static tailoring (i.e., all feedback is based on one baseline assessment).

Nowadays, modern technologies, such as smartphones, smartphone apps and activity trackers, offer new possibilities in health promotion. For example, advantages of using such technologies are: continuously accessible, convenient, accurate and continuous (self-)monitoring of PA, providing highly tailored and real-time feedback, large reach, and interactive features. The high adoption rate of smartphones (97% among adults aged 20–29 years in the Netherlands) and the popularity of health and fitness apps and activity trackers (TelecomNieuwsNet, 2016) suggests that young adults may appreciate and adopt app-based PA interventions.

Early mHealth interventions – interventions focused on text message-delivered interventions or interventions delivered via a personal digital assistant – showed promising results (Fanning et al., 2012; Head et al., 2013; Muntaner et al., 2016; Vandelanotte et al., 2016). A recently published systematic review reviewed studies that used apps in interventions to influence health behavior, including PA (Schoeppe et al., 2016). The majority of the studies that targeted adults reported significant intervention effects (Schoeppe et al., 2016). Furthermore, the majority of the interventions that reported significant changes in behaviors and health-related outcomes included behavior change techniques as goal setting, self-monitoring and feedback on the performance (Schoeppe et al., 2016).

In this context, we developed the “Active2Gether” intervention. A systematic and stepwise approach was used to develop the Active2Gether intervention guided by health behavior theory and scientific evidence (Middelweerd, te Velde, et al., 2018). This resulted in the development of an app suitable for providing highly tailored coaching messages that are framed in an autonomy-supportive style. These coaching messages include behavior change techniques aiming to address relevant behavioral determinants (i.e., self-efficacy, outcome expectations, intentions, impediments, long-term goals, social norm, satisfaction...
and self-regulation skills) and are partly context-specific. A fundamental component of the intervention is the model-based reasoning engine, i.e. a software system that generates conclusions from information stored in the database using logical techniques and a mathematical model that is used to predict behaviors by computer simulations. The reasoning engine is used to tailor the intervention with respect to the type of support provided by the app, to send relevant and context-specific messages to the user, and to tailor the graphs displayed in the app. Detailed information on the development and the technical design of the Active2Gether intervention can be found elsewhere (Klein et al., 2017; Middelweerd, te Velde, et al., 2018).

The primary objective of the Active2Gether intervention was to increase total time spent in MVPA for those who do not meet the Dutch guideline, to maintain PA levels of those who meet the guideline, or to further increase that if they indicate they want to improve further. The secondary aims were: (a) to increase the underlying specific categories of MVPA, i.e. minutes of weekly sports participation, weekly numbers of stairs climbed, and/or weekly minutes of active transport, (b) to enhance the underlying determinants of the PA behaviors.

The aim of the present study was to explore use and effects of the Active2Gether intervention on increased weekly levels of MVPA as well as on psychosocial determinants of MVPA in adults aged 18–30 years compared to two control groups, in a quasi-experiment. Since we could not realize a sufficiently valid and powered research design, this paper is published as an exploratory study online. We decided to publish this paper in this fashion, because the flaws in our research design do not warrant publication in a peer-reviewed journal. However, we do wish to share our results with the scientific community, since our study was registered in the Dutch trial registry (Dutch Trial Registry Registration number NTR5630), and to contribute to avoiding publication bias.

15.2 Methods

15.2.1 Design

A three-arm quasi-experimental trial was conducted to evaluate the short-term effects of the Active2Gether intervention. The trial included baseline, mid-intervention (6 weeks) and post-intervention assessments. Data collection took place between March 2016 and October 2016. The trial was registered (Dutch Trial Registry Registration number NTR5630) and the project protocol was approved by the Ethics Committee of the VU Medical Center Amsterdam. All participants provided written informed consent. The development of the Active2Gether intervention and evaluation plan is described in more detail in an earlier publication (Middelweerd, te Velde, et al., 2018).

15.2.2 Participants

Young adults were recruited by flyers, posters, social media, personal contacts and snow ball strategies. The majority of the participants were recruited through social media (48.4%), through other participants (18.6%) and through flyers and advertisement (11.2%) in the regions of Amsterdam, Leiden and Utrecht, the Netherlands.

Participants registered for the trial through the Active2Gether website by completing a web form asking information about gender, age, type of smartphone they owned (i.e., Android or iOS). Participants were eligible for the study when they met the following criteria: (a) aged 18–30 at time of registration, (b) in possession of a suitable smartphone...
running on Android or iOS, (c) being apparently healthy, (d) Dutch speaking, and (e) signed
the informed consent form. Participants were excluded if they were unable to visit the
research facilities for the intake procedure. Figure 15.1 shows a flow diagram that outlines
the reasons for exclusion or withdrawing from the study.

Figure 15.1: Flow diagram of the participants that were excluded or dropped out.

15.2 Methods

15.2.3 Group allocation

Stratified group allocation was applied, stratified by type of smartphone and gender. As
the Active2Gether app only runs on Android, iPhone users were automatically assigned
to the Fitbit condition, while Android phone users were randomly allocated to one of two
A2G conditions after stratification by gender. The aim was to divide men and women with
an Android phone equally over the two A2G conditions. This was done by using a 1:1
ratio applied to the order of registration. Randomization of Android users after gender
stratification was performed before the participants visited the research facilities.

15.2.4 Intervention

As described above, the participants were allocated to one of the three conditions: (1) the
A2G-Full condition, (2) the A2G-Light condition, and (3) the Fitbit condition.

The participants in the A2G-Full condition received an Android app that provided tai-
lored advice aiming to increase weekly levels of MVPA. To do so, participants were coached
on sports participation, taking the stairs or active transport. Every week, the participants
were asked to choose their coaching domain and to set a weekly goal. Participants received
a suggestion for a coaching domain and a weekly goal based on their previous behavior, but the final decision was up to the user. The participants received a Fitbit One activity tracker that could be synced to the app and allowed the participants to monitor their PA behavior. Lastly, the app sent (daily) coaching messages addressing relevant behavioral determinants, i.e. self-efficacy, outcome expectations, intentions, satisfaction, barriers, and self-regulation skills. The content of the messages was tailored to the user’s behavioral determinants, occupational status, location (i.e., work or university) and weather. Lastly, the app displayed the activity data of the participant, including a graph displaying the activity data of six other participants, preferably friends. The graph with the activity data of others ranked the participants based on their weekly step activity and based on the user’s preferences for social comparison, i.e. upward or downward comparison. Detailed information on the development and the technical design of the Active2Gether intervention can be found elsewhere (Klein et al., 2017; Middelweerd, te Velde, et al., 2018).

The participants in the A2G-Light condition received a slimmed-down version of the A2G-Full app. Similar to the A2G-Full condition, the participants received a Fitbit One that could be synced to the app and allowed the participants to monitor their PA behavior. Also, activity data of six other participants was shown in the same way as in the A2G-Full condition. However, this variant of the Active2Gether app did not send tailored coaching messages.

The participants in the Fitbit condition only received a Fitbit One and the Fitbit app. The Fitbit app is a commercially available app – compatible with iPhones and Android phones – that enabled participants to monitor their step activity and to set activity goals (i.e., goals for number of steps and flights of stairs). Participants did not receive the weekly emails (with a weekly summary of the progress and congratulations on earning badges) that Fitbit sends to their users.

15.2.5 Procedure

Three rounds of assessments were conducted: at baseline, 6-week follow-up (mid-trial) and after completion of the 12-week intervention period. For the majority of the participants, the post-intervention measurement was delayed because of absence during the summer holidays. Participants completed an online questionnaire at all points and wore an ActiGraph accelerometer at baseline and post-intervention follow-up providing objective measurements on levels of PA.

After registering through the Active2Gether website, participants received an email providing detailed information about the study. Participants were asked to visit the research facilities once for an intake of about one hour. During the intake, participants again received detailed information about the study, they signed an informed consent, completed the baseline survey, installed the app(s) that were needed and received a Fitbit One. To complete the baseline measurements, participants were asked to wear an ActiGraph accelerometer for one week to objectively assess their baseline PA levels. After six weeks, participants received an email inviting them to complete the follow-up online questionnaire. At the end of the study, after twelve weeks, participants were asked to complete the final online questionnaire and to wear the ActiGraph accelerometer for another week. The participants did not have to visit the research facilities for the 6-week and post-intervention follow-up assessments: after six weeks, the participants received an email with a link to the 6-week follow-up questionnaire and after twelve weeks, participants received an email with a link
to the post-intervention follow-up questionnaire and were asked to briefly meet one of the researchers in Amsterdam or Utrecht for handing over the ActiGraph and Fitbit devices. Participants who were not able to meet the researchers in person returned the ActiGraph and Fitbit by mail.

Participants \(N_{\text{baseline}} = 13\) (A2G-Full = 2, A2G-Light = 2, Fitbit = 9), \(N_{\text{post-intervention}} = 14\) (A2G-Full = 0, A2G-Light = 3, Fitbit = 11) with insufficient ActiGraph data were asked to wear the accelerometer for another week. After completing the post-intervention follow-up assessment and returning the devices, the participants received a voucher of 20 euros as incentive for participating and 5 additional euros for each participant they brought into the study, ranging from 0–15 additional euros.

15.2 Methods

### 15.2.6 Measurements

#### Physical activity

PA was assessed using two different assessment methods. The ActiGraph accelerometer was used to objectively measure levels of PA to assess intervention effects. The Fitbit One also assesses PA objectively and was primarily used to allow participants to (self-)monitor their PA behavior, but the data were also used to explore possible intervention effects.

Baseline and post-intervention follow-up measurements were conducted using the ActiGraph GT3X+ \((N = 8)\) and ActiGraph wGT3XBT \((N = 32)\) (ActiGraph Inc, USA). The ActiGraphGT3X has a moderate validity and high reliability and is commonly used to assess PA in daily life (Anastasopoulou et al., 2014; Jarrett et al., 2015; J.-M. Lee et al., 2014). The ActiGraph is a three-axial accelerometer that is able to convert accelerations to step counts. The sampling rate was set at 100Hz and afterwards data was aggregated to 1-minute epochs. Participants were instructed to wear the accelerometer on the right hip using an elastic belt for seven consecutive days during waking hours. Furthermore, they were instructed to remove the accelerometer during water activities and sleep. The accelerometer was set up with the specific information – gender, age, height and weight – of the participant.

Choi’s definitions and the “Physical activity” R-package were used to identify non-wear time (e.g., periods of consecutive strings of zero’s for at least 90 minutes; the time window for detecting and handling artefactual movement was set the default at 2 minutes). Interruptions up to 100 counts per minute within the string of zero’s were filtered out (L. Choi et al., 2011).

Troiano’s definitions were used to calculate time spent per activity level using the vertical counts of the ActiGraph; sedentary (<100 counts/minute), light (100–2,019 counts/minute), moderate (2,020–5,998 counts/minute), vigorous (≥5,999 counts/minute) and MVPA (≥2,020 counts/minute) physical activities (Troiano et al., 2008). To adjust for wear time, weekly minutes of MVPA – the sum of all minutes spent in MVPA during the assessment week – was divided by wear time resulting in an average number of MVPA per day during the assessment week.

Participants were asked to wear a Fitbit One during twelve weeks to (self-)monitor their PA behavior. The Fitbit One (Fitbit Inc., San Francisco, CA) is a lightweight tri-axial accelerometer with a built-in altitude monitor. The Fitbit One assesses the step activity, active minutes, number of floors ascended, distance walked and number of calories burned. The Fitbit One can be considered a valid device to assess daily step activity and to assess step activity using smaller time epochs and thus can be used for real-time minute-by-minute self-monitoring, although an overestimation of 677 steps per day by the Fitbit should be
Chapter 15. Exploring use and effects of Active2Gether

taken into account (Ferguson et al., 2015; Gomersall et al., 2016; Middelweerd, Ploeg, et al., 2017; Rosenberger et al., 2016). However, the Fitbit is less suitable for providing instant real-time feedback and daily feedback for PA intensity levels (i.e., moderate, vigorous or MVPA) (Middelweerd, Ploeg, et al., 2017). As there is no algorithm to define non-wear time for the Fitbit data, daily steps <1,000 were treated as non-wear time (J. Choi et al., 2016; Craig et al., 2010; Mutrie et al., 2012). Thus, only days with ≥1,000 steps were included when Fitbit data were used to assess intervention effects and to assess levels of engagement.

**Behavioral determinants**

Behavioral determinants were assessed with an online questionnaire at baseline, 6-week follow-up and post-intervention follow-up. Questionnaires that were used to assess the behavioral determinants were mainly based on existing and previously validated questionnaires.

**Outcome expectations** PA outcome expectations were assessed with six items using a 4-point Likert scale (‘I do not agree at all’ (1) – ‘I totally agree’ (4)). The statements captured expected outcome of PA with respect to health, appearance, weight, feeling fit, relaxation and stress relief (Resnick et al., 2000). A sum score (range = 6–24) was computed for each time point (Cronbach’s α = 0.694–0.830).

**Self-efficacy** Self-efficacy for PA was assessed with thirteen items using a 5-point Likert scale (‘I know I can’t do it’ (1) – ‘I am sure I can do it’ (5)). The questionnaire was developed by Sallis, Pinski, et al. (1988) and translated into Dutch and used by Van Sluijs et al. (2004). A sum score (range = 13–65) was computed for each time point (Cronbach’s α = 0.797–0.883).

**Barriers** Barriers for sports participation (N = 12), active transport (N = 7) and taking the stairs (N = 4) were assessed using a 5-point Likert scale (‘Never’ (1) – ‘Very often’ (5)) (Frank et al., 2009; Sallis, Saelens, et al., 2009). The list of barriers that was assessed was based on an existing questionnaire and previous focus group discussions with the target population (Sallis, 2010). A sum score was computed summing the mean values of the three types of barriers – barriers for sports participation, active transport and taking the stairs – (range = 3–15) for each time point (Cronbach’s α = 0.620–0.717).

**Intention** Intentions were assessed with three items using a 5-point Likert scale (‘Very certainly not’ (1) – ‘Very certainly yes’ (5)). Questions assessed the intentions to be physically active within the next week/month/6 months (Frank et al., 2009; Sallis, Saelens, et al., 2009). For the analysis, intentions to be physically active within the next month and the next 6 months were used.

**Social norm** Injunctive and descriptive social norms were assessed, where injunctive norms refer to the perceptions of what others think you are supposed to do and descriptive norms refer to the perceptions of what others do (Kormos et al., 2015; Reno et al., 1993). Injunctive social norm was assessed with three items stated as “My sibling(s)/fellow students/friends think that I should be sufficiently physically active”. A 5-point Likert Scale
Methods

Self-regulation skills Self-regulation skills were assessed with seven items assessing exercise planning and scheduling, and how the user keeps track of his/her activity and self-determined goals (Rovniak et al., 2002). A 5-point Likert scale (‘Never’ (1) – ‘Very often’ (5)) was used. A sum score (range = 7–35) was computed for each time point (Cronbach’s $\alpha = 0.651–0.697$).

Satisfaction Satisfaction was assessed using one item stating “How satisfied are you with respect to how physically active you are on a scale from 0 to 10?”.

Long-term goals Satisfaction was assessed using one item stating “How motivated are you to be (more) physically active on a scale from 0 to 10?”.

Engagement and usability

Engagement with the intervention was assessed using number of coaching messages read — only for the A2G-Full condition — and Fitbit usage. As all participants were asked to wear the Fitbit during the intervention, we used the number of valid days the Fitbit was worn during 12 weeks (i.e., 84 days).

A purpose-designed feedback questionnaire was used to examine the usability of the intervention. Users’ previous experiences with apps or activity trackers, self-reported usage of the A2G app, several aspects of user-satisfaction — including having encountered technical problems with the A2G or Fitbit app — were assessed at post-intervention follow-up.

Previous experiences with apps were assessed with a single question (“Did you have previous experience with PA apps prior to the current study?”) with three response options (‘Yes, I use a PA app – ‘Yes, I used to use a PA app, but now I don’t’ – ‘No, I have no previous experience with PA apps’). For the analyses, the variable was dichotomized (‘Yes, have previous experiences’ – ‘No, I don’t have any previous experience’).

Previous experiences with activity trackers were assessed with a single question (“Did you have previous experience with activity trackers prior to the current study?”) with three response options (‘Yes, I use a activity tracker’ – ‘Yes, I used to make use of a activity tracker, but now I don’t’ – ‘No, I have no previous experience with activity trackers’). For the analyses, the variable was dichotomized (‘Yes, have previous experiences’ – ‘No, I don’t have any previous experience’).

Usage of the A2G app was assessed for the two A2G conditions using a single question (“How often did you used the Active2Gether app?”), with an 8-point Likert scale (‘Multiple times per day’ (1) – ‘Never’ (8)). For the analyses, the variable was dichotomized (‘Multiple times per day’, ‘Once per day’, and ‘Multiple times per week’ were coded as 1, whereas...
the options ‘Once per week’, ‘Multiple times per month’, ‘Once per month’, ‘Rarely’ and ‘Never’ were coded as 0).

Participants were asked how satisfied they were with the app they used (either one of the two versions of the A2G app or the Fitbit app). A 7-point Likert scale was used to assess level of agreement with the statement “I am pleased with the app” (“I do not agree at all” (1) – ‘I completely agree’ (7)). For the analyses, the variable was dichotomized (“I do not agree at all” – ‘Neutral’ were coded as 0 and ‘I somewhat agree’ – ‘I completely agree’ were coded as 1).

Participants were asked whether they experienced technical problems with the app they used by asking level of agreement with the statement “I experienced technical problems with the app” on a 7-point Likert scale (“I do not agree at all” (1) – ‘I completely agree’ (7)). For the analyses, the variable was dichotomized (“I do not agree at all” – ‘Neutral’ were coded as 0 and ‘I somewhat agree’ – ‘I completely agree’ were coded as 1).

Demographics
Information on age, gender and type of smartphone (iPhone/Android phone) were assessed during registration through the Active2Gether website. Height (self-report), weight (self-report) and being a student (yes/no) were assessed at baseline during the intake session. Height and weight were used to calculate the Body Mass Index (BMI, kg/m$^2$).

15.2.7 Sample size
We used the G*Power software (Faul et al., 2009) and calculated the required sample size for a design with three groups (F-test, ANOVA). As input, we used an effect size of 0.25, which is considered a medium effect size, an alpha of 5% and a power of 80%. Based on these considerations, approximately 53 participants per group were required. Therefore, we aimed to include 159–200 participants.

15.2.8 Statistical analyses
**Intervention effects**
Primary outcome variables were levels of PA at post-intervention follow-up (i.e., mean minutes of MVPA per day and mean steps per day), as measured by the ActiGraph. Secondary outcome variables were scores of behavioral determinants (i.e., outcome expectations, self-efficacy, barriers, social norm, intentions, self-regulation skills, satisfaction and long-term goals) at post-intervention follow-up. Descriptive analyses were conducted for all variables; means and standard deviations (continuous variables) or proportions (categorical variables) were conducted to test for differences between groups at baseline.

For the analyses, the two intervention groups – the A2G-Full and A2G-Light conditions – were compared against a commercially available app, i.e. the Fitbit app. This comparison will provide information on the efficacy of the A2G conditions compared to an existing ‘usual care’ app. In addition, this design gives the opportunity to compare the two A2G conditions. As the difference between these two conditions is the inclusion or absence of the coaching, this comparison will provide information on the efficacy of the coaching part of the A2G app. As participants with an iPhone were automatically assigned to the Fitbit condition and could not be randomly assigned to one of the A2G conditions, additional analyses were conducted to test for differences in intervention effects between the two A2G
Conditions only. Furthermore, there were large differences in the duration of time between start of the intervention and the post-intervention follow-up measurements (i.e., between 12 and 24 weeks). Therefore, additional analyzes were conducted using the Fitbit data for week 1 and week 12 to examine the efficacy of the intervention to increase weekly number of steps at exactly 12-week follow-up.

In a first step, analyses were conducted to examine the efficacy of the intervention to increase weekly minutes of MVPA and weekly number of steps at post-intervention follow-up. To do so, associations were analyzed using linear regression analyses with the intervention conditions entered as dummy variables – the Fitbit condition was coded as the reference group – adjusting for baseline PA (i.e., minutes of MVPA or number of steps) and time between baseline and post-intervention follow-up. In a second step, analyses were conducted to examine the efficacy of the intervention to improve relevant behavioral determinants at post-intervention follow-up. Linear regression analyses with the different determinants as dependent variables, while adjusting for baseline scores and time between baseline and post-intervention measurements, were used. For dichotomous determinant variables (intentions and satisfaction), logistic regression analyses were conducted. These variables were dichotomized as the residuals from the linear regression analyses when using the continuous variables were not normally distributed. All analyses were checked for outliers ($\geq 3\times$standard deviation of the residuals), and when necessary sensitivity analyses were conducted without outliers. The final analyses were conducted without outliers. Four models were run for each outcome variable (i.e., levels of PA and scores of behavioral determinants): (0) a minimal adjusted model (only adjusted for baseline values and time between baseline and post-intervention measurements), (1) a model additionally adjusted for BMI, (2) a model additionally adjusted for student status, and (3a) a model additionally adjusted for BMI and student status (for the intervention effects on PA only) and (3b) a model adjusted for BMI and meeting the PA guidelines (for the intervention effects on behavioral determinants only). Due to the small sample size, no further potential confounders were added to the final model.

**Levels of engagement and usability**

Additional exploratory analyses were conducted to evaluate how the users rated various aspects of the app they had used.

Descriptive analyses were conducted for previous experiences with apps or activity trackers, usage of the A2G app, satisfaction with the A2G or Fitbit app and having encountered technical problems. Chi-squared tests were used to examine differences between groups.

**Non-response analyses**

Non-response analyses were conducted to examine differences between those who had no PA data (assessed with the ActiGraph) for baseline and post-intervention follow-up, those who only had baseline PA data and those who had valid data at both baseline and post-intervention follow-up. No significant differences between the groups were found with respect to age, BMI, being a student and all secondary outcome variables.

Analyses were conducted in STATA 14 (StataCorp. 2015. Stata Statistical Software: Release 14. College Station, TX: StataCorp LP).
15.3 Results

15.3.1 Baseline characteristics

A total of 104 participants (83 females) attended the intake session and completed the baseline questionnaire and 98 participants had valid PA data for the baseline week. On average, participants were 23.4 years old, had a BMI of 22.8 kg/m², 69.2 percent student, 79.8% were female and 31.7% had previous experiences with PA apps. At baseline, participants were moderately to vigorously active for 39.9 minutes per day on average. No significant differences between the A2G conditions and Fitbit condition were found for the baseline characteristics. An overview of the participants’ characteristics is presented in Table 15.1.

Table 15.1: Baseline characteristics of participants in the Active2Gether-Full, Active2Gether-Light and Fitbit condition.

<table>
<thead>
<tr>
<th></th>
<th>Overall</th>
<th>A2G-Full</th>
<th>A2G-Light</th>
<th>Fitbit</th>
<th>p-value¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>N (%)</td>
<td>104 (100)</td>
<td>28 (26.9)</td>
<td>27 (26.0)</td>
<td>49 (47.1)</td>
<td>0.959</td>
</tr>
<tr>
<td>Female [N (%)]</td>
<td>83 (79.8)</td>
<td>21 (75.0)</td>
<td>23 (85.2)</td>
<td>39 (79.6)</td>
<td>0.456</td>
</tr>
<tr>
<td>Age ± SD [years]</td>
<td>23.4 ±3.0</td>
<td>23.7 ±3.2</td>
<td>22.8 ±2.8</td>
<td>23.5 ±3.1</td>
<td>0.758</td>
</tr>
<tr>
<td>Body Mass Index ± SD [kg/m²]</td>
<td>22.8 ±3.4</td>
<td>23.8 ±3.7</td>
<td>22.6 ±3.3</td>
<td>22.3 ±3.3</td>
<td></td>
</tr>
<tr>
<td>Student [N (%)]</td>
<td>72 (69.2)</td>
<td>17 (60.7)</td>
<td>22 (81.5)</td>
<td>33 (67.3)</td>
<td>0.694</td>
</tr>
<tr>
<td>Android phone [N (%)]</td>
<td>57 (54.8)</td>
<td>28 (100)</td>
<td>27 (100)</td>
<td>3 (6.1)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Experiences PA apps [Yes; N(%)]</td>
<td>33 (31.7%)</td>
<td>8 (28.6%)</td>
<td>7 (25.9%)</td>
<td>18 (36.7%)</td>
<td>0.463</td>
</tr>
<tr>
<td>MVPA ± SD [minutes]²</td>
<td>267.7 ±163.8</td>
<td>234.9 ±107.4</td>
<td>258.8 ±202.2</td>
<td>293.1 ±168.5</td>
<td>0.148</td>
</tr>
<tr>
<td>Steps Actigraph ± SD [steps]²</td>
<td>8177.6 ±3272.0</td>
<td>7519.3 ±2884.3</td>
<td>7847.8 ±3546.6</td>
<td>8770.4 ±3307.5</td>
<td>0.099</td>
</tr>
<tr>
<td>Steps Fitbit ± SD [steps]²</td>
<td>9008.9 ±3272.0</td>
<td>8179.9 ±2884.3</td>
<td>9190.7 ±3546.6</td>
<td>9535.5 ±3307.5</td>
<td>0.296</td>
</tr>
<tr>
<td>Weartime Actigraph per day ± SD [minutes/day]</td>
<td>861.9 ±61.3</td>
<td>861.3 ±50.5</td>
<td>865.0 ±58.8</td>
<td>860.5 ±69.6</td>
<td>0.836</td>
</tr>
<tr>
<td>Time between baseline and post-intervention follow-up ± SD [days]</td>
<td>103.4 ±19.5</td>
<td>106.5 ±23.9</td>
<td>109.0 ±21.6</td>
<td>97.7 ±12.6</td>
<td>0.564</td>
</tr>
</tbody>
</table>

Note. Means ± standard deviation (SD) or frequencies (N) and percentages (%) are presented.
¹ Pearson’s Chi-square test with p-value for frequencies and one-way ANOVA for means for differences between A2G-Full and A2G-Light versus Fitbit.
² Baseline minutes of moderate-vigorous physical activity (MVPA); number of steps and weartime were summed for the week and divided by the number of valid days to adjust for weartime.
15.3 Results

15.3.2 Intervention effects on physical activity

PA data (assessed with the ActiGraph) for baseline and post-intervention follow-up was available for 88 participants ($N_{A2G-Full} = 25$, $N_{A2G-Light} = 25$, $N_{Fitbit} = 38$). Table 15.2 shows the means and standard deviations for the outcome measurements for baseline and post-intervention follow-up.

Table 15.2: Characteristics (means ± standard deviation)) at baseline (T1), 6-week follow-up (T2) and post-intervention follow-up (T3).

<table>
<thead>
<tr>
<th></th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minutes of MVPA/day</td>
<td>35.2 ± 15.3</td>
<td>39.7 ± 17.5</td>
<td>38.6 ± 28.1</td>
<td>42.1 ± 20.5</td>
<td>43.5 ± 23.5</td>
<td>44.8 ± 29.5</td>
<td>44.8 ± 15.3</td>
<td>44.8 ± 17.5</td>
<td>44.8 ± 28.1</td>
</tr>
<tr>
<td>ActiGraph1 Steps/day</td>
<td>7519.3 ± 2884.3</td>
<td>7681.5 ± 2463.8</td>
<td>7847.8 ± 3546.6</td>
<td>8366.0 ± 2637.0</td>
<td>8770.4 ± 3070.5</td>
<td>9367.6 ± 4537.3</td>
<td>9367.6 ± 2884.3</td>
<td>9367.6 ± 2463.8</td>
<td>9367.6 ± 3546.6</td>
</tr>
<tr>
<td>Fitbit2 Steps/day</td>
<td>8179.9 ± 2415.9</td>
<td>9392.8 ± 3275.7</td>
<td>9190.7 ± 4610.6</td>
<td>9567.1 ± 3152.2</td>
<td>9535.5 ± 3878.0</td>
<td>9968.1 ± 4506.2</td>
<td>9968.1 ± 2415.9</td>
<td>9968.1 ± 3275.7</td>
<td>9968.1 ± 4610.6</td>
</tr>
<tr>
<td>Self-efficacy (13–65)</td>
<td>42.4 ± 7.6</td>
<td>41.8 ± 7.3</td>
<td>42.8 ± 8.2</td>
<td>40.9 ± 8.3</td>
<td>42.0 ± 6.2</td>
<td>44.5 ± 7.8</td>
<td>44.5 ± 7.6</td>
<td>44.5 ± 6.2</td>
<td>44.5 ± 7.8</td>
</tr>
<tr>
<td>Outcome expectations (6–24)</td>
<td>20.3 ± 2.3</td>
<td>19.7 ± 3.1</td>
<td>19.8 ± 2.4</td>
<td>20.4 ± 3.1</td>
<td>19.3 ± 2.8</td>
<td>20.4 ± 2.6</td>
<td>20.4 ± 2.3</td>
<td>20.4 ± 3.1</td>
<td>20.4 ± 2.8</td>
</tr>
<tr>
<td>Social norm injunctive (3–15)</td>
<td>10.7 ± 2.3</td>
<td>10.9 ± 3.0</td>
<td>10.3 ± 2.7</td>
<td>9.9 ± 2.4</td>
<td>10.2 ± 3.6</td>
<td>9.7 ± 2.6</td>
<td>9.7 ± 2.3</td>
<td>9.7 ± 3.0</td>
<td>9.7 ± 2.6</td>
</tr>
<tr>
<td>Social norm descriptive (4–20)</td>
<td>14.6 ± 2.7</td>
<td>14.9 ± 2.3</td>
<td>14.8 ± 2.7</td>
<td>13.4 ± 2.8</td>
<td>13.2 ± 3.5</td>
<td>13.2 ± 2.9</td>
<td>13.2 ± 2.7</td>
<td>13.2 ± 2.6</td>
<td>13.2 ± 2.7</td>
</tr>
<tr>
<td>Intentions in 1 month (1–5)</td>
<td>4.1 ± 0.7</td>
<td>3.6 ± 1.0</td>
<td>3.5 ± 1.0</td>
<td>3.7 ± 1.1</td>
<td>3.1 ± 1.3</td>
<td>3.4 ± 1.2</td>
<td>3.4 ± 0.8</td>
<td>3.4 ± 1.0</td>
<td>3.4 ± 1.0</td>
</tr>
<tr>
<td>Intentions in 6 months (1–5)</td>
<td>4.4 ± 0.8</td>
<td>3.8 ± 0.9</td>
<td>3.9 ± 0.9</td>
<td>4.1 ± 1.1</td>
<td>3.7 ± 1.1</td>
<td>3.7 ± 0.9</td>
<td>3.7 ± 0.8</td>
<td>3.7 ± 0.9</td>
<td>3.7 ± 0.9</td>
</tr>
<tr>
<td>Barriers (3–15)</td>
<td>8.4 ± 1.7</td>
<td>8.3 ± 1.9</td>
<td>8.2 ± 1.9</td>
<td>7.9 ± 1.5</td>
<td>7.8 ± 1.4</td>
<td>7.9 ± 1.5</td>
<td>7.7 ± 1.5</td>
<td>7.7 ± 1.7</td>
<td>7.7 ± 1.7</td>
</tr>
<tr>
<td>Self-reg. skills (5–25)</td>
<td>18.8 ± 4.3</td>
<td>19.4 ± 3.2</td>
<td>19.3 ± 3.8</td>
<td>19.0 ± 5.5</td>
<td>20.1 ± 5.5</td>
<td>19.3 ± 5.0</td>
<td>21.0 ± 4.6</td>
<td>21.0 ± 4.4</td>
<td>21.0 ± 4.4</td>
</tr>
<tr>
<td>Satisfaction (0–10)</td>
<td>5.5 ± 1.8</td>
<td>6.0 ± 1.7</td>
<td>5.9 ± 2.0</td>
<td>5.5 ± 1.8</td>
<td>5.5 ± 1.7</td>
<td>5.7 ± 1.7</td>
<td>6.0 ± 1.9</td>
<td>6.0 ± 1.9</td>
<td>6.0 ± 1.9</td>
</tr>
<tr>
<td>Long-term goals (0–10)</td>
<td>7.0 ± 2.0</td>
<td>6.9 ± 1.5</td>
<td>7.2 ± 1.5</td>
<td>6.8 ± 1.8</td>
<td>6.7 ± 1.8</td>
<td>7.3 ± 1.2</td>
<td>7.1 ± 1.4</td>
<td>7.1 ± 1.4</td>
<td>7.1 ± 1.4</td>
</tr>
</tbody>
</table>
Chapter 15. Exploring use and effects of Active2Gether

1 Minutes of moderate-vigorous physical activity (MVPA) and number of steps per day assessed with the ActiGraph, at post-intervention follow-up.
2 Number of steps per day assessed with the Fitbit, after 12 weeks.
3 Behavioral determinant (range of sum score).

Regression analyses (based on model 3a: adjusted for BMI and student status) showed no significant intervention effects of the A2G-Full and A2G-Light conditions on levels of PA as compared to the Fitbit condition. Effect sizes were small, and smallest for the A2G-Full condition ($B = 1.2$, $95\% CI = [-8.7,11.1]$). Table 15.3 shows the results of the regression analyses.

Table 15.3: Results of the regression analyses (regression coefficients ($B$) with 95% confidence intervals ($95\% CI$)) to evaluate the intervention effects of the A2G-Full and A2G-Light condition on levels of physical activity at post-intervention follow-up as compared to the Fitbit condition.

### Average minutes of moderate-vigorous physical activity per day

<table>
<thead>
<tr>
<th>Model</th>
<th>$B$ (95% CI)</th>
<th>$B$ (95% CI)</th>
<th>$B$ (95% CI)</th>
<th>$B$ (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fitbit</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>A2G-Full</td>
<td>0.82 [-8.82,10.46]</td>
<td>1.16 [-8.73,11.04]</td>
<td>0.92 [-8.70,10.54]</td>
<td>1.20 [-8.66,11.07]</td>
</tr>
</tbody>
</table>

### Average number of steps per day

<table>
<thead>
<tr>
<th>Model</th>
<th>$B$ (95% CI)</th>
<th>$B$ (95% CI)</th>
<th>$B$ (95% CI)</th>
<th>$B$ (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fitbit</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>A2G-Full</td>
<td>-577.42 [-1913.68,758.85]</td>
<td>-387.88 [-1742.21,966.44]</td>
<td>-575.76 [-1918.33,766.82]</td>
<td>-388.95 [-1750.20,972.31]</td>
</tr>
<tr>
<td>A2G-Light</td>
<td>-128.54 [-1447.23,1190.16]</td>
<td>-45.56 [-1361.36,1270.24]</td>
<td>-70.37 [-1413.70,1272.96]</td>
<td>5.21 [-1334.82,1345.25]</td>
</tr>
</tbody>
</table>

*Note.* Linear regression analyses are presented with regression coefficient ($B$) [95% confidence interval], and all analyses were adjusted for levels of physical activity at baseline and time between baseline and post-intervention follow-up.

Model 0: $y = B_0 + B_1 \times$ Physical activity at post-intervention + $B_2 \times$ Physical activity at baseline + $B_3 \times$ Time until post-intervention follow-up (#days)

Model 1: Model 0 + $B_4 \times$ BMI

Model 2: Model 0 + $B_4 \times$ Student (yes/no)

Model 3: Model 0 + $B_4 \times$ BMI + $B_5 \times$ Student (yes/no)

Additional regression analyses using the ActiGraph data (with adjustment for baseline PA, intervention duration, BMI and being a student) showed a group difference of 2.76 minutes of MVPA per day between the A2G-Full and A2G-Light condition in favor of the A2G-Light condition (Appendix 1). The same regression analyses, but using the Fitbit data at baseline and 12-week follow-up instead, showed a group difference of 533.51 steps per day between A2G-Full and A2G-Light in favor of the A2G-Light condition (Appendix 2).
### 15.3 Results

#### 15.3.3 Intervention effects on behavioral determinants

Survey data for baseline and 12-weeks follow-up was available for 92 participants ($N_{A2G-Full} = 24$, $N_{A2G-Light} = 23$, $N_{Fitbit} = 45$). Table 15.2 shows the means and standard deviations for behavioral determinant scores for baseline, 6-week and post-intervention follow-up.

Linear and logistic regression analyses for the intervention effects on the behavioral determinants at post-intervention follow-up showed no significant intervention effects of the A2G-Full and A2G-Light conditions as compared to the Fitbit condition. For all analyses, small effect sizes were found except for intentions to be physically active within 6 months (OR$_{A2G-Full} = 0.76$, 95%CI = [-0.53,2.05]; OR$_{A2G-Light} = 1.27$, 95%CI = [-0.07,2.62]). Table 15.4 shows the results of the regression analyses.

---

Table 15.4: Results of the linear and logistic regression analyses (regression coefficients (B) or odds ratios (OR) with 95% confidence intervals (95%CI)) to evaluate the intervention effects of the A2G-Full and A2G-Light conditions on behavioral determinants at post-intervention follow-up as compared to the Fitbit condition.

<table>
<thead>
<tr>
<th>Outcome measurement</th>
<th>Condition</th>
<th>Model 0</th>
<th>Model 1: BMI</th>
<th>Model 2: Student</th>
<th>Model 3: BMI-PA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Self-efficacy</strong> B [95% CI]</td>
<td>Fitbit</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>A2G-Full</td>
<td>0.03</td>
<td>0.74</td>
<td>0.14</td>
<td>0.62</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[-2.88, 2.94]</td>
<td>[-2.15,3.63]</td>
<td>[-2.73,3.00]</td>
<td>[-2.24,3.49]</td>
<td></td>
</tr>
<tr>
<td>A2G-Light</td>
<td>-1.52</td>
<td>-1.28</td>
<td>-0.92</td>
<td>-1.54</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[-4.40,1.36]</td>
<td>[-4.08,1.52]</td>
<td>[-3.81,1.98]</td>
<td>[-4.34,1.25]</td>
<td></td>
</tr>
<tr>
<td><strong>Outcome expectations</strong> B [95% CI]</td>
<td>Fitbit</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>A2G-Full</td>
<td>0.44</td>
<td>0.43</td>
<td>0.40</td>
<td>0.41</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[-0.59,1.47]</td>
<td>[-0.63,1.50]</td>
<td>[-0.61,1.41]</td>
<td>[-0.66,1.47]</td>
<td></td>
</tr>
<tr>
<td>A2G-Light</td>
<td>0.07</td>
<td>0.07</td>
<td>-0.12</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[-0.95,1.10]</td>
<td>[-0.97,1.11]</td>
<td>[-1.15,0.91]</td>
<td>[-1.02,1.07]</td>
<td></td>
</tr>
<tr>
<td><strong>Social norm</strong> descriptive B [95% CI]</td>
<td>Fitbit</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>A2G-Full</td>
<td>1.18</td>
<td>1.11</td>
<td>1.26</td>
<td>1.12</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.15,2.20]</td>
<td>[0.05,2.16]</td>
<td>[0.24,2.29]</td>
<td>[0.08,2.16]</td>
<td></td>
</tr>
<tr>
<td>A2G-Light</td>
<td>-0.14</td>
<td>-0.16</td>
<td>0.02</td>
<td>-0.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[-1.13,0.86]</td>
<td>[-1.16,0.84]</td>
<td>[-1.00,1.03]</td>
<td>[-1.05,0.94]</td>
<td></td>
</tr>
<tr>
<td><strong>Social norm</strong> injunctive B [95% CI]</td>
<td>Fitbit</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>A2G-Full</td>
<td>0.11</td>
<td>0.27</td>
<td>0.26</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[-1.64,1.85]</td>
<td>[-1.53,2.06]</td>
<td>[-1.47,2.00]</td>
<td>[-1.73,1.80]</td>
<td></td>
</tr>
</tbody>
</table>
### Chapter 15. Exploring use and effects of Active2Gether

<table>
<thead>
<tr>
<th></th>
<th>A2G-Light</th>
<th>Fitbit</th>
<th>Reference</th>
<th>OR [95%CI]</th>
<th>A2G-Light</th>
<th>Fitbit</th>
<th>Reference</th>
<th>OR [95%CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intentions 1 month</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OR [95%CI]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.45</td>
<td>-2.01</td>
<td>-2.05,1.18</td>
<td></td>
<td>-0.33</td>
<td>-1.97</td>
<td>-1.74,1.56</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[-0.79,1.42]</td>
<td>-0.80,1.44</td>
<td></td>
<td></td>
<td>[-0.83,1.39]</td>
<td>-0.80</td>
<td>-1.24,1.06</td>
<td></td>
</tr>
<tr>
<td><strong>Intentions 6 months</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OR [95%CI]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.01</td>
<td>-1.10</td>
<td>-1.10,1.12</td>
<td></td>
<td>-0.10</td>
<td>-1.24</td>
<td>-1.24,1.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.31</td>
<td>0.28</td>
<td>-0.83,1.39</td>
<td></td>
<td>0.32</td>
<td>0.32</td>
<td>-0.80,1.44</td>
<td></td>
</tr>
<tr>
<td><strong>Barriers B [95% CI]</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.01</td>
<td>-0.60</td>
<td>-0.77,0.37</td>
<td></td>
<td>-0.20</td>
<td>-0.54</td>
<td>-0.54,0.61</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.06</td>
<td>0.53</td>
<td>-0.55,0.54</td>
<td></td>
<td>0.16</td>
<td>0.41</td>
<td>-0.41,0.73</td>
<td></td>
</tr>
<tr>
<td><strong>Self-reg. skills B [95% CI]</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.78</td>
<td>-0.94</td>
<td>-0.96,2.56</td>
<td></td>
<td>0.80</td>
<td>-0.90</td>
<td>-0.90,2.54</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.01</td>
<td>-1.68</td>
<td>-1.69,1.72</td>
<td></td>
<td>0.02</td>
<td>-1.53</td>
<td>-1.53,1.93</td>
<td></td>
</tr>
<tr>
<td><strong>Satisfaction OR [95%CI]</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.37</td>
<td>-1.65</td>
<td>-1.47,1.22</td>
<td></td>
<td>-0.13</td>
<td>-1.52</td>
<td>-1.52,1.11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.72</td>
<td>-2.00</td>
<td>-1.99,0.63</td>
<td></td>
<td>-0.68</td>
<td>-0.43</td>
<td>-0.43,0.83</td>
<td></td>
</tr>
<tr>
<td><strong>Long-term goals B [95% CI]</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.13</td>
<td>-0.83</td>
<td>-0.62,0.77</td>
<td></td>
<td>0.08</td>
<td>-0.84</td>
<td>-0.84,0.56</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.46</td>
<td>-1.17</td>
<td>-1.06,0.32</td>
<td></td>
<td>-0.37</td>
<td>-1.24</td>
<td>-1.24,0.22</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Linear regression analyses are presented with regression coefficient (B) [95% confidence interval (95% CI)] and logistic regression analyses with odds ratio (OR) [95% confidence interval (95% CI)].
### 15.3 Results

Model 0: \( y = B_0 + B_1 \times \text{Physical activity at post-intervention} + B_2 \times \text{Physical activity at baseline} + B_3 \times \text{Time until post-intervention follow-up (#days)} \)

Model 1: Model 0 + \( B_4 \times \text{BMI} \)

Model 2: Model 0 + \( B_4 \times \text{Student (yes/no)} \)

Model 3: Model 0 + \( B_4 \times \text{BMI} + B_5 \times \text{Meeting PA guidelines at baseline (yes/no)} \)

#### 15.3.4 Levels of engagement and usability

For the A2G-Full condition, 1,429 messages were derived, 1,381 messages (i.e., 97% of the messages) were sent and 1,324 messages were successfully received. For five out of 24 users, a derived message was not sent at some point, which could indicate that the app was removed before the end of the study. For nine users, a sent message was not received by the phone, and one user did not receive any messages at all.

For participants in the A2G-Full and Fitbit condition, a decrease is seen (from day 1 to day 84 of the intervention) in the number of participants who recorded valid step activity (>1,000 steps per day) assessed with the Fitbit. At 6-week follow-up (i.e., after 42 days), 68% of the A2G-Full condition, 70% of the A2G-Light condition, and 51% of the Fitbit condition were still using the Fitbit. At 12-week follow-up (i.e., after 84 days), 50% of the A2G-Full condition, 74% of the A2G-Light condition, and 38% of the Fitbit condition were still using the Fitbit. Figure 15.2 shows the number of participants who logged step activity per intervention condition, and a steeper decrease is seen for the Fitbit condition relative to the two A2G conditions.

![Figure 15.2: Fitbit usage in percentage of number of participants per week.](image)

*Note.* The figure shows the percentage of participants who logged at least 1,000 steps per day per condition for 12 weeks (84 days).

The majority of the participants in the A2G-Full and A2G-Light conditions reported that they used the app at least several times per week or more frequently, 63% and 82% respectively (Figure 3). For the Fitbit condition, this held for 73% of the participants. Significant differences were found with respect to how satisfied the participants were with the app they used during the intervention. Majorities of participants in the two A2G
conditions were not satisfied with the app (A2G-Full = 67%, A2G-Light = 64%), whereas 22% of the participants in the Fitbit group were not satisfied with the Fitbit app. More participants in the two A2G conditions (A2G-Full = 54%, A2G-Light = 45%) experienced technical problems with the app compared to the Fitbit condition (23%). Table 15.5 shows the scores on the user evaluations.

A more detailed evaluation of the user experience of the Active2Gether intervention can be found elsewhere (Mollee et al., 2017).

Figure 15.3: Percentages of reported frequency of app usage during the intervention period per intervention group.


### 15.4 Discussion

This study aimed to evaluate whether two versions of the Active2Gether app – a tailored app-based intervention to promote PA – were more effective in increasing levels of PA among young adults than an existing self-monitoring app. The secondary aims of the study were to examine whether the intervention was effective in changing levels of relevant behavioral determinants of PA and how participants used and evaluated the app. No evidence for significant intervention effects on increased PA or more positive determinants of PA were found. The majority of the A2G app users used the app at least several times per week and was not satisfied with the app, and a substantial number of participants experienced technical problems.

The present study was originally designed as a randomized controlled trial with 159–200 participants and a follow-up measurement for all participants at 12 weeks, i.e. immediately after the envisioned intervention period. However, the study as conducted differed substantially from the original protocol.

First of all, the number of participants was lower than envisioned in the study protocol.
Table 15.5: User engagement and usability of the A2G-Full, A2G-Light and Fitbit app.

<table>
<thead>
<tr>
<th></th>
<th>Overall</th>
<th>A2G-Full</th>
<th>A2G-Light</th>
<th>Fitbit</th>
<th>(p)-value(^1)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fitbit usage</strong> [median percentage of days used (range)](^2)</td>
<td>88.1 (0–100)</td>
<td>86.3 (9.5–100)</td>
<td>94.6 (3.6–100)</td>
<td>83.9 (0–100)</td>
<td>0.134</td>
</tr>
<tr>
<td><strong>Previous experiences with PA apps [Yes; N (%)](^3)</strong></td>
<td>33 (36%)</td>
<td>8 (33%)</td>
<td>7 (32%)</td>
<td>18 (40%)</td>
<td>0.469</td>
</tr>
<tr>
<td><strong>Previous experiences with activity trackers [Yes; N (%)](^3)</strong></td>
<td>17 (19%)</td>
<td>6 (25%)</td>
<td>4 (18%)</td>
<td>7 (16%)</td>
<td>0.455</td>
</tr>
<tr>
<td><strong>Satisfied with the A2G or Fitbit app [N (%)](^4)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.000</td>
</tr>
<tr>
<td>Yes</td>
<td>36 (40%)</td>
<td>5 (21%)</td>
<td>5 (23%)</td>
<td>26 (58%)</td>
<td></td>
</tr>
<tr>
<td>Neutral</td>
<td>15 (16%)</td>
<td>3 (13%)</td>
<td>3 (14%)</td>
<td>9 (20%)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>40 (44%)</td>
<td>16 (67%)</td>
<td>14 (64%)</td>
<td>10 (22%)</td>
<td></td>
</tr>
<tr>
<td><strong>Experienced technical problems with the app [N (%)](^5)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.009</td>
</tr>
<tr>
<td>Yes</td>
<td>33 (37%)</td>
<td>13 (54%)</td>
<td>10 (45%)</td>
<td>10 (23%)</td>
<td></td>
</tr>
<tr>
<td>Neutral</td>
<td>3 (3%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>3 (7%)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>54 (60%)</td>
<td>11 (46%)</td>
<td>12 (55%)</td>
<td>31 (70%)</td>
<td></td>
</tr>
<tr>
<td><strong>App usage [N (% Often)](^5)</strong></td>
<td>76 (84%)</td>
<td>14 (63%)</td>
<td>18 (82%)</td>
<td>33 (73%)</td>
<td>0.695</td>
</tr>
</tbody>
</table>

\(^1\) The \(p\)-value is the result of a Chi-square test between A2G-Full and A2G-Light versus Fitbit.  
\(^2\) Percentage of days used = number of days the Fitbit was used (steps > 1000) / 84 days * 100.  
\(^3\) The score was dichotomized: Yes = ‘Yes, I’m currently using one’, ‘Yes, in the past’ and No = ‘No, no experience’.  
\(^5\) The score was dichotomized: Rarely = ‘Never’, ‘Rarely’, ‘Once a month’, ‘Multiple times per month’, ‘Once per week’ and Often = ‘Multiple times per week’, ‘Once a day’, ‘Multiple times per day’.  

Despite our efforts for participant recruitment, fewer people than expected were willing to participate due to lack of interests, lack of time and the perceived burden for the participants. Due to the small sample, the statistical power of the results is too low.

In addition, participants were not assigned to the three conditions based on true randomization. One reason for this is that the two versions of the Active2Gether app (A2G-Full and A2G-Light) were not available for iPhone users, so iPhone users were automatically assigned to the Fitbit condition. Also, the proportion of Android users who registered for the study was lower than expected, so — in order to maintain a balance between the three conditions — they were randomized over the two Active2Gether conditions, rather than
over all three conditions. Therefore, the study would ideally only include Android users, or the Active2Gether intervention should be implemented for iPhones as well.

Third, due to the difficulties with recruiting participants from the target population, the inclusion of the participants was spread over three months. Consequently, some participants were included just at the end of the academic year and thus the beginning of the summer holidays. As a result, the 12-week follow-up measurements were due in the middle of their summer holiday for the majority of the participants. Therefore, the post-intervention measures were delayed, and the time between the baseline and post-intervention follow-up varied widely among the participants.

Lastly, due to malfunction of the PA assessment with the ActiGraph, the baseline measurement had to be redone for a number of participants. Therefore, the baseline measurement of PA for some participants took place during the intervention, rather than at the start.

Despite these major violations of the original study protocol, we do want to discuss the results found in more detail, but this discussion should of course be read and interpreted keeping these differences between the study as designed and conducted in mind.

No statistically significant effects were found and the effect sizes were small: compared to the A2G-Light condition, the A2G-Full condition measured on average 2.76 minutes of MVPA less per day, thus 19.32 minutes of MVPA per week. Also, based on Fitbit registrations, the A2G-Light users took 533.51 more steps per day. Earlier evaluations of app-based interventions also reported mixed results, but the majority of the studies reported significant intervention effects relative to the control group. Those studies reported changes between -15.5% to 34.8% in PA in the intervention groups, of whom the majority evaluated the intervention effects at 8-week follow-up (Fukuoka et al., 2010; Glynn et al., 2014; Hebden et al., 2013; Maher et al., 2015). However, it should be noted that these studies differ with respect how they assess PA: one study used the ActiGraph, two studies used a pedometer to assess step activity, but did not report the validity of the instruments, three studies used self-reported measurements and of these three only one study used a validated questionnaire, and three studies used the built-in smartphone accelerometer to assess PA, but did not report the validity. Because of the different assessment methods used in the different studies, it is difficult to compare the results. Furthermore, the participants in the current study were already active and on average met the guidelines of 30 minutes MVPA per day, whereas the baseline PA levels in other studies were much lower. As it might be difficult to increase weekly levels of MVPA in an already active group, this might partially explain the lack of intervention effect.

The secondary aim of the current study was to examine whether the A2G-Full intervention effectively changes scores in behavioral determinants that were included in the theoretical framework. No significant intervention effects were seen in changes in scores, indicating that sending the tailored coaching messages did not lead to changes in the behavioral determinants. So far, the only study examining the effects of the Fitbit app on social cognitive behavioral determinants showed no significant changes in behavioral determinants after twelve weeks (J. Choi et al., 2016). Other studies making use of apps to change PA made use of self-monitoring features, motivational messages and prompts and offered challenges to increase the levels of PA as well, but did not examine changes in behavioral determinants. However, it remains unclear whether these interventions successfully changed the underlying and relevant behavioral determinants. Therefore, future research is needed to examine whether motivational messages, prompts, challenges and social support features
can be used to change behavioral determinants. To do so, a more iterative assessment of the determinants during the intervention is needed, as done in the Active2Gether intervention. Consequently, this knowledge will contribute to further tailor and personalize app-based interventions to increase levels of PA.

Although 96 participants (92.3%) participated at the post-intervention follow-up assessment, lower rates of engagement with the Fitbit were seen after twelve weeks, especially for the Fitbit condition. However, the overall engagement with the Fitbit was high (median= 88 percent of the days). This is in line with the self-reported app usage: the majority of the participants reported that they used the appointed app several times per week or more throughout the intervention. However, about only 21% and 23% of the participants in the A2G-Full and A2G-Light condition were satisfied with the app, while 58% was satisfied with the Fitbit app in the control condition. Those low scores might be related with the high rates of technical problems that the participants in the A2G conditions encountered. Moreover, it should be noted that the Fitbit used to monitor daily activity did not automatically synchronize with the A2G apps. The participants in the two A2G conditions needed to synchronize the Fitbit through the Fitbit app or Fitbit website. This an additional step can be a burden for the users of the A2G apps and might be more prone to technical errors. The A2G-Full app sent the weekly questions and coaching messages via push messages and the users could only access the app after reading the unread messages. Participants in the A2G-Light condition only received daily or weekly questions via push messages. A more detailed evaluation of the participants’ satisfaction in the usability of the app is published elsewhere (Mollee et al., 2017).

To summarize, the current study showed no significant intervention effects in changes in levels of PA and behavioral determinants compared to the active control groups. Because the study as conducted differed substantially from the study as designed, any attempt to explain these results should be done with utmost caution. First of all, the lack of effects found may be because of the lack of an internally valid research design: we had non-random allocation between the two A2G conditions on the one hand and the ‘control’ Fitbit condition in the other. Also, the number of participants was smaller than we aimed for based on our power-analysis, and there was a large variation in when the post-intervention measurement took place. Because the effect sizes were generally small, it is unlikely that the lack of sufficient power explains the lack of statistically significant differences between the conditions, although the differences between baseline and post-intervention assessment in minutes of MVPA were 12–15% in the A2G conditions, which may be an indication that these A2G interventions do warrant further research. The lack of effect might also be due to the lack of exposure to the interventions; a large majority of the participants did not make use of the app as we assumed was needed to have sufficient influence and impact on determinants and behavior. Such lack of true exposure to m-health and e-health interventions has been found before (Eysenbach, 2005; Schoeppe et al., 2016) and a main focus in further research should be how exposure to and actual use of such interventions can be intensified. Research by Schoeppe et al. (2016) suggests that effects of app-based interventions as part of more comprehensive, multi-component programs that may also include other forms of health education of face-to-face counseling may be more likely to be effective.
15.4.1 **Strengths and limitations**

The main research-design-related limitations of the present study have already been described in the Background to this paper, as well as in the opening paragraphs of this Discussion section: the lack of full randomization, the small sample size, the variation in timing of the post-intervention measurement, and the fact that the baseline measurement of PA for some participants took place during intervention exposure. The initial aim was to include 159–200 participants (minimally 50 participants per condition) and to randomly assign Android users to the three conditions. However, due to the low response rate, only 28 and 27 Android users were assigned to the A2G-Full and A2G-Light condition respectively and none to the Fitbit condition. Consequently, Android users were randomized over the two A2G conditions, and iPhone users were assigned to the Fitbit condition. Additionally, the majority of the participants were highly educated, female and already more physically active than the population at large, which limits the external validity. Furthermore, about half of the participants in the A2G conditions experienced technical problems with their app, however only a few participants informed the researchers that they were having technical problems. Consequently, they might have stopped using the app, without first requesting assistance with solving the problem.

Strengths of this study are the high completion rate for participants (92%) and the fact that the experimental interventions were compared with an existing app (the Fitbit app). Comparing the A2G-Full app with the A2G-Light version further provided information whether sending tailored coaching messages on top of the monitoring and social comparison had an added effect on PA. Another strength was the use of the ActiGraph accelerometer – a valid and reliable accelerometer – to objectively assess baseline and post-intervention follow-up physical activity and the use of existing questionnaires to assess the behavioral determinants. Further evaluation is needed to examine whether sending coaching messages resulted in changes in step activity throughout the study period.

15.5 **Conclusion**

The current study showed no statistically significant effect of the A2G-Full condition as compared the A2G-Light and Fitbit condition. Future work is needed to increase actual use of the apps, to integrate the app in a more comprehensive, multi-component intervention, and in a study with better internal validity.

**Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A2G</td>
<td>Active2Gether</td>
</tr>
<tr>
<td>App</td>
<td>Smartphone application</td>
</tr>
<tr>
<td>BMI</td>
<td>Body mass index</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence interval</td>
</tr>
<tr>
<td>MVPA</td>
<td>Moderate-vigorous physical activity</td>
</tr>
<tr>
<td>PA</td>
<td>Physical activity</td>
</tr>
</tbody>
</table>
References


Chapter 15. Exploring use and effects of Active2Gether


Schoeppe, Stephanie, Stephanie Alley, Wendy Van Lippevelde, Nicola A. Bray, Susan L. Williams, Mitch J. Duncan, and Corneel Vandelanotte (2016). “Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic


Appendix 1

Table 15.6: Results of the regression analyses (regression coefficients (B) with 95% confidence intervals (95%CI)) to evaluate the differences in physical activity at post-intervention follow-up between A2G-Full and A2G-Light.

<table>
<thead>
<tr>
<th></th>
<th>Average minutes of moderate-vigorous physical activity per day</th>
<th>Average number of steps per day</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B (95% CI)</td>
<td>B (95% CI)</td>
</tr>
<tr>
<td>Model 0</td>
<td>Model 1: BMI</td>
<td>Model 2: Student</td>
</tr>
<tr>
<td></td>
<td>B (95% CI)</td>
<td>B (95% CI)</td>
</tr>
<tr>
<td>A2G-Light</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>A2G-Full</td>
<td>-1.90 [-10.84,7.05]</td>
<td>-2.34 [-11.51,6.83]</td>
</tr>
</tbody>
</table>

Note. Linear regression analyses are presented with regression coefficient (B) [95% confidence interval], and all analyses were adjusted for levels of physical activity at baseline and time between baseline and post-intervention follow-up.

Model 0: \( y = B_0 + B_1 \times \text{Physical activity at post-intervention} + B_2 \times \text{Physical activity at baseline} + B_3 \times \text{Time until post-intervention follow-up (#days)} \)

Model 1: Model 0 + B_4 * BMI
Model 2: Model 0 + B_4 * Student (yes/no)
Model 3: Model 0 + B_4 * BMI + B_5 * Student (yes/no)
### Appendix 2

Table 15.7: Results of the linear regression analyses (regression coefficients (B) with 95% confidence intervals (95%CI)) for differences in step activity at post-intervention follow-up between conditions using the Fitbit data ($N = 64$).

#### Average number of steps per day assessed with the Fitbit One for all three conditions

<table>
<thead>
<tr>
<th>Condition</th>
<th>Model 0 B (95% CI)</th>
<th>Model 1: BMI B (95% CI)</th>
<th>Model 2: Student B (95% CI)</th>
<th>Model 3: BMI-Student B (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fitbit</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>A2G-Full</td>
<td>-827.90 [-2849.7,1193.9]</td>
<td>-804.41 [-2878.1,1269.3]</td>
<td>-820.43 [-2858.4,1217.6]</td>
<td>-787.52 [-2879.4,1304.3]</td>
</tr>
</tbody>
</table>

#### Average number of steps per day assessed with the Fitbit One for A2G-Full versus A2G-Light

<table>
<thead>
<tr>
<th>Condition</th>
<th>Model 0 B (95% CI)</th>
<th>Model 1: BMI B (95% CI)</th>
<th>Model 2: Student B (95% CI)</th>
<th>Model 3: BMI-Student B (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A2G-Light</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>A2G-Full</td>
<td>-516.59 [-2226.9,1193.7]</td>
<td>-519.13 [-2294.1,1255.8]</td>
<td>-531.64 [-2266.7,1203.4]</td>
<td>-533.51 [-2334.4,1267.4]</td>
</tr>
</tbody>
</table>

**Note.** Linear regression analyses are presented with regression coefficient (B) [95% confidence interval]. For the analyses, the Fitbit data was used for baseline (day 1–day 7) and 12 weeks follow-up (day 78–day 84).

Model 0: $y = B_0 + B_1 * \text{Physical activity at post-intervention} + B_2 * \text{Physical activity at baseline} + B_3 * \text{Time until post-intervention follow-up (#days)}$

Model 1: $y = B_0 + B_4 * \text{BMI}$

Model 2: $y = B_0 + B_4 * \text{Student (yes/no)}$

Model 3: $y = B_0 + B_4 * \text{BMI} + B_5 * \text{Student (yes/no)}$
### Appendix 3

Table 15.8: Linear and logistic regression analyses for differences in behavioral determinants at post-intervention follow-up between A2G-Full and A2G-Light.

<table>
<thead>
<tr>
<th>Outcome measurement</th>
<th>Condition</th>
<th>Model 0</th>
<th>Model 1: BMI</th>
<th>Model 2: Student</th>
<th>Model 3: BMI-PA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-efficacy $B$ [95% CI]</td>
<td>A2G-Light</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td></td>
<td>A2G-Full</td>
<td>1.55 [-1.79, 4.89]</td>
<td>2.03 [-1.35, 5.42]</td>
<td>1.08 [-2.38, 4.53]</td>
<td>2.15 [-1.25, 5.54]</td>
</tr>
<tr>
<td>Outcome expectations $B$ [95% CI]</td>
<td>A2G-Light</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td></td>
<td>A2G-Full</td>
<td>0.62 [-0.35, 1.59]</td>
<td>0.57 [-0.44, 1.58]</td>
<td>0.82 [-0.14, 1.79]</td>
<td>0.58 [-0.43, 1.59]</td>
</tr>
<tr>
<td>Social norm descriptive $B$ [95% CI]</td>
<td>A2G-Light</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td></td>
<td>A2G-Full</td>
<td>1.10 [0.16, 2.04]</td>
<td>1.01 [0.03, 1.98]</td>
<td>1.18 [0.22, 2.13]</td>
<td>1.02 [0.03, 2.01]</td>
</tr>
<tr>
<td>Social norm injunctive $B$ [95% CI]</td>
<td>A2G-Light</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td></td>
<td>A2G-Full</td>
<td>0.53 [-1.34, 2.39]</td>
<td>0.48 [-1.45, 2.41]</td>
<td>0.54 [-1.47, 2.54]</td>
<td>0.30 [-1.67, 2.26]</td>
</tr>
<tr>
<td>Intentions 1 month OR [95% CI]</td>
<td>A2G-Light</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td></td>
<td>A2G-Full</td>
<td>-0.71 [-2.00, 0.58]</td>
<td>-0.98 [-2.36, 0.40]</td>
<td>-0.82 [-2.15, 0.51]</td>
<td>-0.96 [-2.34, 0.43]</td>
</tr>
<tr>
<td>Intentions 6 months OR [95% CI]</td>
<td>A2G-Light</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td></td>
<td>A2G-Full</td>
<td>-0.12 [-1.61, 1.37]</td>
<td>-0.26 [-1.82, 1.31]</td>
<td>-0.33 [-1.93, 1.27]</td>
<td>-0.33 [-1.92, 1.26]</td>
</tr>
<tr>
<td>Barriers $B$ [95% CI]</td>
<td>A2G-Light</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td></td>
<td>A2G-Full</td>
<td>-0.02 [-0.71, 0.67]</td>
<td>-0.13 [-0.82, 0.57]</td>
<td>-0.08 [-0.80, 0.63]</td>
<td>-0.13 [-0.84, 0.57]</td>
</tr>
</tbody>
</table>
Chapter 15. Exploring use and effects of Active2Gether

<table>
<thead>
<tr>
<th></th>
<th>A2G-Light</th>
<th>Reference</th>
<th>Reference</th>
<th>Reference</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Self-reg. skills</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>B [95% CI]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A2G-Full</td>
<td>0.45</td>
<td>0.40</td>
<td>0.26</td>
<td>0.39</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[-1.53,2.43]</td>
<td>[-1.65,2.45]</td>
<td>[-1.80,2.32]</td>
<td>[-1.69,2.48]</td>
<td></td>
</tr>
<tr>
<td><strong>Satisfaction</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>OR [95%CI]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A2G-Full</td>
<td>0.27</td>
<td>0.46</td>
<td>0.14</td>
<td>0.53</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[-1.11,1.65]</td>
<td>[-0.98,1.90]</td>
<td>[-1.24,1.52]</td>
<td>[-0.94,1.99]</td>
<td></td>
</tr>
<tr>
<td><strong>Long-term goals</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>B [95% CI]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A2G-Full</td>
<td>0.20</td>
<td>0.26</td>
<td>0.24</td>
<td>0.29</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[-0.59,0.98]</td>
<td>[-0.54,1.07]</td>
<td>[-0.58,1.05]</td>
<td>[-0.51,1.08]</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Linear regression analyses are presented with regression coefficient (B) [95% confidence interval (95%CI)] and logistic regression analyses with odds ratio (OR) [95% confidence interval (95%CI)], and all analyses were adjusted for levels of physical activity at baseline and time between baseline and post-intervention follow-up.

**Model 0:** \( y = B_0 + B_1 \times \text{Physical activity at post-intervention} + B_2 \times \text{Physical activity at baseline} + B_3 \times \text{Time until post-intervention follow-up (#days)} \)

**Model 1:** Model 0 + \( B_4 \times \text{BMI} \)

**Model 2:** Model 0 + \( B_4 \times \text{Student (yes/no)} \)

**Model 3:** Model 0 + \( B_4 \times \text{BMI} + B_5 \times \text{Meeting PA guidelines at baseline (yes/no)} \)
16. Evaluation of user appreciation and adherence of Active2Gether

Abstract
Physical inactivity is an increasingly serious global health problem, which implies a strong need for effective and engaging interventions. Smartphone technology offers new possibilities to address physical activity promotion. For app-based interventions to have an impact, both the effectiveness and user appreciation of the app are important. In this paper, we explore the user appreciation of the Active2Gether intervention, which offers personalized coaching to increase physical activity levels in daily life. The results are compared to the evaluation of a simplified version of the Active2Gether app (in which no coaching messages are sent) and the Fitbit app. Overall, the results reveal that users of a physical activity app appreciate a coaching feature to be included (on top of self-monitoring functionalities), but are also critical of how it is implemented (in terms of the number and content of the messages). The results also show that it is important to find a balance in the number of messages sent: too many messages seem to be perceived as annoying, but on the other hand, such system-initiated user interaction seems to reduce dropout.
This chapter appeared as:

16.1 **Introduction**

Despite the well-known health and well-being benefits of physical activity (I.-M. Lee et al., 2012; World Health Organization, 2010), about 50% of the adult population in western countries are less physically active than recommended by health authorities (World Health Organization, 2014). Moreover, engagement in moderate to vigorous physical activity decreases with age, in particular when transitioning from adolescence into (young) adulthood (Bell and C. Lee, 2005; Kwan et al., 2012). Thus, effective interventions are needed to encourage young adults to become or remain physically active. Nowadays, smartphone technology offers new possibilities to address physical activity promotion, as smartphones are accessible at all times, convenient, accurate, can be used to (self-) monitor the levels of physical activity and can provide highly tailored and real-time feedback. The high adoption rate of smartphones (97% among adults aged 20–29 years) and the popularity of health and fitness apps and activity trackers in the Netherlands (TelecomNieuwsNet, 2016) suggest that young adults will appreciate and adopt a physical activity intervention that makes use of smartphone technology.

For app-based interventions to have an impact, it is not only important that they are effective, but also that they are accepted by users. After all, if people are unwilling to use a certain app, it won’t be possible to prove its effectiveness. Research has shown that discontinuation of app use has several possible reasons, among which a lack of user friendliness and low engagement ((CHIC), 2011). A recent systematic review reported that 17 out of 23 app-based health intervention studies found significant intervention effects on lifestyle behavior outcomes and related health outcomes (Schoeppe et al., 2016). Additionally, some studies demonstrated perceived effectiveness of such apps. For example, King et al. reported that 69% of participants mentioned that the apps motivated them to be more physically active and 71% reported that the apps helped them to exercise regularly (King et al., 2013). Of the studies that found significant intervention effects, three studies examined associations between app usage and changes in the behavioral and health outcomes. All three showed that higher app usage was associated with improved physical activity and healthy eating (Schoeppe et al., 2016).

The Active2Gether (A2G) intervention is an example of such an app-based physical activity intervention. This app-based intervention is linked with a Fitbit One activity tracker and aims to encourage young adults to adopt and maintain a physically active lifestyle, by focusing on the domains of active transport, stair walking and leisure time sports activities. To do so, it classifies users into one of three awareness categories (in need of education about a healthy level of physical activity, open for coaching and in need of positive feedback to maintain behavior), it helps users to select the most promising coaching domain (active transport, stairs, sports), it helps users to set a goal and it sends motivational messages. Detailed information on the coaching system can be found in the subsequent section.

This study focuses on the **user evaluation** of the Active2Gether coaching system. We compare the evaluation of this app to two other related apps: a simplified variant of the same app (without coaching functionality), and a commercially available physical activity app (the Fitbit app). The objective of the study presented in this paper is threefold: (1) to evaluate how users used the app (adherence, interaction rates), (2) to assess how users evaluated the app with respect to perceived effectiveness, user friendliness etc., and (3) to evaluate the users’ appreciation of the coaching messages sent. By evaluating different aspects of the apps, we form an idea of what is and is not appreciated by users of physical activity apps,
which is vital information for developers of such systems.

16.2 **Active2Gether system**

The Active2Gether personalized coaching system evaluated in this paper aims to encourage young adults to adopt and maintain a physically active lifestyle, by focusing on the domains of *active transport, stair walking* and *leisure time sports activities*.

16.2.1 **Initial assessment**

New users start by filling out an online intake questionnaire, including questions about their daily life (e.g., occupation, significant locations) and about psychological factors underlying their physical activity behavior (e.g., self-efficacy, intentions, perceived barriers). Then, the users start a one-week assessment, to gauge their current physical activity level. The physical activity data is collected by means of a *Fitbit One* activity monitor, in combination with prompted daily user input about active transport and sports activities. The Fitbit One was chosen because of its relatively long battery life and possibility to synchronize the data continuously through Bluetooth LE.

16.2.2 **Awareness classification**

After the assessment week, the users are assigned to one of three awareness categories, namely *education*, *coaching* or *feedback*. This classification is based on whether they meet the Dutch physical activity guidelines (World Health Organization, 2010) and whether they think they should be more active. This classification is repeated every three weeks, in order to tailor the system to the user’s latest awareness state.

If users do not meet the guidelines, but think they are sufficiently active, they receive *educational* messages to inform them about healthy levels and health benefits of physical activity. If users meet the guidelines and do not see the need to be more active, they receive affirmative *feedback* messages to maintain their current level of physical activity. If users don’t meet the guidelines and understand that their physical activity level should increase, or if they meet the guidelines but still want to be more active, they enter the *coaching* phase.

16.2.3 **Domain selection and goal setting**

In the coaching phase, the system first suggests the user to select one of the three possible domains (i.e., active transport, stair walking or sports activities) to focus on for the next week. To do so, the user’s behavior in each of the three domains is compared to what could be expected for this particular user based on personal context information. The domain with the lowest evaluation, and thus the largest potential for improvement, is suggested to the user, although they are free to select another domain instead.

After the domain selection, the user is prompted to set a domain-specific goal. If the user met his previous goal for this domain, the system suggests to increase it, and otherwise the user is suggested to keep the same goal.

16.2.4 **Identification of promising coaching determinants**

Then, the system runs simulations of a computational model to estimate what types of coaching messages are expected to be most effective for the user. To do so, the user receives
a number of questions to assess the current state of personal determinants underlying physical activity behavior (e.g., self-efficacy, intentions). These states are translated into numerical values and inputted to the computational model, which describes the dynamics between those determinants and their effect on the behavior (Mollee and van der Wal, 2013). Using simulations, the system determines what the effect of improvement in each of the determinants on the behavior would be, and selects the three most promising determinants to be targeted in the coaching accordingly.

16.2.5 Coaching messages
Based on the selected domain and the identified most promising coaching determinants, the coaching messages are filtered to remove any messages that are irrelevant or not applicable to the user. At certain times (up to a maximum of three times per day), a message is selected from the remaining set of messages and sent to the user. Additionally, users may also receive messages to remind them to synchronize their data or to charge their Fitbit.

The coaching cycle (consisting of domain selection, goal setting, and identification of promising coaching determinants) is repeated weekly, in order to tailor the coaching to the user’s current state and needs at all times.

16.2.6 Active2Gether app
The Active2Gether app shows a picture of a virtual coach with a welcome message that depends on the user’s choice for a coaching domain, as well as the current daily number of steps and stairs and the user’s progress towards the general weekly goal of 70,000 steps. Below that, the app shows an ordered graph with the user’s total step count of the past seven days, among the data of up to six other users. These users are selected based on the user’s preferred (upward or downward) direction of social comparison. Where possible, the app shows the data of Facebook connections, and if not available, the data of anonymized other users is shown. The same data is also accessible to the users by logging in to the Active2Gether website.

As explained above, users may receive different types of messages from the system. These pop up on the smartphone with a push notification, and are presented as overlay on top of the dashboard. As long as the app is not opened to read the message, the user receives a notification every 15 minutes.

More detailed information about the design of the Active2Gether system can be found in (Klein et al., 2015).

16.3 Methods
This section describes the context in which the user evaluation was conducted, as well as the process of data collection and preprocessing. First, we describe the user study in which the data was collected. Then, we describe the conditions of this study in more detail. Finally, we describe the aim and content of the analyses.

16.3.1 User study
Participants were recruited at two university campuses in the Netherlands, as well as through referral of other participants. Interested participants were eligible if they were young adults (18 to 30 years old), healthy, and in possession of a smartphone running on Android or iOS.
Participants were assigned to one of three conditions, using a stratified randomization procedure based on their gender, type of smartphone and befriended participants. Each condition received (a variant of) a physical activity app: (1) Active2Gether Full, (2) Active2Gether Light, or (3) Fitbit. As the Active2Gether app was only available for Android smartphones, participants with an iPhone were automatically assigned to the Fitbit condition. The two other conditions were balanced on gender. Where possible, friends of participants were assigned to the same condition, in order to prevent them from comparing their apps during the study.

All participants were asked to fill out an online intake questionnaire, including questions about demographics, occupation, context, physical activity level and psychological constructs related to motivation to engage in physical activity. After the intake, the participants received a Fitbit One activity tracker, and were given instructions on how to install their assigned physical activity app and how to set up the synchronization. The participants used the app for a period of twelve weeks or longer, depending on their availability for the final appointment. After twelve weeks, the participants received a link to the final questionnaire, including questions about their experience with the app. At the final appointment with the researchers, they received €20 in gift vouchers as incentive for their participation.

16.3.2 Experimental conditions

As mentioned above, the participants were assigned to one of three conditions, each associated with (a variant of) a physical activity app. Figure 16.1 shows screenshots of the two different apps that were used.

Active2Gether-Full
Participants in the Active2Gether Full condition (A2G-Full) received the Active2Gether app, as described in Section 16.2.

In order to facilitate timely data synchronization, the participants were instructed to install the Fitbit app as well, as this app enables synchronization of the Fitbit One activity tracker with the smartphone through Bluetooth LE. However, they were urged to only use the Active2Gether app, and not to view or use the Fitbit app instead.

Active2Gether-Light
Participants in the Active2Gether Light condition (A2G-Light) also received the Active2Gether app. However, in contrast to the participants in the Active2Gether Full condition, they were not sent any coaching messages. Apart from that, their app provided the same functionalities and layout as the full Active2Gether app.

Fitbit
Participants in the Fitbit condition were coached with the Fitbit app. The dashboard of the Fitbit app shows the users their current daily step and stairs data, as well as distance travelled, number of active minutes and calories burned. By clicking on any of these data tiles, the users can view graphs of their data on different levels of aggregation, varying from 5-minute epochs to yearly statistics. In addition, the Fitbit app allows users to log food and water intake, to log sports activities, and to monitor their sleep. The participants were neither encouraged nor discouraged to use these additional functionalities. Similar to the Active2Gether app, Fitbit offers the users a website with a dashboard of their activity data as well.
Data collection

The user study yielded different types of data that are of interest when evaluating the users’ experience with the Active2Gether and Fitbit apps.

Intake questionnaire

First, the intake questionnaire provided information about the participants’ demographics and baseline physical activity level. The demographics included information about gender, age, and height and weight.

The baseline physical activity level was obtained using a short version of the IPAQ (The IPAQ Group, no date), and interpreted with the Combi Norm. The Combi Norm states that people should meet at least one of two other norms, namely the Fit Norm (Pollock et al., 1998) or the Dutch Norm for Healthy Exercise (Nederlandse Norm voor Gezond Bewegen, NNGB) (Kemper et al., 2000). In short, the Fit Norm requires to engage in vigorous-intensity physical activity for at least 20 minutes for at least three times per week. The NNGB states that adults should carry out at least 30 minutes of moderate to vigorous physical activity on a minimum of five days per week.

Final questionnaire

Second, a final questionnaire was used for information about the participants’ subjective experiences. One question asked about prior experience with physical activity apps and activity trackers. An answer “yes” indicated that users are already using an app or tracker, “some” meant that they have tried an app or tracker before but were not currently using
any, and “no” means that they had no prior experience. In addition, the final questionnaire contained Likert items about different aspects of the users’ appreciation of the apps. For example, the participants were asked to evaluate the number of questions and messages sent by the app. The question items about user appreciation were based on (Lund, 2001) and (Sauro, 2015), and included statements like “the app is easy to use” and “I would recommend this app to my friend”. In addition, the questionnaire allowed the users to name their three most and least favorite features or aspects of the app they used during the study. Also, the questionnaire included questions for the users in the Active2Gether conditions about the number, content and tone of the messages and/or questions sent. Finally, the participants were also asked whether they experienced problems with the battery life of their phone (due to their assigned coaching app) or technical problems of any other kind.

**Fitbit activity tracker data**
Third, the Fitbit collects different types of physical activity data, such as steps, floors climbed, distance traveled and calories burned. The presence of step data was used as an indicator of dropouts: if no Fitbit data is synchronized, it indicates that the participant is no longer using the app.

**Active2Gether app data**
Finally, the Active2Gether app provided some information about the frequency of interaction with the users. The questions and messages sent to the user were logged, as well as whether they successfully reached the user’s phone. First, this shows how much interaction the user had with the app. Also, logs of whether a message or question was successfully sent and received could indicate if the users experienced some technical problems or if they possibly removed the Active2Gether app from their smartphone.

### Data analysis
In order to evaluate the Active2Gether app, we explored different aspects of the use of the intervention by the end user.

#### App use and dropouts
First, we investigated the dropout of participants based on their Fitbit data. This was done through a Kaplan-Meier survival analysis. The difference between the three groups in the survival curves was tested with a log-rank test. Also, the number of days that participants were using their app (based on their Fitbit data) was determined. As this data was not normally distributed, differences between the conditions were tested by comparing mean ranks with a Kruskal-Wallis test and Mann Whitney U post-hoc tests.

#### Interaction frequency through questions and messages
For users in the two Active2Gether conditions, we also investigated how much (system-initiated) interaction they had with the app in terms of received questions (A2G-Full, A2G-Light) and messages (A2G-Full).

#### User experience
The final questionnaire contained 20 Likert items about user appreciation of the apps. A factor analysis revealed that the data could be summarized in four factors. All four factors showed good to excellent internal consistency: $\alpha = [.943, .901, .908, .814]$. Discussion
between JM and StV resulted in the following labels of the four factors: (1) satisfaction, (2) user friendliness, (3) perceived effectiveness, and (4) professionalism.

Examples of statements covered by each of the four factors are the following: (1) satisfaction: “the app meets my expectations”, (2) user friendliness: “I can easily find the information I’m looking for”, (3) perceived effectiveness: “the app motivates me to achieve my goals”, and (4) professionality: “the app looks professional”.

Differences between the user appreciation scores in the three conditions were assessed by means of a one-way Anova and Tukey post-hoc tests.

The questions about the experience of technical problems were answered on a 7-point Likert scale. It was considered as an occurrence of problems if the participant had selected one of the three answer options that reflected some extent of experiencing technical issues.

**Evaluation of questions and messages (A2G)**

The participants’ evaluation of the number of questions and messages sent by the app was given on a 5-point Likert scale in the final questionnaire. The answer options ‘too many’ and ‘far too many’ were aggregated into one category, as well as the options ‘too few’ and ‘far too few’. Then, the percentages of the answer selected were calculated for the two Active2Gether conditions as descriptive statistics.

For participants in the Active2Gether Full condition, the final questionnaire contained eight Likert items about the coaching messages. A factor analysis revealed that the data could be summarized in two factors. One negatively worded item was reversed. Both factors showed acceptable to good internal consistency: $\alpha = [.822, .760]$. The factors were labeled by JM and AM as capturing (1) the tone of voice and (2) the content of the messages.

Examples of statements covered by the two factors are: (1) tone of voice: “the messages seem credible and trustworthy”, and (2) content: “the messages are relevant to my personal situation”.

**Positive and negative aspects**

In the question about the most positive and most negative aspects of the app, the participants could list up to three positive and three negative features in free text. To analyze the participants’ feedback, two lists of categories were created while reading the responses (i.e., one for positive and one for negative aspects). The categories of positive aspects were (1) self-monitoring or insight, (2) social comparison, (3) coaching (messages), (4) goal setting, (5) clear, neat layout, (6) reminder, (7) perceived effect, (8) variety of data, and (9) other. The categories of negative aspects were (1) push notifications, (2) synchronization problems, (3) technical/battery problems, (4) inaccuracy of measurements, (5) lack of coaching, (6) excess or repetition of messages/questions, (7) irrelevance of coaching suggestions/messages, (8) missing functionalities, (9) unsatisfactory layout or user friendliness, (10) perceived demotivational effect, (11) use of activity tracker, and (12) other.

Then, all response items were classified using these lists and counted per category. This implies that one participant could mention more than one aspect in the same category, which would also be counted twice. Then, to compensate for the different numbers of participants in the three conditions, the counts were divided by the number of people in the corresponding condition to obtain a percentage.

All analyses were performed using SPSS 23.0 and Microsoft Excel 2010.
16.4 Results

This section gives an overview of the participants who partook in the user study, and shows the results of the app evaluation, as outlined in Section 16.3.

16.4.1 Participant characteristics

Originally, 104 people signed up for participation. Eleven participants dropped out before the end of the study, for example because of technical problems (e.g., smartphone lacked storage or battery capacity to run the apps), because they strongly disliked wearing the activity tracker, or because participation in the study collided with other obligations.

Table 16.1 shows the remaining number of participants in each of the conditions, as well as the median age, the age range, the number and percentage of female participants, and the number of participants meeting the norm for physical activity.

<table>
<thead>
<tr>
<th>All</th>
<th>A2G-Full</th>
<th>A2G-Light</th>
<th>Fitbit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of participants</td>
<td>92</td>
<td>24</td>
<td>23</td>
</tr>
<tr>
<td>Age – median</td>
<td>23</td>
<td>23</td>
<td>23</td>
</tr>
<tr>
<td>Age – range</td>
<td>[18-31]</td>
<td>[19-30]</td>
<td>[18-30]</td>
</tr>
<tr>
<td>Number of females</td>
<td>72 (78%)</td>
<td>17 (71%)</td>
<td>19 (83%)</td>
</tr>
<tr>
<td>Number meeting norm</td>
<td>50 (54%)</td>
<td>13 (54%)</td>
<td>9 (39%)</td>
</tr>
</tbody>
</table>

Table 16.2 shows how much prior experience the participants in each condition have with physical activity apps or trackers.

16.4.2 App use and dropouts

Table 16.3 shows the mean, median and range of the number of days that participants were using the app, how many participants dropped out per condition, and the percentage of participants that was still uploading Fitbit data after twelve weeks. Participants were marked as dropouts if they consecutively did not upload any Fitbit data for at least one day before the end of the experiment.

A Kruskal-Wallis test showed that the mean rank of the number of days that participants used the app (i.e., uploaded their Fitbit data) differed significantly between the three conditions, $\chi^2(2) = 10.671, p = .005$. Mann-Whitney U post-hoc tests revealed that differences existed between the conditions Active2Gether Full and Fitbit ($U = 372.5, p = .022$) as well as between the Active2Gether Light and Fitbit groups ($U = 292, p = .001$), but not between the two Active2Gether conditions ($U = 219, p = .102$).

Figure 16.2 shows the Kaplan-Meier survival curves for the three conditions based on the availability of Fitbit data. The log-rank test revealed that the survival functions show
Table 16.2: Participants’ prior experience with physical activity apps and trackers.

<table>
<thead>
<tr>
<th>Experience</th>
<th>All</th>
<th>A2G-Full</th>
<th>A2G-Light</th>
<th>Fitbit</th>
</tr>
</thead>
<tbody>
<tr>
<td>apps – yes</td>
<td>15</td>
<td>6</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>(16%)</td>
<td>(25%)</td>
<td>(9%)</td>
<td>(16%)</td>
</tr>
<tr>
<td>apps – some</td>
<td>18</td>
<td>2</td>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>(20%)</td>
<td>(8%)</td>
<td>(22%)</td>
<td>(24%)</td>
</tr>
<tr>
<td>apps – no</td>
<td>58</td>
<td>16</td>
<td>15</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>(63%)</td>
<td>(67%)</td>
<td>(65%)</td>
<td>(60%)</td>
</tr>
<tr>
<td>trackers – yes</td>
<td>9</td>
<td>4</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>(10%)</td>
<td>(17%)</td>
<td>(4%)</td>
<td>(9%)</td>
</tr>
<tr>
<td>trackers – some</td>
<td>8</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>(9%)</td>
<td>(8%)</td>
<td>(13%)</td>
<td>(7%)</td>
</tr>
<tr>
<td>trackers – no</td>
<td>74</td>
<td>18</td>
<td>18</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>(80%)</td>
<td>(75%)</td>
<td>(78%)</td>
<td>(84%)</td>
</tr>
</tbody>
</table>

Table 16.3: Dropouts and participants that were still using the app after 12 weeks.

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>A2G-Full</th>
<th>A2G-Light</th>
<th>Fitbit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of participants</td>
<td>92</td>
<td>24</td>
<td>23</td>
<td>45</td>
</tr>
<tr>
<td>Days using the app – mean</td>
<td>70.5</td>
<td>79.0</td>
<td>81.0</td>
<td>60.6</td>
</tr>
<tr>
<td>Days using the app – median</td>
<td>84</td>
<td>84</td>
<td>84</td>
<td>84</td>
</tr>
<tr>
<td>Days using the app – range</td>
<td>[0-84]</td>
<td>[12-84]</td>
<td>[21-84]</td>
<td>[0-84]</td>
</tr>
<tr>
<td>Number of dropouts</td>
<td>35</td>
<td>8</td>
<td>3</td>
<td>24</td>
</tr>
<tr>
<td>Percentage using the app at 12 weeks</td>
<td>62.0%</td>
<td>66.7%</td>
<td>87.0%</td>
<td>46.7%</td>
</tr>
</tbody>
</table>

statistically significant differences between the three conditions, $\chi^2(2) = 12.381, p = .002$. 

16.4.3 Interaction frequency through questions and messages

In total, 8,556 questions were successfully received by the 47 participants in the Active2Gether Full and Active2Gether Light conditions, over the course of twelve weeks. All questions that were derived were also sent, and 112 questions were sent but not received. The 24 participants in the A2G-Full condition received 1,324 messages successfully. In contrast, 48 were derived but not sent and 57 were sent but not received.

Further analysis shows that Active2Gether (Full or Light) participants received an average 182 questions during the twelve-week period. All derived questions were successfully sent, but for 22 out of 47 users, a question was not received by the phone at some point. For one user, no questions were derived and therefore this user did not receive any questions. This suggests some technical problems or unsuccessful installation of the app.

Similarly, the logs show that participants in the Active2Gether Full condition received an average of 55 messages in twelve weeks. For five out of 24 users, a derived message was
not sent at some point, which could indicate that the app was removed before the end of the study. For nine users, a sent message was not received by the phone, and one user did not receive any messages at all.

16.4.4 User experience

Of the 92 participants that completed the study, 90 filled out the complete final questionnaire. Two participants started, but did not finish the questionnaire.

As described in the Section 16.3, a factor analysis was performed that revealed four factors. Table 16.4 and Figure 16.3 show the average scores on those factors for the three conditions. A one-way Anova showed that the user experience ratings for all four factors differed between the three conditions; satisfaction: $F(2, 88) = 20.455, p < .001$; user friendliness: $F(2, 88) = 4.755, p = .011$; perceived effectiveness: $F(2, 88) = 5.541, p = .005$; professionality: $F(2, 88) = 15.224, p < .001$. Tukey post-hoc tests revealed that differences existed between the conditions Active2Gether Full and Fitbit ($p < .001$; $p = 0.032$; $p = .014$; $p = .009$), and between the Active2Gether Light and Fitbit groups ($p < .001$; $p = .039$; $p = .033$; $p < .001$), but not between the two Active2Gether conditions ($p = .751$; $p > .999$; $p = .977$; $p = .091$).

The differences between the three conditions are also apparent in the overall user experience rating, $F(2, 88) = 14.809, p < .001$. A Tukey post-hoc test showed that the difference between the Active2Gether Full and Fitbit groups is significant ($p < .001$), as well as between the Active2Gether Light and Fitbit condition ($p < .001$), but not between the two Active2Gether conditions ($p = .748$).

Figure 16.4 shows the percentage of participants that expressed battery problems or other technical issues.
Table 16.4: Average scores on user appreciation (range [1,7]).

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>A2G-Full</th>
<th>A2G-Light</th>
<th>Fitbit</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Satisfaction (mean, sd)</td>
<td>3.73</td>
<td>3.01</td>
<td>2.73</td>
<td>4.61</td>
</tr>
<tr>
<td></td>
<td>(1.56)</td>
<td>(1.43)</td>
<td>(1.38)</td>
<td>(1.24)</td>
</tr>
<tr>
<td>(2) User friendliness (mean, sd)</td>
<td>5.11</td>
<td>4.71</td>
<td>4.71</td>
<td>5.51</td>
</tr>
<tr>
<td></td>
<td>(1.29)</td>
<td>(1.38)</td>
<td>(1.69)</td>
<td>(1.07)</td>
</tr>
<tr>
<td>(3) Perc. effectiveness (mean, sd)</td>
<td>4.23</td>
<td>3.66</td>
<td>3.75</td>
<td>4.77</td>
</tr>
<tr>
<td></td>
<td>(1.60)</td>
<td>(1.79)</td>
<td>(1.64)</td>
<td>(1.40)</td>
</tr>
<tr>
<td>(4) Professionality (mean, sd)</td>
<td>4.45</td>
<td>4.18</td>
<td>3.44</td>
<td>5.09</td>
</tr>
<tr>
<td></td>
<td>(1.36)</td>
<td>(1.34)</td>
<td>(1.44)</td>
<td>(1.04)</td>
</tr>
<tr>
<td>Overall (mean, sd)</td>
<td>4.38</td>
<td>3.89</td>
<td>3.66</td>
<td>4.99</td>
</tr>
<tr>
<td></td>
<td>(1.23)</td>
<td>(1.43)</td>
<td>(1.24)</td>
<td>(1.00)</td>
</tr>
</tbody>
</table>

Figure 16.3: Average scores on four factors of user appreciation: (1) satisfaction, (2) user friendliness, (3) perceived effectiveness, and (4) professionality.

16.4.5 Evaluation of questions and messages (A2G)

Figure 16.5 shows the percentage of participants that perceived the number of questions received as ‘too many’, ‘just right’ or ‘too few’.

In addition to the questions, the users received messages through the app. The Active2Gether Light participants did not receive motivational coaching messages, but were only sent messages about their Fitbit’s low battery life or overdue data synchronization. Figure 16.6 shows the percentage of participants that selected certain answers.

Table 16.5 shows the average scores on the two factors on which the messages in the Active2Gether Full condition were evaluated.
Figure 16.4: Percentage of participants with battery problems and/or other technical problems.

Figure 16.5: User evaluation of number of questions in Active2Gether conditions.

Table 16.5: Average scores on eight statements about evaluation of messages (range [1,5]).

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>St. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tone of voice</td>
<td>3.04</td>
<td>0.848</td>
</tr>
<tr>
<td>Content</td>
<td>2.32</td>
<td>0.800</td>
</tr>
<tr>
<td>Overall</td>
<td>2.68</td>
<td>0.726</td>
</tr>
</tbody>
</table>

16.4.6 Positive and negative aspects

As explained in Section 16.3, the reported positive and negative aspects of the apps were classified into overarching categories.

Table 16.6 shows the categories of positive aspects that were mentioned most often in the Active2Gether Full condition, and Table 16.7 lists the categories of negative aspects.
Table 16.6: Most often reported positive aspects in the Active2Gether Full condition, with count and percentage.

<table>
<thead>
<tr>
<th>#</th>
<th>Aspect</th>
<th>Count</th>
<th>Perc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Self-monitoring or insight</td>
<td>13</td>
<td>54%</td>
</tr>
<tr>
<td>2</td>
<td>Coaching (messages)</td>
<td>10</td>
<td>42%</td>
</tr>
<tr>
<td>3</td>
<td>Social comparison</td>
<td>9</td>
<td>38%</td>
</tr>
</tbody>
</table>

Table 16.7: Most often reported negative aspects in the Active2Gether Full condition, with count and percentage.

<table>
<thead>
<tr>
<th>#</th>
<th>Aspect</th>
<th>Count</th>
<th>Perc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Technical/battery problems</td>
<td>15</td>
<td>63%</td>
</tr>
<tr>
<td>2</td>
<td>Excess or repetition of messages/questions</td>
<td>12</td>
<td>50%</td>
</tr>
<tr>
<td>3</td>
<td>Irrelevance of coaching suggestions/messages</td>
<td>9</td>
<td>38%</td>
</tr>
</tbody>
</table>

Table 16.8 lists the categories of positive aspects that were mentioned most often in the Active2Gether Light condition, and Table 16.9 enumerates the categories of negative aspects.

Table 16.10 shows the categories of positive aspects that were mentioned most often in the Fitbit condition, and Table 16.11 lists the most common categories of negative aspects.
Table 16.8: Most often reported positive aspects in the Active2Gether Light condition, with count and percentage.

<table>
<thead>
<tr>
<th>#</th>
<th>Aspect</th>
<th>Count</th>
<th>Perc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Self-monitoring or insight</td>
<td>17</td>
<td>77%</td>
</tr>
<tr>
<td>2</td>
<td>Layout</td>
<td>14</td>
<td>61%</td>
</tr>
<tr>
<td>3</td>
<td>Social comparison</td>
<td>7</td>
<td>32%</td>
</tr>
</tbody>
</table>

Table 16.9: Most often reported negative aspects in the Active2Gether Light condition, with count and percentage.

<table>
<thead>
<tr>
<th>#</th>
<th>Aspect</th>
<th>Count</th>
<th>Perc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Missing functionalities</td>
<td>14</td>
<td>61%</td>
</tr>
<tr>
<td>2</td>
<td>Synchronization problems</td>
<td>11</td>
<td>48%</td>
</tr>
<tr>
<td>3</td>
<td>Technical/battery problems</td>
<td>9</td>
<td>39%</td>
</tr>
</tbody>
</table>

Table 16.10: Most often reported positive aspects in the Fitbit condition, with count and percentage.

<table>
<thead>
<tr>
<th>#</th>
<th>Aspect</th>
<th>Count</th>
<th>Perc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Self-monitoring or insight</td>
<td>41</td>
<td>91%</td>
</tr>
<tr>
<td>2</td>
<td>Layout</td>
<td>23</td>
<td>51%</td>
</tr>
<tr>
<td>3</td>
<td>Variety of data</td>
<td>19</td>
<td>42%</td>
</tr>
</tbody>
</table>

Table 16.11: Most often reported negative aspects in the Fitbit condition, with count and percentage.

<table>
<thead>
<tr>
<th>#</th>
<th>Aspect</th>
<th>Count</th>
<th>Perc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Inaccuracy of measurements</td>
<td>26</td>
<td>58%</td>
</tr>
<tr>
<td>2</td>
<td>Synchronization problems</td>
<td>11</td>
<td>24%</td>
</tr>
<tr>
<td>3</td>
<td>Missing functionalities</td>
<td>8</td>
<td>18%</td>
</tr>
</tbody>
</table>

16.5 Discussion and conclusions

This section provides an interpretation of the results. The most important findings are discussed, as well as their implications for the design of physical activity apps or interventions.

Participants of the study were young adults (18 to 31 years old), and a majority was female (78%). Approximately half of the participants were sufficiently physically active.
according to health recommendations, which is in line with overall findings about the adult population in western countries (World Health Organization, 2014). The majority of the participants had no prior experience with physical activity apps (63%) or activity trackers (80%).

The Fitbit data showed that the dropout of the intervention was lower in the Active2Gether conditions (both variants) than in the Fitbit condition. The percentage of users after 12 weeks was highest in the Active2Gether Light condition (87.0%) and lowest in the Fitbit condition (46.7%). This pattern was already visible in earlier weeks of the intervention. Interestingly, this cannot be explained by the experience of technical or battery problems (see Figure 16.4), as those factors were higher in the two Active2Gether conditions than in the Fitbit condition.

Research has shown that dropout numbers in health interventions are very diverse, e.g. from 6% in an 8-week physical activity intervention (King et al., 2016) to 73% in a 14-week healthy lifestyle intervention (Naimark et al., 2015). This makes it difficult to compare the results, but it is interesting to see that the dropout between the Active2Gether conditions and the Fitbit condition differed so strongly, even though the setup of the study was otherwise exactly the same. This could suggest that the Active2Gether conditions offer something that retains the users’ interest for a longer period of time. Other research has shown that adherence in health apps is generally quite low: 26% of health apps is only used once after downloading, and 74% of health app users indicated to have stopped using the app within ten times of using it ((CHIC), 2011). In light of these findings, the adherence in the current study was very acceptable.

The systems logs showed that the participants in the two Active2Gether conditions received approximately 182 questions during the twelve-week period. Almost 99% of the questions were received successfully. Approximately half of the users perceived the number of questions as too high, which could be resolved by replacing some user input by automated registration (e.g., of sports activities and transport options).

Similarly, over 92% of the derived coaching messages were received successfully by the users in the Active2Gether Full condition. Over the twelve-week period, these participants received an average of 55 coaching messages. This is less than the system allows (i.e., up to three messages per day), which indicates that there were not always relevant messages available for the user, and the set of messages should be extended to cover more combinations of context variables. However, the participants in the two Active2Gether conditions also indicated that the number of messages was too high (57% and 38%). Since 38% of the users in the Active2Gether Light condition also perceived the number of messages as too high, even though they only received messages about the status of their Fitbit battery and data synchronization, it is possible that the coaching messages were not the main contributor to these sentiments. Also, it is possible that participants did not clearly distinguish between questions and messages, and perceived the overall number of app-initiated interactions as too high.

Over all four factors of user appreciation, the Fitbit app was rated higher than the two Active2Gether conditions. Generally, the full Active2Gether app scored slightly better than the simplified version, although these differences were not significant. Reasons for the relatively low scores could be explained by the feedback on the apps’ negative aspects. Both Active2Gether conditions reported quite some technical problems (63% and 39%, respectively), for example with respect to their smartphone’s reduced battery life. For the
Chapter 16. Evaluation of user appreciation and adherence of Active2Gether

full Active2Gether app, the repetition in the questions and messages was disliked (50%) and the messages were perceived as not very personal or relevant (38%). The main criticism on the Active2Gether Light app was its simplicity (61%). The participants in the Fitbit condition complained most often about its inability to reflect certain activities (58%), as well as delays in or problems with synchronization (24%) and lacking functionalities (18%).

On the other hand, the feedback on the positive aspects shows that participants in all three conditions highly value the possibility to review their behavior (58%, 74% and 91%, respectively). In the Fitbit condition, the percentage is probably higher because of the option to view activity data in more detail (i.e., per 5 minutes) and in different types of parameters (i.e., active minutes, calories burned, etc.). These aspects are mentioned by 42% of users in the Fitbit condition. In addition, participants in both Active2Gether conditions appreciated the comparison to other users (38% and 30%), and the clean layout of the app (29% and 61%). Finally, users of the full Active2Gether app praised the coaching aspect (42%).

One of the key strengths of this study is that the user evaluation was based on considerable use of the app, as the participants were asked to use their app for at least twelve weeks. This allows for a substantiated evaluation. In addition, since a variety of different aspects of the apps were considered, the evaluation in this paper gives a rather complete picture of the likes and dislikes of the participants. While the focus of the evaluation is on one of the apps, the full Active2Gether app, the comparison to its simplified version and a commercially available app provides more insight in the aspects that are appreciated by users.

A limitation of the present study is that the results might not be easily transferable to the general population. It covered only young adults (18 to 31 years old), and the majority of the participants was female (78%). Also, all participants signed up voluntarily, so they probably were already intrinsically motivated to improve their physical activity levels through an app-based intervention. Moreover, it is possible that participants who use a physical activity app in the context of an experiment perceive their experience differently from users who download the app solely for their own use. In addition, although different aspects of the apps were evaluated, it is difficult to say which aspects or features contributed to specific scores on their satisfaction, user friendliness, perceived effectiveness and professionalism. Further (qualitative) research should reveal exactly which aspects were liked and disliked by the users. Also, since the Active2Gether app was only available for Android smartphones, the assignment of participants to conditions was not completely random, which in theory could have influenced the results. Finally, although the subjective user evaluation is very important for the user experience and adherence, it does not necessarily imply the apps’ effectiveness as well. In order to develop and offer successful physical activity interventions, both the user experience and effectiveness should be ensured.

Overall, we can conclude that users of a physical activity app want a coaching feature to be included (on top of self-monitoring functionalities), but are also critical of how it is implemented (in terms of number and content of the messages). It is important that the coaching is perceived as personal and relevant, and it should be sufficiently diverse in order not to become too repetitive. Thus, it is important to find a (personal) balance in the number of messages: too many messages seem to be annoying, but on the other hand, such system-initiated user interaction seems to reduce dropout. Further research should reveal how this perfect balance can be achieved.
Acknowledgments

The authors would like to thank prof. dr. Johannes Brug for his feedback on the manuscript.
References


REFERENCES


Part Six

17 Discussion & Conclusion ................. 339
17.1 Summary of research contributions
17.2 Discussion of research questions and results
17.3 Ethical aspects
17.4 Future work
17.5 Conclusion
17. Discussion & Conclusion

The research presented in this thesis explored new approaches to behavior change interventions for physical activity, by applying techniques from modern (mobile) technology and artificial intelligence as well as scientific knowledge from human-directed disciplines within psychology and social sciences. Because of the interdisciplinary nature of this subject, it was addressed by focusing on different aspects from various perspectives.

This chapter reflects on how the previous chapters contribute to the research objective outlined in Chapter 1. In Section 17.1, the overall contributions of the work presented in this thesis are summarized. Section 17.2 presents a discussion of the results, by revisiting the research questions from Section 1.3 of the Introduction. Section 17.3 reflects upon the ethical aspects of the Active2Gether intervention and Section 17.4 looks ahead at future work. Finally, Section 17.5 presents the conclusion of this thesis.

17.1 Summary of research contributions

The research presented in this thesis helps to advance the field of innovative healthy lifestyle promotion, by investigating and developing methods that make coaching more personal, more relevant and overall more ‘intelligent’. Because of the interdisciplinary nature of the research objective, the contributions of this thesis are diverse. This section outlines its overall research contributions, thereby explaining how this research contributes to the advancement of intelligent behavior change applications.

First of all, an inventory is made of the state-of-the-art of apps that promote physically active lifestyles. This is done from two perspectives. First, it is investigated to what extent behavior-change techniques are applied in the current offer of physical activity apps. This gives insight into how evidence-based these apps are. Second, the extent to which technological features are implemented is explored. This helps to understand how much of the technological potential is currently used in such coaching apps. Overall, the research into the state-of-the-art contributes to an understanding of the progress of physical activity coaching apps so far, as well as insight into missed opportunities. In combination with an exploration of user preferences for physical activity coaching apps, this provides a good starting point for further research and development of innovative healthy lifestyle promotion.

Second, some of the research in this thesis contributes to insights in human behavior and factors that influence behavior change processes. This is done, for example, through investigating whether people who are willing to join an online social community in a physical activity promotion program benefit more from it than people who don’t sign up for the community. The results of the analyses do indeed point in that direction. Another new
inspection is that it is important to consider people’s preferred direction of social comparison when implementing a social comparison feature, as presenting data of other users for comparison may lead to adverse effects if it contradicts with the user’s preference. Furthermore, computational models of behavior change processes also contribute to understanding human behavior in this area. This thesis presents a computational model of influences on physical activity behavior, which can be a helpful tool to further explore the dynamics through simulations and property analysis. The preliminary validation of this model adds to its reliability, and therefore to the value of its simulations. Finally, two validations of an existing computational model of social contagion help to support the credibility of this model, as well as to understand the role of the social contagion process in collective behavior change attempts.

The work on the design and development of innovative behavior change approaches is the third overall contribution of this thesis. Clearly, the documentation of the overall design and implementation of the Active2Gether system provides valuable information. It may serve as inspiration for future developments of behavior change systems, and the related lessons learnt point to the more or less promising directions. In addition, some behavior change approaches (that may or may not be included in the final version of the Active2Gether system) are explored in more detail. For example, a potential way to apply the computational model of social contagion to achieve behavior change within a social network is investigated. Similarly, secondary aspects of the computational model of influences on physical activity behavior, that come into play when applying it in a real-life behavior change system, are considered. Such explorations help to understand which endeavors might prove helpful in the design of behavior change interventions. Finally, the user evaluation of the Active2Gether system (in comparison with a commercially available app) helps to understand which aspects are appreciated by end users. This information can help to increase engagement, and thereby adherence and overall effectiveness. The fact that the evaluation is based on real use of the apps for a considerable amount of time, rather than based on hypothetical use scenarios, makes it an even more valuable contribution to research on behavior change applications.

The fourth overall contribution worth mentioning is the collection of several datasets in the context of this thesis. These datasets serve different purposes: they form the basis for answering specific research questions, they are used as input to the Active2Gether system that was developed in the context of this thesis, or they are used to evaluate the effectiveness and user appreciation of the Active2Gether system. Overall, the datasets provide very rich information regarding physical activity behavior and a diversity of related factors, such as the users’ significant locations, the availability of stairs at and transport options to these locations, friendship connections between users, and extensive psychological questionnaires about constructs like self-efficacy, intentions, perceived barriers, et cetera. Therefore, they did not only contribute to the research in this thesis, but they allow for a wide range of further analyses that were beyond the scope or time limit of this thesis. These additional investigations will likely lead to further insight in human behavior change dynamics.

17.2 Discussion of research questions and results

The research objective of this thesis is to investigate how mobile technology and artificial intelligence techniques can be applied in the design of a behavior change system that aims
to increase physical activity levels in young adults. In doing so, four subquestions were formulated, and each of the previous parts focused on answering one of these research questions. Table 17.1 shows an overview of the parts, research questions and chapters that together form this thesis. In the following subsections, the research questions are revisited one by one, the results related to these questions are discussed, and the implications and limitations of the work are explored.

Table 17.1: Overview of the parts, research questions and chapters in this thesis.

<table>
<thead>
<tr>
<th>Part I</th>
<th>Part II</th>
<th>Part III</th>
<th>Part IV</th>
<th>Part V</th>
<th>Part VI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>RQ 1</td>
<td>RQ 2</td>
<td>RQ 3</td>
<td>RQ 4</td>
<td>Discussion</td>
</tr>
<tr>
<td>Chapter 1</td>
<td>Chapter 2</td>
<td>Chapter 5</td>
<td>Chapter 8</td>
<td>Chapter 13</td>
<td>Chapter 17</td>
</tr>
<tr>
<td>Chapter 3</td>
<td>Chapter 6</td>
<td>Chapter 9</td>
<td>Chapter 10</td>
<td>Chapter 14</td>
<td></td>
</tr>
<tr>
<td>Chapter 4</td>
<td>Chapter 7</td>
<td>Chapter 10</td>
<td>Chapter 15</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chapter 11</td>
<td>Chapter 16</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chapter 12</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

17.2.1 Research question 1

What are requirements for mobile behavior change interventions for physical activity based on the state of the art of such interventions and user preferences of the target population?

The research in this thesis started by investigating the state of the art and user preferences for physical activity interventions. This way, a starting point was set for further development of such interventions, and promising directions or possible pitfalls were identified.

Reviewing the actual offer of physical activity apps, rather than the (only sparsely available) literature on this topic, led to valuable insights into the characteristics of currently available apps. The results presented in Chapter 2 show that the analyzed apps on average use five different behavior behavior change techniques, and none of the apps use more than eight or less than two, out of a possible 23 techniques. The most frequently used techniques were providing feedback, self-monitoring and goal setting, whereas some techniques were not used in any of the reviewed apps. Since research shows that the use of behavior change techniques is associated with effectiveness (Webb et al., 2010), these findings suggest that incorporating more of such techniques is a promising endeavor for the development of physical activity apps.

Similarly, the results in Chapter 3 reveal that the reviewed apps on average incorporate eight technological features out of a possible 37, and the numbers ranged between 0.5 and 19.5 (with 0.5 scores indicating disagreement between reviewers). The features that were identified most often were user input (to log activities or to form a personal profile), a textual/numerical overview of the user’s behavior and progress, sharing achievements or workouts in internal or external social networks, and general advice on physical activity. Many features were discovered only rarely, among which adaptation, integration with external sources, feedback based on the user’s physical/social context, encouragement
through gamification or some form of punishment, and the possibility to contact an expert through the app.

When exploring the requirements for mobile healthy lifestyle interventions, the possibilities (in terms of evidence-based behavior change techniques or technological capabilities) are not the only important factor. The needs, expectations and preferences of intended users are also important, as users’ acceptance of the behavior change app influences their app usage and engagement, and thereby indirectly affects its chances of effectiveness. The focus group interviews discussed in Chapter 4 investigated these aspects, and revealed that people in the target population see the value in a physical activity app. They indicated that they prefer apps that motivate or coach them, provide tailored feedback toward personally set goals, and that enable competition with friends through a ranking or earning rewards. These features were especially important to participants who did not meet the guidelines for healthy physical activity levels; sufficiently active participants expressed interest in detailed activity information, for example to see how they could intensify their training sessions or improve their athletic performances. Also, they were not willing to share accomplishments regarding physical activities through general social media, but rather in a private online community. These are important insights for the development of new physical activity interventions.

For both app reviews, an evident limitation is that they create a snapshot of the state of the art of physical activity apps at a certain time point. This means that the results that were valid at the time of the analysis could get outdated after a while, and regular replications of the study should ensure whether the results are still relevant. In addition, since the total offer of health and fitness apps has grown to unmanageable numbers, it is impossible to cover all existing apps. Instead, a search and inclusion strategy is needed to navigate through the available apps and focus on a subset. Depending on the search strategy and inclusion criteria, the set of apps may be biased. The app review described in Chapter 2 applied a relatively strict selection on the apps (e.g., requiring some form of individual tailoring to the user), which could explain why slightly higher numbers of behavior change techniques were identified in the app analysis compared to similar studies (Breton et al., 2011; Cowan et al., 2013). Another limitation of both reviews is that the results depend on the recognizability of the behavior change techniques (in Chapter 2) or the technological features (in Chapter 3), as well as on the reviewers’ interpretation. Therefore, it is important to carefully design and operationalize the scoring framework, thereby ensuring high inter-rater reliability and replicability of the results.

One of the strengths of the focus group interviews is that the results are based on opinions of people from the target population of the Active2Gether system (i.e., Dutch young adults). Therefore, their opinions can be incorporated to the design fairly directly, without having to take possible differences in user preferences between subpopulations into account. However, this also implies that the results cannot be generalized to beyond the interviewed subpopulation. In order to establish whether the findings hold for other groups as well, the study should be repeated with participants from more diverse backgrounds. In addition, enriching the qualitative findings with empirical data of app usage should reveal whether these user preferences actually transfer to app engagement and adherence.

Altogether, the three studies presented in Part II contribute to a deeper understanding of the requirements for a new generation of app-based physical activity interventions. The results show that it is important to make better use of available knowledge and technology: on the one hand, incorporating more behavior change techniques (that are associated with
effectiveness), and on the other hand, implementing more technology (that enables smarter and more tailored support). Newly developed interventions that comply with these findings arguably will progress toward the look and feel of a virtual personal coach. That is also exactly one of the preferences expressed by intended users, and therefore also likely increases the probability of users’ engagement in the intervention.

17.2.2 Research question 2

*What role can dynamic computational models play in the development of an intelligent mobile intervention for physical activity?*

Computational modeling is a technique from artificial intelligence and computational science that is often used to study, predict and better understand behavior of complex systems. It requires knowledge of the domain under investigation, and an iterative approach of design and analysis of the modeled relationships (see also Section 1.4.3). In this thesis, several steps were undertaken to investigate the potential role of a computational model in a mobile intervention for physical activity.

First of all, Chapter 5 proposes a computational model of the process that the behavior change intervention is aimed to influence, i.e. the interplay of cognitive and social concepts affecting physical activity behavior. For the creation of this model, it was necessary to gain a thorough understanding of the literature on this process (Bandura, 1998, 2004), in order to be able to translate the knowledge into a formal specification. The resulting model was used to run simulations of many different scenarios that cannot be manipulated easily in reality. For example, the simulations show that for an active person, the perception of higher impediments leads to a decrease in physical activity. This effect is more apparent for a high value of the parameter that describes the effect of the impediments on the behavior. On the contrary, the presence of impediments causes a boost in the behavior of an inactive person. This effect is greater for a low influence of the impediments on the behavior. This suggests that encountering (and overcoming) obstacles could in some cases help to build confidence and as a result lead to an increase in physical activity. This is a good example of how carefully conducted simulations can lead to insights in the modeled behavior.

In addition, the computational model was analyzed by means of specification and analysis of properties that express dynamic patterns that are expected to emerge (Bosse, Jonker, et al., 2009). This analysis was done to automatically check whether the model behaves correctly and to test certain hypotheses, by running a large number of simulations and verifying such properties against the simulation traces. This process would be very time consuming if done by hand, and the use of automated property checking also enables quick verification of complex properties that would require deep logical thinking. For example, automated property checking revealed that all simulations of inactive persons who experience facilitating circumstances (and no impediments) show an increase in physical activity over the course of the simulation trace.

Although the model analysis or verification and expert validation discussed in Chapter 5 are valuable first steps in the evaluation of the presented computational model, the value of the model was further strengthened by validating the output of the model based on empirical data. This was done in Chapter 6 by testing the accuracy of the model to predict changes in physical activity levels over a period of two to twelve weeks. The predictions of the model were compared to empirical physical activity data of 108 different participants, in this case
the daily numbers of steps measured with a Fitbit One activity monitor. The results of this validation study show that the computational model performs rather well in predicting the changes in physical activity levels. The predictions showed a weak to moderate positive correlation with the actual data, which was statistically significant ($p < .05$) for all predicted weeks. In contrast, a simple alternative model, based on randomly predicted end values for the physical activity behavior, had both weaker and non-significant correlations with the empirical data. Still, the random model performed better on some of the weeks than on others. This indicates that some characteristics of the data could have made it easier to predict the change correctly, and further investigation should find out why this happened. Nonetheless, the validation study presented in Chapter 6 contributes to the trustworthiness of the computational model that was part of the reasoning engine of the Active2Gether system, and therefore is a valuable step towards a more reliable and effective behavior change intervention.

However, when applying a computational model in a behavior change system, not only the goodness of fit matters. In the Active2Gether system, the model was incorporated in its reasoning engine, as a mechanism that predicts the most promising coaching strategy. (See Chapter 14 and maybe Section 17.2.4 for more details on the practical application of the model.) While the validity of the model’s predictions is undeniably essential, diversity in the content of the selected coaching messages is also important, in order to ensure the users’ engagement. Therefore, Chapter 7 describes an endeavor to find a set of parameters through a parameter tuning algorithm (simulated annealing), that yields values that stay close to the values based on literature but increases the diversity of the model outcomes as well. However, more importantly than finding this solution to the practical problem that arose when turning conceptual ideas into concrete implementations, the case study showcases a new application of parameter tuning in the field of computational modeling. The results from various simulation experiments with the model led to new insights in its behavior, and the global patterns of the resulting parameter sets provided information about (or could be explained by) the structure of the model and the meaning of its concepts and relations. This novel application of parameter tuning techniques is a new approach in the toolset of computational modelers, and therefore constitutes a valuable scientific contribution as well.

One of the limitations of the computational model presented in Chapter 5, and ultimately applied in the Active2Gether system as described in Chapter 14, is the fact that its design relied on a number of assumptions. Although the concepts, relations and parameter values of the model are based –where possible– on available literature on this topic (Bandura, 1998, 2004; Dzewaltowski et al., 1990; Petosa et al., 2003; Rovniak et al., 2002), formalizing the theory inevitably implies making decisions and simplifications to fill in the gaps (e.g., the numerical representation of the psychological concepts, the relationships as differential equations). This is not uncommon in computational modeling of complex behavioral processes, but it does mean that verification and validation of the model is important. The automated property verification and the expert validation in Chapter 5 provide initial evidence for the correctness of the model and face validity, and the study described in Chapter 6 takes the validation to the next level by comparing the model’s predictions to empirical data. Still, the presented model validation also has its limitations. The analyses reveal that the model performs rather well in predicting changes on a weekly level, but they do not disclose whether it also correctly represents the behavior on a more detailed (i.e., daily) level. Also, it would be relevant to find out whether the results are transferable to longer
periods than the currently tested twelve weeks. The model’s performance on predicting the underlying psychological constructs would be an interesting further exploration as well. In terms of validating this model for its application in a behavior change system, an evident improvement of the current study would be to investigate whether the model also correctly predicts changes in physical activity, while taking the effect of a certain coaching strategy into account. If simulation results including such effects are proven valid, the reliability of the decisions of the reasoning engine also increases.

The main limitations of the parameter tuning study presented in Chapter 7 are concerned with the fact that the results are based on only one case study, with one model, one evaluation measure (and related cost function) and one parameter tuning algorithm. Therefore, further investigation should reveal whether this novel use of parameter tuning techniques to study model behavior is also successful when extended to other applications.

Overall, in order to provide an answer to research question 2, Part III provides an in-depth exploration of a computational model for psychosocial influences on physical activity behavior, with a special focus on how to apply it in a real-life behavior change system. Initial steps towards validation of the model show promising results, thereby justifying its incorporation in the reasoning engine of the Active2Gether system. Applying a parameter tuning algorithm to increase the diversity of the simulation outcomes further consolidates the model’s suitability for application in the reasoning engine.

17.2.3 Research question 3

*How can an individual’s social network be used to influence his/her physical activity behavior?*

As explained in Chapter 1, social processes play an important role in achieving and maintaining a healthy lifestyle (Zimmerman and Connor, 1989). These processes rely on several motivational mechanisms, such as social norms, observational learning, social facilitation, social support and social comparison (Bandura, 1998; Buunk et al., 2013; Cheng et al., 2014; Festinger, 1954; McNeill et al., 2006). Some of these concepts were incorporated in the computational model that was discussed in Part III. The chapters in Part IV describe the role of several other social aspects in behavior change for physical activity.

First of all, Chapter 8 investigates the effect of changes in the structure of a social network on the diffusion of emotions through this network. Although the spread of emotions might not seem directly relevant for physical activity promotion, the mechanisms studied in this work also apply to attitudes and behaviors (Christakis and Fowler, 2013), and are therefore interesting foundations for network interventions that aim to stimulate behavior change. In that context, Chapter 8 proposes a method for finding effective network interventions to influence specific individuals. That method analyzes the structure of the network around the targeted individual to find strong transitive connections to people with a negative influence and weak transitive connections to people with a positive influence, and thereby determines which ties should be strengthened or weakened to achieve the most optimal effect on the targeted individual. The effect of these interventions was analyzed by simulating the diffusion of emotional values (e.g., about intentions and goals) through the social network, based on a model of social contagion (Bosse, Duell, et al., 2009). The simulation experiments demonstrate that it is possible to design a behavior change mechanism that influences specific
persons by affecting the social interactions between people and themselves. The simulations show that changing connections closer to the targeted individual yield a larger influence than changing connections further from the target, even if such further connections are the strongest or weakest links in a chain. A comparison of the effect of the proposed interventions with all possible interventions shows that they are among the most optimal possible interventions, i.e. in the 98th or 99th percentile. In addition, it was shown that targets with fewer connections are easier to influence than highly connected individuals. The proposed method provides support for the feasibility of behavior change systems that aim to support specific individuals by altering (online) interactions between people in their social network.

The proof of concept presented in Chapter 8 was based on simulations of a computational model of social contagion (Bosse, Duell, et al., 2009), applied to partially hypothetical data of a social network. After all, the structure of the network was adopted from the classical Zachary karate club network (Zachary, 1977), but the weights of the connections as well as the initial ‘emotion’ values were assigned randomly. Therefore, it is important to assess whether one of the most important building blocks of these experiments (i.e., the computational model) is valid, by applying it to real data and evaluating its accuracy. The work in Chapter 9 provides a first step in that direction. Here, physical activity data was collected for a network of 20 people over a period of 30 days. Based on the first seven days of data, the model predicted whether each participant’s physical activity level would increase or decrease, and the predictions were compared to the actual data. The results show that the contagion model predicted the direction of the change correctly in 80% of the cases. If only participants with significant trendlines in their physical activity data were considered, this performance went up to 87%. In addition, the mean squared error of the model’s predictions was close to the error of trendlines of the data (0.4272 and 0.3613, respectively), and much lower than the error based on extrapolations of the first seven days (6.8992). Considering that the model’s predictions are only based on seven days of data and the trendlines are derived from the complete dataset, we conclude that the model is able to predict changes in physical activity levels in a social network rather well.

Another relevant question regarding social processes in physical activity promotion is whether participation in social components of physical activity interventions influences the effectiveness of such interventions. Previous research had shown that partaking in the online community is associated with a higher physical activity level (Groenewegen et al., 2012), but the question remained whether it also makes a difference in the effect of a promotion program. In Chapter 10, this was investigated by analyzing a dataset of 50,000 people participating in a corporate health program. Approximately 5,000 of these participants chose to join a built-in online community to share their achievements. From that dataset, ten connected components were selected, together with a number of non-connected users (who did not join the community) with otherwise similar characteristics. The results demonstrate that the physical activity level of people that were willing to join a community showed an increase that was significantly greater compared to the other users. Since the data sets were balanced for possibly confounding factors like gender, time of the year and corporation, it is very likely that people’s willingness to become member of the community was the dominant factor that made a difference for their increase in physical activity. Although this study does not prove that an online community increases the effectiveness of a physical activity promotion program overall, it does show that it makes a difference for users who want to
participate in such an online social network. Further research should reveal whether actual participation in a community (rather than willingness to participate) has a similar (or smaller or greater) impact on the program’s effectiveness. Also, by assigning users randomly to each of these two conditions, the possibility that the differences are caused by other underlying personality characteristics should be excluded.

Although the analysis in Chapter 10 shows that the willingness to partake in an online community is associated with an increased effectiveness of a physical activity promotion program, it does not explain what the cause of this effect is. Therefore, the work in Chapter 11 builds upon the earlier findings, by investigating whether the computational model of social contagion (previously applied in Chapter 8 and studied in Chapter 9) is able to capture the dynamics in physical activity levels for users who were part of the online community of the corporate health program. In order to do so, we compared model predictions of such a contagion model (enriched with an expected linear increase as result of participating in the intervention) with predictions of a simple linear model to a dataset of 2,472 interconnected users. The results show that the enriched social contagion model performs better at describing the pattern seen in the empirical data than the linear model, both in terms of a significant difference in mean absolute error and correlation with the empirical data. These results indicate that some of the dynamics of the physical activity levels in the network could be explained by social contagion processes. This is vital information for designers of health interventions with a social component, as such models can then be used to maximize the benefits of social influence.

The preceding endeavors were all more or less concerned with social contagion processes. In addition to that, this thesis also explored social comparison as mechanism to influence behavior change for physical activity. Chapter 12 aimed to investigate the effect of showing users their preferred direction of social comparison (i.e., upward or downward) or showing the opposite direction. The results demonstrate that showing social comparison in the direction that users do not prefer is counter-effective. Therefore, this study shows that it is important to know users’ preferences: if not to enhance motivational effects, then at least to avoid adverse effects of showing social comparison that discourages them. Additional striking outcomes are that significant effects were found based on very small sample sizes, and that participants seemed able to indicate their preference reasonably well by simply answering one dichotomous question. Especially in light of the observation that many existing physical activity promotion programs implement social comparison in an upward direction, the results of this study argue that designers of such interventions should be aware of the risk of presenting a one-type-fits-all operationalization of social comparison to their users.

This work on answering research question 3 has some limitations. First, some analyses were done based on hypothetical data. This is the case with the simulation experiments of social network interventions in Chapter 8 (where the connection strengths and emotion values were assigned at random), and with the simulations of the contagion model in Chapter 11 (where the connection strengths were based on generic values). The parameters of the social contagion model in Chapter 9 were derived from real data, but were based on heuristics. Validation of these heuristics would further strengthen the reliability of the outcomes of that study. Second, in both studies in which the accuracy of the social contagion model was tested (described in Chapter 9 and Chapter 11), certain assumptions implied limitations on the findings. In Chapter 9, social contagion was the only factor considered as
influence on the physical activity behavior in the group, and Chapter 11 only considered connections in the online community (and not in real life) in the social contagion process. Both simplifications imply that the models described an incomplete picture of the social processes taking place. A third limitation has to do with the generalizability of the results. For example, the analyses in Chapter 10 were based on whether or not users joined an online community at some point in time, but not necessarily during the same period as the analyzed data. Therefore, it is important to be very careful when interpreting the results and when drawing conclusions about the effect of such social features in general. In addition, the results of Chapter 11 cannot easily be transferred to the general population, as the analyses are based on data collection in the context of a physical activity promotion program. Similarly, the findings in Chapter 12 were based on a small sample of male amateur soccer players, so the results should be interpreted with caution. Finally, it is important to realize that the research presented in this thesis (although it provides valuable insights on social aspects in behavior change interventions) only scratches the surface of the vast amount of social processes in behavior change. Further research should continue studying these and other mechanisms in more detail.

In conclusion, the work presented in Part IV provides more insight in the importance of social processes in physical activity interventions, thereby contributing to answering research question 3. It was shown that users who choose to join an online community (at some point in time) benefit more from a physical activity promotion program than users who do not participate in the community. Also, a computational model of social contagion was supported by the results of two different data analysis studies. In addition, this part contains an investigation of how to exploit such processes in behavior change systems, specifically by altering connections in a user’s social network to steer social contagion or by selecting other users for social comparison while taking the user’s preference into account. This provides important guidelines for developers of behavior change interventions.

17.2.4 Research question 4

To what extent can the answers to the questions above be used to design, implement, exploit and evaluate a personalized mobile intervention for physical activity promotion?

The previous research questions each highlight parts of the multi-faceted subject of innovations in behavior change for physical activity. This research question, addressed in the chapters in Part V, deals with bringing all components together for the design, implementation, exploitation and evaluation of a personalized mobile intervention for physical activity promotion.

First of all, Chapter 13 describes the step-by-step approach to designing the Active2Gether intervention. This consists of defining the intervention’s primary and secondary objectives, establishing the theoretical framework, developing the content of the intervention, pilot testing the intervention, employing the intervention, and finally evaluating its effectiveness. Planning and reporting this development process meticulously stimulated approaching it in a structured manner, thereby ensuring that the applied techniques were grounded on evidence and well-integrated into one overall reasoning engine. The resulting chapter serves as a guideline for the design of behavior change interventions, by providing a starting point. Also, the detailed documentation increases the transparency of the current intervention, which helps to interpret the results of studies on its effectiveness and user
17.2 Discussion of research questions and results

evaluation.

Whereas Chapter 13 summarizes the components and reasoning mechanisms of the Active2Gether system, it does not provide details on its implementation. Therefore, Chapter 14 describes the technical development of the Active2Gether intervention and its underlying system, together with a reflection on the technical design choices. A recurring theme in the lessons learnt is the need for more flexibility than the system currently offers. To name a few examples, more flexibility would produce an improvement for the data collection (i.e., relying more on unintrusive measurements and less on user input), for the implementation of the ranking feature (i.e., less strict application of the user’s preference for upward or downward social comparison) and for the suggestion of a coaching domain (i.e., considering all activities related to a certain physical activity domain, rather than only activities associated with the reported significant locations). Also, the chapter suggests some ideas to make the coaching messages more personally and contextually relevant to the users, which we believe would improve the system’s effectiveness and user experience as well. As with the intervention design described in Chapter 13, such comprehensive documentation may serve as a source of inspiration or point of reference for the development of other behavior change systems. Additionally, the lessons learnt help others to identify potential pitfalls and opportunities to improve upon our system.

The previous two chapters provide insights into the design and implementation of the Active2Gether intervention. The intervention was employed in a user study to assess its effectiveness and to investigate the user experience. In Chapter 15, a study of the effectiveness of the Active2Gether system in increasing the physical activity levels of the participants is presented. The results show no significant effect on the physical activity levels in the two intervention groups (Active2Gether-Full and Active2Gether-Light), and no significant change in the underlying behavioral determinants either. Still, a clinically relevant increase of 4.2 to 4.4 daily minutes of moderate to vigorous physical activity was observed, which does suggest that there was an upward trend in the physical activity behavior. Therefore, it would be interesting to repeat the study over a longer period of time and with larger sample sizes, in order to see whether a significant effect would be found in more optimal circumstances.

For app-based interventions to be successful in supporting users to achieve or maintain healthy behavior, not only the effectiveness of the app is important. The users’ appreciation of the app should not be overlooked, as it can influence engagement with and adherence to the app. In Chapter 16, we explore the user appreciation of the Active2Gether intervention. The results show that the participants wanted a coaching feature to be included (on top of self-monitoring functionalities), but were also critical of how it was implemented (in terms of number and content of the messages). The coaching should be perceived as personal and relevant, and it should be sufficiently diverse in order not to become too repetitive. Thus, a (personal) balance in the number of messages should be found: too many messages seemed to be annoying, but on the other hand, such system-initiated user interaction appeared to reduce dropout. Overall, the user evaluation of the Active2Gether system (in comparison with the commercially available Fitbit app) helps to understand which aspects were appreciated by end users. This information could help to increase engagement, and thereby adherence and overall effectiveness. The fact that this evaluation was based on real use of the apps for a considerable amount of time, rather than based on hypothetical use scenarios, makes it an even more valuable contribution to research on behavior change applications.
During the process of designing, implementing and evaluating the Active2Gether system, several opportunities for improvement came to light. Unfortunately, the user study was not able to prove that Active2Gether was effective in increasing the users’ number of active minutes. This does not necessarily mean that the underlying design principles are faulty, but maybe that they were not executed well enough to see significant effects. The reflections on the technical design choices in Chapter 14 provide several suggestions to improve the current system. In addition, the user evaluation study reported in Chapter 16 also indicates how the user experience could be improved. For example, the system should be more technically stable, the physical activity data in the dashboard should be more detailed and updated more often, and the coaching should be more personally and contextually relevant. It is promising that, despite the small sample sizes and the short exposure to the intervention, small but clinically relevant changes in daily number of minutes spent on moderate to vigorous physical activity were observed. Repeating the user study, after implementing the identified improvements in the Active2Gether system, with larger sample sizes and over a longer period of time, should show whether it is indeed a worthwhile endeavor to incorporate evidence-based techniques and methods from artificial intelligence to improve mobile behavior change interventions.

Altogether, the chapters in Part V provide a complete description of the process of developing and evaluating an innovative mobile behavior change intervention for physical activity. Where relevant and possible, the results from the research presented in the preceding parts were incorporated into the system’s design. Also, several improvements for the development of such systems were suggested. Therefore, this part contributes on the one hand by proposing a design for a behavior change system with innovative elements, and on the other hand by providing hands-on guidelines for future work in this domain.

### 17.3 Ethical aspects

When using technology to monitor and influence people’s behavior, it is important to take ethical considerations into account. After all, a considerable amount of personal data is collected and used to derive tailored feedback that is aimed to influence the user’s behavior. In order to respect the users’ privacy and autonomy, this should be done carefully and with high regard for ethical implications. Therefore, this section reflects on some ethical aspects that are relevant with respect to the design and employment of the Active2Gether system.

The reflection is based on the eFRIEND ethical framework for intelligent environments development (Jones et al., 2015). This framework is a combination several other frameworks addressing ethical issues, and extends it with previously overlooked items. It is focused on so-called intelligent environments, i.e. context-sensitive services to humans in the physical space. Although the Active2Gether system does not fall exactly within the definition of an intelligent environment (Augusto et al., 2013), the ethical principles are largely similar. Each of the following subsections addresses one of the general principles from the framework (Jones et al., 2015), and a brief summary is presented at the end of the section.

#### 17.3.1 Non-maleficence and beneficence

The first ethical principle concerns the non-maleficent and beneficent purpose of the system. Clearly, the Active2Gether system was designed to increase the users’ welfare and quality of life by encouraging healthy behavior. By supporting users to choose a beneficial coaching
domain and to set a realistic goal, the system proactively offers opportunities to assist the users. The current implementation of the system does not include a mechanism to detect extreme physical activity levels and to advise users to slow down, but the system’s architecture does allow implementing this in a next version. Still, the Active2Gether system is built in such a way that it stimulates physical activity in the users’ daily life, and it does not reward extreme physical activities. Also, new goals are suggested in small increments, and the reasoning engine allows the users to stop being coached towards higher physical activity levels once they have reached the norms.

17.3.2 User-centricity

Another ethical principle is that the users’ wishes should be central in the development process. In order to develop a behavior change system that meets the users’ needs and wishes, focus group interviews were conducted to inform design choices (see Chapter 4). Also, as described in Chapter 15, the system was first pilot tested within the Active2Gether team and with people from the target population, in order to solve any technical issues and to tweak some functionalities. This way, we attempted to keep the user in the design loop. Still, the development process would ideally have had more iterations with the users. The findings from the evaluation study (in Chapter 16) provided helpful starting points for improving the system, and one can only imagine that more feedback moments would lead to further improvements of the intervention.

17.3.3 Multiple user groups

The principle of multiple user groups is related to the situation in which an intelligent system is used by multiple people at the same time. In case of intelligent environments, this is a valid consideration. However, as the Active2Gether system is designed to be run on the smartphone for one individual, this ethical aspect is not relevant.

17.3.4 Privacy

A term that is unquestionably brought up when discussing ethics of technology, is privacy. In a system like the Active2Gether app, privacy could be ensured by letting users decide the level of detail in which data is collected. This was taken into account during the development of the social comparison feature. For people granted permission to view their Facebook connections and who are connected to another Active2Gether user on Facebook, their full name is shown in the overview. Users who don’t know each other or who did not allow the Active2Gether system to look up their Facebook connections, only the first two letters of their first name are shown, thereby maintaining a level of anonymity.

An evident example that would benefit from more flexible is the location monitoring, which could be turned off or set to other time intervals or accuracy levels. Unfortunately, the current implementation of the Active2Gether system does not support such custom settings. Although users can turn off the location monitoring, this will undeniably deteriorate the quality of the coaching as well as the user experience. Similarly, physical activity data could also be stored at higher levels of aggregation, according to the users’ preferences. The privacy aspect is therefore an important focal point for improving the ethical aspects of the Active2Gether intervention.
17.3.5 Data protection

A related ethical concern is data protection. It involves the users being in charge of their data and information sharing, as well as informed consent to the use of personal data and good practices of data protection. The Active2Gether system did show the physical activity data on the dashboard, and a separate personal profile page showed some additional background information, such as the current coaching domain, the current goal and an overview of the reported significant locations. The users could, for example, change their current goal, but changing information about their significant locations required involvement from the team. Also, deleting their personal data or their complete account was not possible for the users themselves. Such functionalities should therefore be implemented and/or professionalized before employing the system in real-life scenarios.

Still, Active2Gether adhered to recognized principles of data protection. This was documented in a study protocol before commencement of the final study, and it was approved by the medical ethical committee of the VU medical center. Also, all participants were informed about the data collection, and they signed an informed consent before being admitted to the final study. When employing the Active2Gether in real-life settings outside academic purposes, such a procedure should still be in place to fully inform users of the personal data that is stored during their use of the system.

17.3.6 Security

Another ethical aspect related to privacy and data protection is the system’s security. It concerns the need to protect users and their information, the reliability and stability of the system and the security of the data transfer. In the Active2Gether system, unauthorized access to the users’ personal data is avoided by protecting all accounts and dashboards with a password. In addition, safe data transfer is ensured by making use of a secure HTTPS connection. Also, personally identifiable information (including the users’ email addresses and first names) and passwords are encrypted, using either a MD5 hash function or AES encryption (Daemen and Rijmen, 2013; Rivest, 1992). However, when employing an intervention like Active2Gether in real life, data that could be used indirectly to identify users should also be encrypted, in order to reduce the risk of any security breaches.

Although there were no signs of the Active2Gether system being unreliable or unstable during its use in the final study, more mechanisms should be in place to ensure its stability at potential higher demands.

17.3.7 Autonomy

The Active2Gether system aims to support its users’ in their efforts to achieve behavior change, while respecting their independence and autonomy. An evident example of this intent is the fact that the reasoning engine presents the users with suggestions, which can be overruled by them. For example, based on the physical activity data and context information, the system suggests a coaching domain to focus on, but the users decide whether they want to follow the advice or not. The same holds for setting a goal: the system does suggests a goal based on the user’s physical activity data, but the user can easily adjust it before confirmation.

Still, the user does not have unlimited influence. In the current implementation, it is not possible to put the system ‘on hold’ (i.e., not receive any coaching messages and/or
questions). It is imaginable that such a feature would be relevant if the user’s regular daily life is interrupted, for example when he/she is ill or on holidays. In addition, as suggested in some of the previous sections, the user’s autonomy would be further enhanced by enabling more custom settings for data collection and information sharing.

17.3.8 Transparency
Unlike with some intelligent environments, users of the Active2Gether system do not have to be given notice of the existence of the system, since it is impossible to be unaware of it. However, it would be recommended to inform users about the system’s capabilities, in order to shape their expectations and to inform them of any potential weaknesses. For example, as long as the system does not include a mechanism to detect extreme physical activity behavior, users should be pointed at their own responsibility to behave responsibly. This also holds for the other end of the scale: the system aims to support users to live a healthier life, but people in need of medical intervention should not rely on the Active2Gether system alone, as that goes beyond the aim and capability of the intervention.

In terms of data collection and processing, the Active2Gether system is fairly transparent. Since a substantial part of the intervention relies on self-monitoring, much of the collected data is relayed back to the user via the dashboard. That way, users are also made aware that this data is being collected. This also holds for the location data, although to a lesser extent. This data is not shown to the user, but it is used as trigger to ask the users questions about their travel modes to certain locations. Because of that, the users should be aware that their location is being monitored in the background. Also, all users have to grant permission to use their smartphone’s location services, so it never happens unnoticed. In case the user input on travel options would become obsolete because of other reliable methods for collecting this data, it would be recommended to implement a feature that reminds the user of this data being collected (e.g., showing the location data on the dashboard as well), in order to increase the system’s transparency.

One aspect of the Active2Gether system that could benefit from more transparency is the dashboard panel that shows a ranking of the weekly number of steps of the user and six other users. The selection of these other users is based on an algorithm that places high priority on the user’s preference for the direction of social comparison (cf. Chapter 12 and Chapter 14). This implies that users are often found at the top or bottom position of the ranking, regardless of (changes in) their physical activity behavior. The basis for this selection mechanism was not transparent to the users, and was also sometimes perceived as demotivational, as efforts to increase physical activity were not rewarded by a higher position in the ranking.

17.3.9 Equality, dignity and inclusiveness
The final ethical principle considers equality, dignity and inclusiveness. There is an ethical consideration with respect to the intended user groups. As argued throughout this thesis, the Active2Gether system focuses on young adults (aged between 18 and 30 years) and is only available for smartphones running on Android. Therefore, many of the intermediate studies and the final study were conducted with participants from this age group. In practice, many participants were highly educated and/or university students and living in urban areas. Consequently, the design of the Active2Gether system is optimized for this population, and might be less effective or appealing for other groups. This is an undesirable consequence,
especially since people with a low socioeconomic status are overall less physically active than the general population (Giles-Corti and Donovan, 2002), and they are harder to reach with health interventions (Iliffe et al., 2017). Ideally, the Active2Gether system should be extended and tested for other population groups and operating systems as well, so it can be equally beneficial for every potential user.

On the other hand, the system is inclusive in the sense that it is (potentially) very affordable. In principle, the intervention does not require any human involvement, and is therefore quite cheap to maintain and employ for many users. Therefore, the cost for downloading and using the Active2Gether app could be kept at a minimum (or even free). Currently, the Active2Gether system does rely on measurements from a Fitbit device, which might be a financial burden for some potential users, but an extension of the system to allow using measurements from the smartphone itself is feasible.

Summary

Overall, the Active2Gether intervention respects the ethical principles from the eFRIEND framework well. However, there is still some room for improvement if the intervention would become available to the public outside the academic setting. For example, more data stored on the system’s server should be secured by means of encryption, more custom settings for data collection and information sharing should be available to the users, and the intervention should be made suitable for a broader audience than the current target population.

17.4 Future work

The research presented in this thesis explored several aspects of using technology for behavior change support. Especially in such a relatively new discipline, there are always many opportunities for further investigation. This section addresses a number of promising directions for future research to follow up on this thesis.

First of all, the research into the requirements for mobile behavior change interventions (presented in Part II) is inevitably linked to the current state of the techniques and science. As modern technology and understanding of human behavior formation progresses, new theoretical insights and technological developments lead to new user preferences, and consequently new requirements for such interventions. Therefore, such analyses should be regularly replicated, in order to validate or update the latest findings. In addition, there is a need for in-depth reviews of effectiveness of mobile behavior change interventions. This would contribute to understanding which mobile behavior change interventions (or on a more detailed level, which behavior change techniques) are successful in effectuating behavior change. If sufficient data is available, such analyses could also reveal which type of intervention or technique is effective for which type of user. That way, a new generation of interventions could be designed in an even more tailored fashion.

With respect to the role of computational models in the development of mobile interventions for physical activity, additional validation studies of such models are recommended to further consolidate their reliability. Naturally, more reliable and realistic computational models imply to more veracious new insights in human behavior through simulations, as well as more effective behavior change support systems. Also, it would be interesting to compare the validity of computational models of behavior change based on different
theoretical frameworks. In addition, it is highly plausible that there are individual differences in the psychosocial and behavioral processes represented by these computational models, which can be characterized by differences in their parameter values. It should be investigated whether this is indeed the case, and if so, individual differences should be taken into account. These individual parameter values could either be estimated based on responses to (personality) questionnaires, or learned automatically from behavioral data. Allowing individually set parameters could then lead to better performance of the models and more truth value in their simulations.

Another interesting direction for future research in the domain of computational models of behavior change processes is related to the way they are generated. Even with parameters tuned individually based on user data, a computational model as presented in Part III follows a top-down approach. It is designed based on theory, which means that it makes of existing knowledge and that its internal dynamics and simulation outputs are interpretable. However, with the increasing availability of relevant data, it is interesting to investigate more bottom-up approaches as well. In other words, such models could be learned from data. This does require a large body of data from a large number of participants, but it could lead to unforeseen insights and potentially higher validity. Apart from building models of behavioral processes from data, similar methods could be used to directly derive prediction mechanisms for determining coaching actions of behavior change systems. If such approach is successful, this could render the use of computational models in reasoning engines of behavior change systems unnecessary, but at the same time, it could also lead to new insights that can be incorporated in top-down generated computational models.

As mentioned before, many social processes have been established to play a role in behavior formation and behavior change. The work in this thesis only scratches the surface of this area of research, so further research into these and other social influences should be conducted to investigate which processes could be beneficial (or detrimental) to people’s attempts at behavior change. On the one hand, such investigations would contribute to further building theory on this topic, by gaining new insights in human behavior. On the other hand, the research should also focus on development of behavior change interventions, by exploring how these processes could be applied to help change behavior and how the underlying principles could be operationalized.

Another valuable opportunity for future research concerns the further development of the Active2Gether system or similar innovative behavior change interventions. This thesis described the design of such an intervention in detail, and provided many ideas for improvement as well, for example in Chapter 14 and Chapter 16. In addition, the previous section discussed some guidelines for better adherence to ethical norms.

One of the main directions for potential improvement of the Active2Gether system is related to the augmentation of the ‘intelligence’ in its reasoning engine. Incorporating more intelligent methods could make the system more flexible and adaptive to the users, which would contribute to the intervention being perceived as a virtual personal coach. Such intelligence could be achieved by incorporating more data-based methods and using feedback mechanisms to update the coaching offered to the users. Feedback could be obtained directly from the users (e.g., by letting them indicate whether they liked a certain coaching message), or from user data about their behavior or underlying psychological determinants (i.e., by deriving whether some action had a visible effect). This way, the system could adapt to each individual users and to the users’ progress over time.
In addition, if sufficient data is available, principles used in recommender systems could be applied. For example, collaborative filtering is an algorithm based on the assumption that if two users have a similar opinion on something, then the opinion of one of those users on something else is a good predictor for the opinion of the other user. In other words, if a certain coaching message is effective for one user, and this user is similar to another user, then the message will probably also work for the other user. However, such approaches require a vast amount of data over a longer period of time, and probably a less complex system, in order to be able to straightforwardly link observations of effectiveness to specific messages or behavior change techniques.

Another approach to more intelligent coaching relies on more contextual awareness and personal relevance of the messages. For example, geo-fencing (i.e., monitoring a virtual perimeter in a geographic) could be used to trigger sending certain coaching messages, for example when a user arrives home. In addition, the coaching messages could incorporate more information about the users’ behavioral patterns, such as the days when they usually exercise or the time of the day when they usually leave home on workdays. This again could be learned from user data by automatically recognizing patterns in time series of their daily activities. Also, direct triggers to send certain coaching messages from the user data could be implemented, such as when reaching a daily or weekly goal. This does not necessarily require very sophisticated methods, but it could improve the perception of the intervention as a virtual coach. Also, such messages (that are triggered by locations, behavioral patterns or physical activity data) are likely to contain practical and personal feedback, and thereby increase engagement with the intervention. The emergence of the Internet of Things (IoT) leads to the availability of many new types of (more detailed) data that are relevant for monitoring and influencing behavior (Gubbi et al., 2013; Whitmore et al., 2015). This development will likely further advance the use of data-based techniques in behavior change systems.

A key strength of the Active2Gether system is that it is based on an architecture that can be adapted relatively easily. Each of its building blocks (such as the mechanism for data collection, the reasoning engine, or the dashboard interface), can be improved, extended or completely replaced with alternative solutions. That way, it allows for flexible implementation of the improvements described above.

17.5 Conclusion

The research presented in this thesis investigates several aspects of using technology to stimulate behavior change for physical activity, and it proposes the design of an intelligent physical activity promotion app (i.e., the Active2Gether system). In doing so, techniques from (mobile) technology and artificial intelligence are applied, as well as scientific knowledge from human-directed disciplines within psychology and social sciences. The role of social processes in establishing (and maintaining) healthy behavior is considered in particular.

In Part II, the state of the art of mobile behavior change interventions for physical activity and user preferences of the target population are discussed, in order to gain insight into the requirements for such interventions. The results show that it is important to make better use of available knowledge and technology: on the one hand, incorporating more behavior change techniques (that are associated with effectiveness), and on the other hand,
implementing more technology (that enables smarter and more tailored support). Also, intended users expressed their preference for an intervention that has the impression of a virtual personal coach.

Part III discusses the role of computational models in the development of mobile interventions for physical activity. Together, the chapters provide an in-depth exploration of a computational model for psychosocial influences on physical activity behavior, with a focus on how to apply it in a real-life behavior change system. Initial validation studies of the model show promising results, thereby justifying its incorporation in Active2Gether’s reasoning engine. Applying a parameter tuning algorithm to increase the diversity of the simulation outcomes further consolidates the model’s suitability for application in the reasoning engine.

In Part IV, the role of social processes in establishing behavior change is studied. It is shown that users who choose to join an online community (at some point in time) benefit more from a physical activity promotion program than users who do not participate in the community. Also, the validity of a computational model of social contagion is supported by the results of two different data analysis studies. In addition, this part contains an exploration of ways to exploit such processes in behavior change systems, i.e. by altering connections in a user’s social network to steer social contagion and by selecting other users for social comparison while taking the user’s preference into account. These findings provide important guidelines for developers of behavior change interventions.

Part V provides a complete description of the process of developing and evaluating an innovative mobile behavior change intervention for physical activity. Where relevant and possible, the results from the research presented in the preceding parts were incorporated into the system’s design. Also, several improvements for the development of such systems are suggested.

Overall, the work presented in this thesis contributes to the scientific advancement of the domain of intelligent behavior change interventions for physical activity, by investigating several approaches to incorporate the use of technology in analyzing, understanding and supporting human behavior. In addition, this thesis presents practical steps and insights with respect to the development of such an intelligent physical activity promotion intervention. We hope and expect that this work contributes to the further development of sophisticated physical activity interventions, and thereby to a healthier society.
References


Breton, Emily R., Bernard F. Fuemmeler, and Lorien C. Abroms (2011). “Weight loss — there is an app for that! But does it adhere to evidence-informed practices?” In: Translational behavioral medicine 1.4, pages 523–529.


Giles-Corti, Billie and Robert J. Donovan (2002). “Socioeconomic status differences in recreational physical activity levels and real and perceived access to a supportive physical environment”. In: Preventive medicine 35.6, pages 601–611.


Summary

Moving forward: Supporting physical activity behavior change through intelligent technology

Physical activity is an important prerequisite for global health. Despite the well-known benefits for both physical and mental health, approximately 50% of the adult population in western countries are less physically active than recommended by health authorities. Therefore, effective and engaging interventions are needed to increase and maintain physical activity levels. It is believed that modern (mobile) technology provides an opportunity to support people to become or remain physically active.

The research presented in this thesis investigates several aspects of using technology to stimulate behavior change for physical activity, and it proposes the design of an intelligent physical activity promotion app. In doing so, techniques from (mobile) technology and artificial intelligence are applied, as well as scientific knowledge from human-directed disciplines within psychology and social sciences. The role of social processes in establishing (and maintaining) healthy behavior is considered in particular.

First, the state of the art of mobile behavior change interventions for physical activity and user preferences of the target population are discussed, in order to gain insight into the requirements for such interventions. The results show that it is important to make better use of available knowledge and technology: on the one hand, incorporating more behavior change techniques (that are associated with effectiveness), and on the other hand, implementing more technology (that enables smarter and more tailored support). Research into the user preferences showed that they would like an intervention that has the impression of a virtual personal coach.

Second, the role of computational models in the development of mobile interventions for physical activity is discussed. A computational model for psychosocial influences on physical activity behavior is explored in depth, with a focus on how to apply it in a real-life behavior change system. Initial validation studies of the model show promising results, thereby justifying its incorporation in Active2Gether’s reasoning engine. Also, a parameter tuning algorithm is applied to increase the diversity of the simulation outcomes, which provides further support for applying the model in the reasoning engine.

Next, the role of social processes in establishing behavior change is studied. It is shown that users who choose to join an online community (at some point in time) benefit more from a physical activity promotion program than users who do not participate in the community. Also, the validity of a computational model of social contagion is supported by the results of two different data analysis studies. In addition, this thesis presents an exploration of ways to
exploit such processes in behavior change systems, i.e. by altering connections in a user’s social network to steer social contagion and by selecting other users for social comparison while taking the user’s preference into account. These findings provide important guidelines for developers of behavior change interventions.

Finally, this thesis provides a complete description of the process of developing and evaluating an innovative mobile behavior change intervention for physical activity. Also, several improvements for the development of such systems are suggested.

Overall, the work presented in this thesis contributes to the scientific advancement of the domain of intelligent behavior change interventions for physical activity, by investigating several approaches to incorporate the use of technology in analyzing, understanding and supporting human behavior. In addition, this thesis presents practical steps and insights with respect to the development of such an intelligent physical activity promotion intervention. We hope and expect that this work contributes to the further development of sophisticated physical activity interventions, and thereby to a healthier society.
Samenvatting

In beweging: Toepassingen van intelligente technologie voor het stimuleren van lichaamsbeweging

Voldoende lichaamsbeweging is een belangrijke voorwaarde voor gezondheid. Ondanks het feit dat de voordelen van beweging voor de fysieke en mentale gezondheid algemeen bekend zijn, beweegt ongeveer de helft van de volwassen bevolking in westerse landen minder dan door gezondheidsinstanties wordt aanbevolen. Er zijn daarom effectieve en boeiende interventies nodig om lichaamsbeweging te stimuleren. Moderne (mobiele) technologie biedt nieuwe mogelijkheden om mensen te helpen voldoende actief te worden of te blijven.

Dit proefschrift behandelt verschillende aspecten van het gebruik van technologie om gedragsverandering op het gebied van lichaamsbeweging te stimuleren. Daarnaast presenteert het een ontwerp van een intelligente app die de gebruikers motiveert meer te bewegen (het Active2Gether systeem). Daarbij worden technieken uit (mobiele) technologie en kunstmatige intelligentie toegepast, evenals wetenschappelijke kennis uit mensgerichte disciplines binnen de psychologie en de sociale wetenschappen. Ook wordt er uitgebreid aandacht besteed aan de rol van sociale processen bij het tot stand brengen en handhaven van gezond gedrag.

In het eerste deel wordt de stand van zaken rondom mobiele interventies voor gedragsverandering op het gebied van lichaamsbeweging in kaart gebracht, evenals de behoeften, wensen en voorkeuren van beoogde gebruikers. Hiermee wordt inzicht verkregen in de eisen aan zulke interventies. Uit de resultaten blijkt dat het belangrijk is om meer en beter gebruik te maken van de beschikbare kennis en technologie: enerzijds door meer gedragsveranderingstechnieken (die samenhangen met effectiviteit) toe te passen, en anderzijds door meer technologieën (die slimmere en persoonlijkere ondersteuning mogelijk maken) te implementeren. Onderzoek naar de voorkeuren van gebruikers liet zien dat ze een interventie willen die de rol heeft van een virtuele persoonlijke coach.

Ten tweede wordt de rol van computationele modellen in de ontwikkeling van mobiele interventies voor lichaamsbeweging besproken. Een computationeel model van psychosociale invloeden op lichamelijk gedrag wordt uitgebreid behandeld, met de nadruk op de toepassen ervan in een daadwerkelijk systeem voor gedragsverandering. Eerste studies naar de validiteit van het model tonen veelbelovende resultaten. Daarnaast wordt een parameter tuning algoritme toegepast om de variatie in de simulatie-uitkomsten te vergroten. Deze beide stappen dragen bij aan de betrouwbaarheid van het model, waarmee de keuze om het model te integreren in de zogenaamde ‘reasoning engine’ van het Active2Gether systeem kan worden verantwoord.
De rol van sociale processen in het tot stand brengen van gedragsverandering wordt ook behandeld. Data analyse van gebruikers van een online coachingprogramma laat zien dat gebruikers die er (ten tijde van de interventie of op een later moment) voor kiezen om deel te nemen aan de online community meer profiteren van het programma dan gebruikers die dat niet doen. Ook dragen twee data analyse studies bij aan de validiteit van een computationeel model van de verspreiding van emoties, attitudes en gedragingen door sociale netwerken, ook wel ‘social contagion’ (‘sociale besmetting’) genoemd. Daarnaast worden verschillende methoden om dergelijke processen toe te passen in gedragsveranderingssystemen onderzocht, bijvoorbeeld het wijzigen van verbindingen in een sociaal netwerk van een gebruiker om het proces van sociale besmetting te sturen en het selecteren van andere gebruikers voor sociale vergelijking, waarbij rekening wordt gehouden met de voorkeur van de gebruiker. Deze bevindingen dragen bij aan waardevolle richtlijnen voor ontwikkelaars van interventies voor gedragsverandering.

Tot slot presenteert dit proefschrift een volledige beschrijving van het proces van het ontwikkelen en evalueren van een innovatieve mobiele interventie voor gedragsverandering op het gebied van lichaamsbeweging. Daarbij wordt ook een aantal ideeën voor mogelijke verbeteringen van zulke systemen aangedragen.

Dit proefschrift draagt bij aan het vakgebied van intelligente interventies voor gedragsverandering op het gebied van lichaamsbeweging, door te onderzoeken hoe toepassingen van technologie kunnen helpen het gedrag van mensen te analyseren, te interpreteren en te beïnvloeden. Daarnaast presenteert dit proefschrift praktische stappen en inzichten rondom de ontwikkeling van een dergelijke intelligente interventie voor lichaamsbeweging. Wij hopen en verwachten dat dit werk daarom bijdraagt aan de verdere ontwikkeling van ge-avanceerde interventies die lichaamsbeweging stimuleren, en daarmee tot een gezondere samenleving.
V pohybu: Využití inteligentní technologie pro změnu chování v oblasti fyzické aktivity

Dostatečný pohyb je důležitým aspektem zdraví. Navzdory skutečnosti, že prospěšnost pohybu pro tělesné a duševní zdraví je dobře známa, přibližně 50% dospělé populace v západních zemích se pohybuje méně než doporučují lékaři. Účinné zásahy jsou tedy nezbytné k podpoře pohybu. Moderní mobilní technologie nabízí lidem nové možnosti stát se fyzicky aktivními, a nebo aktivními zůstat.

Tato disertační práce se zabývá různými aspekty využití technologie ke stimulaci změn chování v oblasti fyzické aktivity. Kromě toho se zde představuje návrh inteligentní aplikace, která motivuje uživatele, aby se více pohyboval: systém Active2Gether. Přítom je použito mobilní technologie a umělé inteligence, stejně jako vědeckých poznatků z psychologie a společenských věd. Velká pozornost se věnuje také roli sociálních procesů při vytváření a udržování zdravého životního stylu.

V první části je zmapována oblast mobilních intervencí na změny chování v oblasti fyzické aktivity, stejně jako potřeby, přání a preference cílených uživatelů. Výsledky této sekce ukazují, že je důležité více a lépe využít dostupných znalostí a technologie: na jedné straně uplatnit více behaviorálních technik v souvislosti s účinností, na druhé straně je třeba začlenit víc technologie, což umožní inteligentnější a individualizovanou pomoc. Výzkum totiž ukazuje, že uživatelé dávají přednost asistenci, která vyvolává dojem virtuálního osobního trenéra.

Druhá část práce pojednává o úloze výpočetních modelů ve vývoji mobilních intervencí v oblasti fyzické aktivity. Práce se zabývá rozborem výpočetního modelu interakce psychologických a sociálních vlivů na fyzickou aktivitu s důrazem na jeho uplatnění při změně chování. Počáteční studie o účinnosti modelu ukazují slibné výsledky. Tudíž, začlenění tohoto modelu do systému Active2Gether je oprávněné.

Ve třetí části se disertační práce zabývá vlivem okolí na změny chování. Rozbor dat uživatelů online koučovacího programu ukazuje, že pro uživatele, kteří se zapojili do online komunity je program účinnější. Další dva rozbory dat objasňují mechanismus šíření emocí, postojů a chování prostřednictvím sociálních sítí, též známý jako „sociální nákaza“. Vedle toho jsou zkoumány různé metody v procesech ovlivňování chování. Tyto poznatky pak významně přispívají při vývoji intervencí na změně chování.

Na závěr tato disertační práce představuje úplný popis procesu vývoje a vyhodnocení výsledků inovativního mobilního systému na změnu chování v oblasti fyzické aktivity. Navíc je zde prezentována řada návrhů na možné zlepšení takovýchto systémů.
Tato disertace přispívá v oblasti inteligentních intervencí na změny chování v oblasti fyzické aktivity tím, že zkoumá, jak může současná technologie pomoci analyzovat, interpretovat a ovlivňovat chování lidí. Kromě toho tato práce shrnuje praktické kroky a poznatky ve vývoji inteligentní intervence. Je proto možné očekávat, že tato práce přispěje k dalšímu rozvoji systémů, které stimulují tělesnou aktivitu a tím vytvářejí zdravější společnost.
In contrast to popular belief, doing PhD research does not have to be a solitary journey. It certainly was not in my case: this thesis is the result of many collaborations with and much support from a wide range of people.

First of all, I want to thank Jan for giving me the opportunity to do the research that resulted in this thesis. You gave me the freedom to become an independent researcher, which helped me to grow as a person and build confidence in my own abilities at the same time. Even though we have not collaborated intensively, it was pleasant to work with you.

I am also grateful to Michel for his guidance in shaping the path of my PhD research. While working with you, I learned a lot about project management, about combining practical work and scientific value, and about keeping the bigger picture in mind. Your enthusiasm for taking on new challenges has been contagious and motivational.

A special thanks goes out to Natalie, who has become a close friend both within and outside work. It is safe to say that I would not have started my journey as a PhD student without your example, and I’m very glad for having your support and friendship throughout.

I would also like to sincerely thank the members of my reading committee for spending some of their precious time on reviewing my thesis. I am very grateful for your feedback and your positive remarks.

The Active2Gether project – and consequently much of the contents of this thesis – has been a team effort. Therefore, I want to thank all members of the Active2Gether project and all other co-authors for our successful collaborations: Anouk, Saskia, Hans, Adnan, Eric, Michel, Aart, Natalie, Jan, Maartje, Mirjam, Daniëlle, Robin and Anita. Some people deserve a special mention: Anouk and Saskia for their valuable contributions from their field of expertise, and Adnan and Eric for their substantial collaboration on the technical parts and data analyses.

In addition, I’ve had a lot of support from other colleagues and fellow PhD students. This includes current and former members of the Behavioural Informatics research group, colleagues in the Artificial Intelligence section, researchers in other projects of the Healthy Lifestyle Solutions program and many people I’ve met during conferences and courses. It was great to share experiences, have interesting discussions and getting to know all of you.

It goes without saying that my family and friends have been a great support system over the last few years. You have all been invaluable to me, whether for your patience when I needed to vent or for your company during fun and relaxing times when my thesis was the last thing on my mind. You know who you are, thank you.

Amsterdam, September 12th 2017
List of publications

2017

- Anouk Middelweerd, Saskia J. te Velde, Julia S. Mollee, Michel C.A. Klein, and Johannes Brug (2018). “Development of Active2Gether: An app-based intervention combining evidence-based behavior change techniques with a model-based reasoning system to promote physical activity among young adults”. In: *Journal of Medical Internet Research*

* For articles marked with an asterisk, the authors can be regarded to have made equal contributions to the work, and are therefore in alphabetical order.

2016


2015


2014


2013


2012


<table>
<thead>
<tr>
<th>Year</th>
<th>Title</th>
<th>Author</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011-01</td>
<td>Variational Algorithms for Bayesian Inference in Latent Gaussian Models</td>
<td>Botond Cseke</td>
<td>RUN</td>
</tr>
<tr>
<td>2011-02</td>
<td>Organizing Agent Organizations. Syntax and Operational Semantics of an Organization-Oriented Programming Language</td>
<td>Nick Tinnemeier</td>
<td>UU</td>
</tr>
<tr>
<td>2011-03</td>
<td>Compositional Design and Verification of Component-Based Information Systems</td>
<td>Jan Martijn van der Werf</td>
<td>TUe</td>
</tr>
<tr>
<td>2011-04</td>
<td>Insights in Reinforcement Learning; Formal analysis and empirical evaluation of temporal-difference learning algorithms</td>
<td>Hado van Hasselt</td>
<td>UU</td>
</tr>
<tr>
<td>2011-05</td>
<td>Enterprise Architecture Coming of Age - Increasing the Performance of an Emerging Discipline</td>
<td>Base van der Raadt</td>
<td>VU</td>
</tr>
<tr>
<td>2011-06</td>
<td>Semantically-Enhanced Recommendations in Cultural Heritage</td>
<td>Yujia Cao</td>
<td>UT</td>
</tr>
<tr>
<td>2011-07</td>
<td>Multimodal Information Presentation for High Load Human Computer Interaction</td>
<td>Yiwen Wang</td>
<td>TUe</td>
</tr>
<tr>
<td>2011-08</td>
<td>BDI-based Generation of Robust Task-Oriented Dialogues</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011-09</td>
<td>Designing for Awareness: An Experience-focused HCI Perspective</td>
<td>Tim de Jong</td>
<td>OU</td>
</tr>
<tr>
<td>2011-10</td>
<td>Cloud Content Contention</td>
<td>Bart Bogaut</td>
<td>UvT</td>
</tr>
<tr>
<td>2011-11</td>
<td>AirPort under Control, Multiagent Scheduling for Airport Ground Handling</td>
<td>DhaVal Vyaa</td>
<td>UT</td>
</tr>
<tr>
<td>2011-12</td>
<td>Behavioral Finance and Agent-Based Artificial Markets</td>
<td>Milan Lovric</td>
<td>EUR</td>
</tr>
<tr>
<td>2011-13</td>
<td>The Meaning of Structure: the Value of Link Evidence for Information Retrieval</td>
<td>Marijn Koolen</td>
<td>UvA</td>
</tr>
<tr>
<td>2011-14</td>
<td>Selective Search in Games of Different Complexity</td>
<td>Maarten Schadd</td>
<td>UM</td>
</tr>
<tr>
<td>2011-15</td>
<td>Exploring Topic Structure: Coherence, Diversity and Relatedness</td>
<td>Jiyn He</td>
<td>UvA</td>
</tr>
<tr>
<td>2011-16</td>
<td>Strategic Decision-Making in complex games</td>
<td>Mark Ponsen</td>
<td>UM</td>
</tr>
<tr>
<td>2011-17</td>
<td>The Mind’s Eye on Personal Profiles</td>
<td>Ellen Rusman</td>
<td>OU</td>
</tr>
<tr>
<td>2011-18</td>
<td>Guiding service-oriented software engineering - A view-based approach</td>
<td>Qing Gu</td>
<td>VU</td>
</tr>
<tr>
<td>2011-19</td>
<td>Modularization and Specification of Service-Oriented Systems</td>
<td>Linda Terlouw</td>
<td>TUD</td>
</tr>
<tr>
<td>2011-20</td>
<td>System Evaluation of Archival Description and Access</td>
<td>Junie Zhang</td>
<td>UvA</td>
</tr>
<tr>
<td>2011-21</td>
<td>Finding People and their Utterances in Social Media</td>
<td>Wouter Weerkamp</td>
<td>UU</td>
</tr>
<tr>
<td>2011-22</td>
<td>Behavior Generation for Interpersonal Coordination with Virtual Humans On Specifying, Scheduling and Realizing Multimodal Virtual Human Behavior</td>
<td>Herwin van Welbergen</td>
<td>UT</td>
</tr>
<tr>
<td>2011-23</td>
<td>Analysis and Validation of Models for Trust Dynamics</td>
<td>Syed Waqar ul Gounain Jaffry</td>
<td>VU</td>
</tr>
<tr>
<td>2011-24</td>
<td>Dynamic website optimization through autonomous management of design patterns</td>
<td>Aniel Bhulai</td>
<td>VU</td>
</tr>
<tr>
<td>2011-25</td>
<td>Effective Focused Retrieval by Exploiting Query Context and Document Structure</td>
<td>Rianne Kaptein</td>
<td>UvA</td>
</tr>
<tr>
<td>2011-26</td>
<td>Discrimination-aware Classification</td>
<td>Ellen Rusman</td>
<td>OU</td>
</tr>
<tr>
<td>2011-27</td>
<td>Affective Signal Processing (ASP): Unraveling the mystery of emotions</td>
<td>Egon van den Broek</td>
<td>UvA</td>
</tr>
<tr>
<td>2011-28</td>
<td>Computational and Game-Theoretic Approaches for Modeling Bounded Rationality</td>
<td>Ludo Wiltman</td>
<td>EUR</td>
</tr>
<tr>
<td>2011-29</td>
<td>Methodological Advances in Bibliometric Mapping of Science</td>
<td>Nees-Jan van Eck</td>
<td>EUR</td>
</tr>
<tr>
<td>2011-30</td>
<td>Arguing to Motivate Decisions</td>
<td>Tom van der Weide</td>
<td>UU</td>
</tr>
<tr>
<td>2011-31</td>
<td>Explaining Agent Behavior in Virtual Training</td>
<td>Paolo Turrini</td>
<td>UU</td>
</tr>
<tr>
<td>2011-32</td>
<td>Maaike Harbers</td>
<td>UU</td>
<td></td>
</tr>
<tr>
<td>2011-33</td>
<td>Experiments in serious game design: a cognitive</td>
<td>Erik van der Spek</td>
<td>UU</td>
</tr>
</tbody>
</table>
2011-37 Adriana Burlutiu (RUN)
Machine Learning for Pairwise Data, Applications for Preference Learning and Supervised Network Inference

2011-38 Nyree Lemmens (UM)
Bee-inspired Distributed Optimization

2011-39 Joost Westra (UU)
Organizing Adaptation using Agents in Serious Games

2011-40 Viktor Clerc (VU)
Architectural Knowledge Management in Global Software Development

2011-41 Luan Ibraimi (UT)
Cryptographically Enforced Distributed Data Access Control

2011-42 Michal Sindlar (UU)
Explaining Behavior through Mental State Attribution

2011-43 Henk van der Schuur (UU)
Process Improvement through Software Operation Knowledge

2011-44 Boris Reuderink (UT)
Robust Brain-Computer Interfaces

2011-45 Herman Stehouwer (UvT)
Statistical Language Models for Alternative Sequence Selection

2011-46 Beibei Hu (TUD)
Towards Contextualized Information Delivery: A Rule-based Architecture for the Domain of Mobile Police Work

2011-47 Azizi Bin Ab Aziz (VU)
Exploring Computational Models for Intelligent Support of Persons with Depression

2011-48 Mark Ter Maat (UT)
Response Selection and Turn-taking for a Sensitive Artificial Listening Agent

2011-49 Andrea Niculescu (UT)
Conversational interfaces for task-oriented spoken dialogues: design aspects influencing interaction quality

2012-01 Terry Kakeeto (UvT)
Relationship Marketing for SMEs in Uganda

2012-02 Muhammad Umair (VU)
Adaptivity, Emotion, and Rationality in Human and Ambient Agent Models

2012-03 Adam Vanya (VU)
Supporting Architecture Evolution by Mining Software Repositories

2012-04 Jurriaan Sauer (UU)
Development of Content Management System-based Web Applications

2012-05 Marijn Plomp (UU)
Maturing Interorganisational Information Systems

2012-06 Wolfgang Reinhardt (OU)
Awareness Support for Knowledge Workers in Research Networks

2012-07 Rianne van Lam balgen (VU)
When the Going Gets Tough: Exploring Agent-based Models of Human Performance under Demanding Conditions

2012-08 Gerben de Vries (UvA)
Kernel Methods for Vessel Trajectories

2012-09 Ricardo Neisse (UT)
Trust and Privacy Management Support for Context-Aware Service Platforms

2012-10 David Smits (TUt)
Towards a Generic Distributed Adaptive Hypermedia Environment

2012-11 J.C.B. Rantham Prabhakara (TUe)
Process Mining in the Large: Preprocessing, Discovery, and Diagnostics

2012-12 Kees van der Slauijs (TUe)
Model Driven Design and Data Integration in Semantic Web Information Systems

2012-13 Suleman Shahid (UvT)
Fun and Face: Exploring non-verbal expressions of emotion during playful interactions

2012-14 Evgeny Knutov (TUe)
Generic Adaptation Framework for Unifying Adaptive Web-based Systems

2012-15 Natalie van der Wal (VU)
Social Agents. Agent-Based Modelling of Integrated Internal and Social Dynamics of Cognitive and Affective Processes.

2012-16 Fiemke Both (VU)
Helping people by understanding them - Ambient Agents supporting task execution and depression treatment

2012-17 Amal Elgammal (UvT)
Towards a Comprehensive Framework for Business Process Compliance

2012-18 Eltjo Poort (VU)
Improving Solution Architecting Practices

2012-19 Helen Schonenberg (TUe)
What’s Next? Operational Support for Business Process Execution

2012-20 Ali Bahramisharif (RUN)
Covert Visual Spatial Attention, a Robust Paradigm for Brain-Computer Interfacing

2012-21 Roberto Cornacchia (TUe)
Querying Sparse Matrices for Information Retrieval

2012-22 Thijs Vis (UvT)
Intelligence, politie en veiligheidsdienst: verenigbare grootheden?

2012-23 Christian Muehl (UT)
Toward Affective Brain-Computer Interfaces: Exploring the Neurophysiology of Affect during Human Media Interaction

2012-24 Laurens van der Werff (UT)
Evaluation of Noisy Transcripts for Spoken Document Retrieval

2012-25 SiIja Eckartz (UT)
Managing the Business Case Development in Inter-Organizational IT Projects: A Methodology and its Application

2012-26 Emile de Maat (UvA)
Making Sense of Legal Text

2012-27 Hayrettin Gurkok (UT)
Mind the Sheep! User Experience Evaluation & Brain-Computer Interface Games

2012-28 Nancy Pascall (UvT)
Engendering Technology Empowering Women

2012-29 Almer Tigelaar (UT)
Peer-to-Peer Information Retrieval

2012-30 Alina Pomeranz (TUe)
Designing Human-Centered Systems for Reflective Decision Making

2012-31 Emily Bagaru kaya (RUN)
A Learning by Construction Approach for Higher Order Cognitive Skills Improvement, Building Capacity and Infrastructure

2012-32 Wietseke Visser (TUe)
Qualitative multi-criteria preference representation and reasoning

2012-33 Nory Sie (OUN)
Coalitions in Cooperation Networks (COCOON)

2012-34 Pavol Jancura (RUN)
Evolutionary analysis in PPI networks and applications
2012-35 Evert Haasdijk (VU)
Never Too Old To Learn – On-line Evolution of Controllers in Swarm- and Modular Robotics

2012-36 Denis Ssebugwawo (RUN)
Analysis and Evaluation of Collaborative Modeling Processes

2012-37 Agnes Nakakawa (RUN)
A Collaboration Process for Enterprise Architecture Creation

2012-38 Selmar Smit (VU)
Parameter Tuning and Scientific Testing in Evolutionary Algorithms

2012-39 Hassan Fatemi (UT)
Risk-aware design of value and coordination networks

2012-40 Agus Gunawan (UvT)
Information Access for SMEs in Indonesia

2012-41 Sebastian Kelle (OU)
Game Design Patterns for Learning

2012-42 Dominique Verpoorten (OU)
Reflection Amplifiers in self-regulated Learning

2012-43 Withdrawn

2012-44 Anna Tordai (VU)
On Combining Alignment Techniques

2012-45 Benedikt Kratz (UvT)
A Model and Language for Business-aware Transactions

2012-46 Simon Carter (UvA)
Exploration and Exploitation of Multilingual Data for Statistical Machine Translation

2012-47 Manos Tsagkias (UvA)
Mining Social Media: Tracking Content and Predicting Behavior

2012-48 Jorn Bakker (TUe)
Handling Abrupt Changes in Evolving Time-series Data

2012-49 Michael Kaisers (UM)
Learning against Learning - Evolutionary dynamics of reinforcement learning algorithms in strategic interactions

2012-50 Steven van Kervel (TUD)
Ontology driven Enterprise Information Systems Engineering

2012-51 Jeroen de Jong (TUD)
Heuristics in Dynamic Scheduling; a practical framework with a case study in elevator dispatching

2013-01 Viorel Milea (EUR)
News Analytics for Financial Decision Support

2013-02 Erietta Liarou (CWI)
MonetDB/DataCell: Leveraging the Column-store Database Technology for Efficient and Scalable Stream Processing

2013-03 Szymon Klarmann (VU)
Reasoning with Contexts in Description Logics

2013-04 Chetan Yadati (TUD)
Coordinating autonomous planning and scheduling

2013-05 Dulce Pumareja (UT)
Groupware Requirements Evolutions Patterns

2013-06 Romulo Goncalves (CWI)
The Data Cyclotron: Juggling Data and Queries for a Data Warehouse Audience

2013-07 Giel van Lankveld (UvT)
Quantifying Individual Player Differences

2013-08 Giel van Lankveld (UvT)
Making enemies: cognitive modeling for opponent agents in fighter pilot simulators

2013-09 Fabio Gori (RUN)
Metagenomic Data Analysis: Computational Methods and Applications
2013-35 Abdallah El Ali (UvA)
Minimal Mobile Human Computer Interaction

2013-36 Than Lam Hoang (TUe)
Pattern Mining in Data Streams

2013-37 Dirk Börner (OUN)
Ambient Learning Displays

2013-38 Eelco den Heijer (VU)
Autonomous Evolutionary Art

2013-39 Joop de Jong (TUD)
A Method for Enterprise Ontology based Design of Enterprise Information Systems

2013-40 Pim Nijssen (UM)
Monte-Carlo Tree Search for Multi-Player Games

2013-41 Jochem Liem (UvA)
Supporting the Conceptual Modelling of Dynamic Systems: A Knowledge Engineering Perspective on Qualitative Reasoning

2013-42 Léon Planken (TUD)
Algorithms for Simple Temporal Reasoning

2013-43 Marc Bron (UvA)
Exploration and Contextualization through Interaction and Concepts

2014-01 Nicola Barile (UU)
Studies in Learning Monotone Models from Data

2014-02 Fiona Tuliyano (RUN)
Combining System Dynamics with a Domain Modeling Method

2014-03 Sergio Raul Duarte Torres (UT)
Information Retrieval for Children: Search Behavior and Solutions

2014-04 Hanna Joachmann-Mannak (UT)
Websites for children: search strategies and interface design - Three studies on children’s search performance and evaluation

2014-05 Jurriaan van Reijisen (UU)
Knowledge Perspectives on ADVancing Dynamic Capability

2014-06 Damian Tamburri (VU)
Supporting Networked Software Development

2014-07 Arya Adriansyah (TUE)
Aligning Observed and Modeled Behavior

2014-08 Samur Araujo (TUD)
Data Integration over Distributed and Heterogeneous Data Endpoints

2014-09 Philip Jackson (UvT)
Toward Human-Level Artificial Intelligence: Representation and Computation of Meaning in Natural Language

2014-10 Ivan Salvador Razo Zapata (VU)
Service Value Networks

2014-11 Jaanneke van der Zwaan (TUD)
An Empathic Virtual Buddy for Social Support

2014-12 Willem van Willigen (VU)
Agent-Based Support for Behavior Change: Models and Applications in Health and Safety Domains

2014-13 Arlette van Wissen (VU)
An Empathic Virtual Buddy for Social Support

2014-14 Naser Ayat (UvA)
On Entity Resolution in Probabilistic Data

2014-15 Jasmina Maric (UvT)
Making brain-computer interfaces better: improving usability through post-processing.

2014-16 Krystyna Milian (VU)
Supporting Trail recruitment and design by automatically interpreting eligibility criteria

2014-17 Kathrin Dentler (VU)
Computing healthcare quality indicators automatically; Secondary Use of Patient Data and Semantic Interoperability

2014-18 Mattijs Ghijsen (UvA)
Methods and Models for the Design and Study of Dynamic Agent Organizations

2014-19 Vinicius Ribeiro (TUe)
Adaptive Hypermedia Courses: Qualitative and Quantitative Evaluation and Tool Support

2014-20 Mena Habib (UT)
Named Entity Extraction and Disambiguation for Informal Text: The Missing Link

2014-21 Kassidy Clark (TUD)
Negotiation and Monitoring in Open Environments

2014-22 Marieke Peeters (UU)
Personalized Educational Games - Developing agent-supported scenario-based training

2014-23 Eleftherios Sidirougos (UvA/CWI)
Space Efficient Indexes for the Big Data Era

2014-24 Davide Ciolli (VU)
Trust ing Semi-structured Web Data

2014-25 Martijn Lappenschaar (RUN)
New network models for the analysis of disease interaction

2014-26 Tim Baarslag (TUD)
What to Bid and When to Stop

2014-27 Rui Jorge Almeida (EUR)
Conditional Density Models Integrating Fuzzy and Probabilistic Representations of Uncertainty

2014-28 Anna Chmielowiec (VU)
Decentralized k-Clique Matching

2014-29 Jaap Kabbeldijk (UU)
Variability in Multi-Tenant Enterprise Software

2014-30 Peter de Cock (UvT)
Anticipating Criminal Behaviour

2014-31 Leo van Moergeestel (UU)
Agent Technology in Agile Multiparallel Manufacturing and Product Support

2014-32 Martijn van Eijkhof (CWI/TUD)
Contextual Multidimensional Relevance Models

2014-33 Kevin Vlaanderen (UU)
Supporting Process Improvement using Method
<table>
<thead>
<tr>
<th>Title</th>
<th>Author</th>
<th>Year</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Web, Speech Technologies and Rural Development in West Africa - An ICT4D Approach</td>
<td>Ravi Khadka (UU)</td>
<td>2016</td>
<td>(UU)</td>
</tr>
<tr>
<td>Revisiting Legacy Software System Modernization</td>
<td>Steffen Michels (RUN)</td>
<td>2016</td>
<td>(RUN)</td>
</tr>
<tr>
<td>Hybrid Probabilistic Logics - Theoretical Aspects, Algorithms and Experiments</td>
<td>Guangliang Li (UvA)</td>
<td>2016</td>
<td>(UvA)</td>
</tr>
<tr>
<td>Socially Intelligent Autonomous Agents that Learn from Human Reward</td>
<td>Berend Weel (VU)</td>
<td>2016</td>
<td>(VU)</td>
</tr>
<tr>
<td>Towards Embodied Evolution of Robot Organisms</td>
<td>Albert Menoño Peñuela (VU)</td>
<td>2016</td>
<td>(VU)</td>
</tr>
<tr>
<td>Refining Statistical Data on the Web</td>
<td>Julia Efremova (Tu/e)</td>
<td>2016</td>
<td>(Tu/e)</td>
</tr>
<tr>
<td>Mining Social Structures from Genealogical Data</td>
<td>Daan Odijk (UvA)</td>
<td>2016</td>
<td>(UvA)</td>
</tr>
<tr>
<td>Context &amp; Semantics in News &amp; Web Search</td>
<td>Alejandro Moreno Celleri (UT)</td>
<td>2016</td>
<td>(UT)</td>
</tr>
<tr>
<td>From Traditional to Interactive Playspaces: Automatic Analysis of Player Behavior in the Interactive Tag Playground</td>
<td>Grace Lewis (VU)</td>
<td>2016</td>
<td>(VU)</td>
</tr>
<tr>
<td>Software Architecture Strategies for Cyber-Foraging Systems</td>
<td>Fei Cai (UvA)</td>
<td>2016</td>
<td>(UvA)</td>
</tr>
<tr>
<td>Query Auto Completion in Information Retrieval</td>
<td>Brend Wanders (UT)</td>
<td>2016</td>
<td>(UT)</td>
</tr>
<tr>
<td>Repurposing and Probabilistic Integration of Data; An Iterative and data model independent approach</td>
<td>Julia Kiseleva (TUe)</td>
<td>2016</td>
<td>(TUe)</td>
</tr>
<tr>
<td>Using Contextual Information to Understand Searching and Browsing Behavior</td>
<td>Dilhan Thilakarathne (VU)</td>
<td>2016</td>
<td>(VU)</td>
</tr>
<tr>
<td>In or Out of Control: Exploring Computational Models to Study the Role of Human Awareness and Control in Behavioural Choices, with Applications in Aviation and Energy Management Domains</td>
<td>Wen Li (TUD)</td>
<td>2016</td>
<td>(TUD)</td>
</tr>
<tr>
<td>Understanding Geo-spatial Information on Social Media</td>
<td>Mingxin Zhang (TUD)</td>
<td>2016</td>
<td>(TUD)</td>
</tr>
<tr>
<td>Large-scale Agent-based Social Simulation - A study on epidemic prediction and control</td>
<td>Nicolas Honing (TUD)</td>
<td>2016</td>
<td>(TUD)</td>
</tr>
<tr>
<td>Peak reduction in decentralised electricity systems - Markets and prices for flexible planning</td>
<td>Ruud Mattheij (UvT)</td>
<td>2016</td>
<td>(UvT)</td>
</tr>
<tr>
<td>The Eyes Have It</td>
<td>Mohammad Khelghati (UT)</td>
<td>2016</td>
<td>(UT)</td>
</tr>
<tr>
<td>Deep web content monitoring</td>
<td>Eelco Vriezeekolk (UT)</td>
<td>2016</td>
<td>(UT)</td>
</tr>
<tr>
<td>Assessing Telecommunication Service Availability</td>
<td>Peter Bloem (UvA)</td>
<td>2016</td>
<td>(UvA)</td>
</tr>
<tr>
<td>Risks for Crisis Organisations</td>
<td>Dennis Schunselaar (TUe)</td>
<td>2016</td>
<td>(TUe)</td>
</tr>
<tr>
<td>Configurable Process Trees: Elicitation, Analysis, and Enactment</td>
<td>Zhuochun Ren (UvA)</td>
<td>2016</td>
<td>(UvA)</td>
</tr>
<tr>
<td>Monitoring Social Media: Summarization, Classification and Recommendation</td>
<td>Daphne Karrerman (UT)</td>
<td>2016</td>
<td>(UT)</td>
</tr>
<tr>
<td>Beyond R2D2: The design of nonverbal interaction behavior optimized for robot-specific morphologies</td>
<td>Giovanni Sileno (UvA)</td>
<td>2016</td>
<td>(UvA)</td>
</tr>
<tr>
<td>Aligning Law and Action - a conceptual and computational inquiry</td>
<td>Andrea Minuto (UT)</td>
<td>2016</td>
<td>(UT)</td>
</tr>
<tr>
<td>Materials that Matter - Smart Materials meet Art &amp; Interaction Design</td>
<td>Merijn Bruinjes (UT)</td>
<td>2016</td>
<td>(UT)</td>
</tr>
<tr>
<td>Believable Suspect Agents; Response and Interpersonal Style Selection for an Artificial Suspect</td>
<td>Christian Detweiler (TUD)</td>
<td>2016</td>
<td>(TUD)</td>
</tr>
<tr>
<td>Accounting for Values in Design</td>
<td>Thomas King (TUD)</td>
<td>2016</td>
<td>(TUD)</td>
</tr>
<tr>
<td>Governing Governance: A Formal Framework for Analysing Institutional Design and Enactment Governance</td>
<td>Spyros Martzoukos (UvA)</td>
<td>2016</td>
<td>(UvA)</td>
</tr>
<tr>
<td>Combinatorial and Compositional Aspects of Bilingual Aligned Corpora</td>
<td>Saskia Koldijk (RUN)</td>
<td>2016</td>
<td>(RUN)</td>
</tr>
<tr>
<td>Context-Aware Support for Stress Self-Management: From Theory to Practice</td>
<td>Thibault Sellam (UvA)</td>
<td>2016</td>
<td>(UvA)</td>
</tr>
<tr>
<td>Automatic Assistants for Database Exploration</td>
<td>Bram van der Laar (UT)</td>
<td>2016</td>
<td>(UT)</td>
</tr>
<tr>
<td>Experiencing Brain-Computer Interface Control</td>
<td>Jorge Gallego Perez (UT)</td>
<td>2016</td>
<td>(UT)</td>
</tr>
<tr>
<td>Robots to Make you Happy</td>
<td>Christina Weber (UL)</td>
<td>2016</td>
<td>(UL)</td>
</tr>
<tr>
<td>Real-time foresight - Preparedness for dynamic innovation networks</td>
<td>Tanja Buttlar (TUD)</td>
<td>2016</td>
<td>(TUD)</td>
</tr>
<tr>
<td>Collecting Lessons Learned</td>
<td>Gleb Polevoy (TUD)</td>
<td>2016</td>
<td>(TUD)</td>
</tr>
<tr>
<td>Participation and Interaction in Projects. A Game-Theoretic Analysis</td>
<td>Yan Wang (UT)</td>
<td>2016</td>
<td>(UT)</td>
</tr>
<tr>
<td>The Bridge of Dreams: Towards a Method for Operational Performance Alignment in IT-enabled Service Supply Chains</td>
<td>Jan-Iaap Oerlemans (UL)</td>
<td>2017</td>
<td>(UL)</td>
</tr>
<tr>
<td>Investigating Cybercrime</td>
<td>Sjoerd Timmer (UU)</td>
<td>2017</td>
<td>(UU)</td>
</tr>
<tr>
<td>Designing and Understanding Forensic Bayesian Networks using Argumentation</td>
<td>Daniël Harold Telgen (UU)</td>
<td>2017</td>
<td>(UU)</td>
</tr>
<tr>
<td>Grid Manufacturing: A Cyber-Physical Approach with Autonomous Products and Reconfigurable Manufacturing Machines</td>
<td>Mrunal Gawade (CWI)</td>
<td>2017</td>
<td>(CWI)</td>
</tr>
<tr>
<td>Multi-Core Parallelism in a Column-Store</td>
<td>Mahdieh Shadi (UvA)</td>
<td>2017</td>
<td>(UvA)</td>
</tr>
<tr>
<td>Collaboration Behavior</td>
<td>Damir Vancic (EUR)</td>
<td>2017</td>
<td>(EUR)</td>
</tr>
<tr>
<td>Intelligent Information Systems for Web Product Search</td>
<td>Noelia Bertens (UU)</td>
<td>2017</td>
<td>(UU)</td>
</tr>
<tr>
<td>Insight in Information: from Abstract to Anomaly</td>
<td>Rob Konijn (VU)</td>
<td>2017</td>
<td>(VU)</td>
</tr>
<tr>
<td>Detecting Interesting Differences: Data Mining in Health Insurance Data using Outlier Detection and Subgroup Discovery</td>
<td>Dong Nguyen (UT)</td>
<td>2017</td>
<td>(UT)</td>
</tr>
<tr>
<td>Text as Social and Cultural Data: A Computational Perspective on Variation in Text</td>
<td>Robert van Delden (UT)</td>
<td>2017</td>
<td>(UT)</td>
</tr>
<tr>
<td>Steering Interactive Play Behavior</td>
<td>Floriant Kunneman (RUN)</td>
<td>2017</td>
<td>(RUN)</td>
</tr>
<tr>
<td>Modelling patterns of time and emotion in Twitter #anticipointment</td>
<td>Sander Leemans (TUe)</td>
<td>2017</td>
<td>(TUe)</td>
</tr>
<tr>
<td>Robust Process Mining with Guaranteans</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year</td>
<td>Title</td>
<td>Authors</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>----------------------------------------------------------------------</td>
<td>--------------------------------</td>
<td></td>
</tr>
<tr>
<td>2017-13</td>
<td>Social Touch Technology - Extending the reach of social touch through haptic technology</td>
<td>Gijs Huisman (UT)</td>
<td></td>
</tr>
<tr>
<td>2017-14</td>
<td>You Are Who You Play You Are: Modelling Player Traits from Video Game Behavior</td>
<td>Shoshannah Tekofsky (UvT)</td>
<td></td>
</tr>
<tr>
<td>2017-15</td>
<td>Memory-Based Text Correction</td>
<td>Peter Berck (RUN)</td>
<td></td>
</tr>
<tr>
<td>2017-16</td>
<td>Understanding and Modeling Users of Modern Search Engines</td>
<td>Aleksandr Chuklin (UvA)</td>
<td></td>
</tr>
<tr>
<td>2017-17</td>
<td>Crowdsourced Online Dispute Resolution</td>
<td>Daniel Dimov (UL)</td>
<td></td>
</tr>
<tr>
<td>2017-18</td>
<td>Entity Associations for Search</td>
<td>Ridho Reindanda (UvA)</td>
<td></td>
</tr>
<tr>
<td>2017-19</td>
<td>Proximity of Terms, Texts and Semantic Vectors in Information Retrieval</td>
<td>Jeroen Vuurens (TUD)</td>
<td></td>
</tr>
<tr>
<td>2017-20</td>
<td>Fostering Engagement in Knowledge Sharing: The Role of Perceived Benefits, Costs and Visibility</td>
<td>Mohammadbashir Sedighi (TUD)</td>
<td></td>
</tr>
<tr>
<td>2017-21</td>
<td>Meta Matters in Interactive Storytelling and Serious Gaming (A Play on Worlds)</td>
<td>Jeroen Linssen (UT)</td>
<td></td>
</tr>
<tr>
<td>2017-22</td>
<td>Logics for causal inference under uncertainty</td>
<td>Sara Magliacane (VU)</td>
<td></td>
</tr>
<tr>
<td>2017-23</td>
<td>Entities of Interest—Discovery in Digital Traces</td>
<td>David Graus (UvA)</td>
<td></td>
</tr>
<tr>
<td>2017-24</td>
<td>Use of Affordances for Efficient Robot Learning</td>
<td>Chang Wang (TUD)</td>
<td></td>
</tr>
<tr>
<td>2017-25</td>
<td>Knowledge Representation for Clinical Guidelines, with applications to Multimorbidity Analysis and Literature Search</td>
<td>Versuska Zamborlini (VU)</td>
<td></td>
</tr>
<tr>
<td>2017-26</td>
<td>Socially intelligent robots that understand and respond to human touch</td>
<td>Merel Jung (UT)</td>
<td></td>
</tr>
<tr>
<td>2017-27</td>
<td>Investigating Positioning and Gaze Behaviors of Social Robots: People’s Preferences, Perceptions and Behaviors</td>
<td>Michiel Joosse (UT)</td>
<td></td>
</tr>
<tr>
<td>2017-28</td>
<td>Architecture Practices for Complex Contexts</td>
<td>John Klein (VU)</td>
<td></td>
</tr>
<tr>
<td>2017-29</td>
<td>From IT-Business/Strategic Alignment to Performance: A Moderated Mediation Model of Social Innovation, and Enterprise Governance of IT</td>
<td>Adel Alhuraibi (UvT)</td>
<td></td>
</tr>
<tr>
<td>2017-30</td>
<td>The Power of Facial Expressions</td>
<td>Wilma Latuny (UvT)</td>
<td></td>
</tr>
<tr>
<td>2017-31</td>
<td>Advances in computational methods for QFT calculations</td>
<td>Ben Ruijl (UL)</td>
<td></td>
</tr>
<tr>
<td>2017-32</td>
<td>Access to and Retrievability of Content in Web Archives</td>
<td>Thaeer Samar (RUN)</td>
<td></td>
</tr>
<tr>
<td>2017-33</td>
<td>Towards a Design Rationale for Software Documentation: A Model of Computer-Mediated Activity</td>
<td>Brigit van Loggem (OU)</td>
<td></td>
</tr>
<tr>
<td>2017-34</td>
<td>The Evaluation Framework for Learning Analytics</td>
<td>Martine de Vos (VU)</td>
<td></td>
</tr>
<tr>
<td>2017-35</td>
<td>Interpreting natural science spreadsheets</td>
<td>Yuanhao Guo (UL)</td>
<td></td>
</tr>
<tr>
<td>2017-36</td>
<td>Shape Analysis for Phenotype Characterisation from High-throughput Imaging</td>
<td>Yuanhao Guo (UL)</td>
<td></td>
</tr>
<tr>
<td>2017-37</td>
<td>WiBAF: A Within Browser Adaptation Framework that Enables Control over Privacy</td>
<td>Alejandro Montes García (TUe)</td>
<td></td>
</tr>
<tr>
<td>2017-38</td>
<td>Normative Social Applications</td>
<td>Alex Kayal (TUD)</td>
<td></td>
</tr>
<tr>
<td>2017-39</td>
<td>Exploiting properties of the human auditory system and compressive sensing methods to increase noise robustness in ASR</td>
<td>Sara Ahmadi (RUN)</td>
<td></td>
</tr>
<tr>
<td>2017-40</td>
<td>Steer your Mind: Computational Exploration of Human Control in Relation to Emotions, Desires and Social Support for Applications in Human-Aware Support Systems</td>
<td>Altaf Hussain Abro (VU)</td>
<td></td>
</tr>
<tr>
<td>2017-41</td>
<td>Minding a Healthy Lifestyle: An Exploration of Mental Processes and a Smart Environment to Provide Support for a Healthy Lifestyle</td>
<td>Adnan Manzoors (VU)</td>
<td></td>
</tr>
<tr>
<td>2017-42</td>
<td>Causal discovery from mixed and missing data with applications on ADHD datasets</td>
<td>Elena Sokolova (RUN)</td>
<td></td>
</tr>
<tr>
<td>2017-43</td>
<td>Semantic Mapping in Video Retrieval</td>
<td>Maaik de Boer (RUN)</td>
<td></td>
</tr>
<tr>
<td>2017-44</td>
<td>Understanding User Stories - Computational Linguistics in Agile Requirements Engineering</td>
<td>Garm Lucassen (UU)</td>
<td></td>
</tr>
<tr>
<td>2017-45</td>
<td>Decentralized Runtime Norm Enforcement</td>
<td>Han van der Aa (VUA)</td>
<td></td>
</tr>
<tr>
<td>2017-46</td>
<td>Sensor-based Learning Support</td>
<td>Jan Schneider (OU)</td>
<td></td>
</tr>
<tr>
<td>2017-47</td>
<td>Crowd Knowledge Creation Acceleration</td>
<td>Yie Yang (TUD)</td>
<td></td>
</tr>
<tr>
<td>2017-48</td>
<td>Collaborative inquiry-based learning</td>
<td>Angel Suarez (OU)</td>
<td></td>
</tr>
<tr>
<td>2018-01</td>
<td>Comparing and Aligning Process Representations</td>
<td>Han van der Aa (VUA)</td>
<td></td>
</tr>
<tr>
<td>2018-02</td>
<td>Multi-perspective Process Mining</td>
<td>Felix Mannhardt (TUE)</td>
<td></td>
</tr>
<tr>
<td>2018-03</td>
<td>Causal Models For Well-Being: Knowledge Modeling, Model-Driven Development of Context-Aware Applications, and Behavior Prediction</td>
<td>Steven Bosens (UT)</td>
<td></td>
</tr>
<tr>
<td>2018-04</td>
<td>Flexible Coordination Support for Diagnosis Teams in Data-Centric Engineering Tasks</td>
<td>Jordan Janeiro (TUD)</td>
<td></td>
</tr>
<tr>
<td>2018-05</td>
<td>Supporting the Complex Dynamics of the Information Seeking Process</td>
<td>Hugo Huurdeman (UvA)</td>
<td></td>
</tr>
<tr>
<td>2018-06</td>
<td>Model-Driven Information Security Risk Assessment of Socio-Technical Systems</td>
<td>Dan Ionita (UT)</td>
<td></td>
</tr>
<tr>
<td>2018-07</td>
<td>A formal account of opportunism in multi-agent systems</td>
<td>Jieting Luo (UU)</td>
<td></td>
</tr>
<tr>
<td>2018-08</td>
<td>Advances in Model Learning for Software Systems</td>
<td>Rick Smeters (RUN)</td>
<td></td>
</tr>
<tr>
<td>2018-09</td>
<td>Data Assimilation in Discrete Event Simulations</td>
<td>Xu Xie (TUD)</td>
<td></td>
</tr>
<tr>
<td>2018-10</td>
<td>Moving forward: supporting physical activity behavior change through intelligent technology</td>
<td>Julia S. Mollee (VU)</td>
<td></td>
</tr>
</tbody>
</table>