Theory-Based and Tailor-Made:

Motivational Messages for Behavior Change Technology

Roelof de Vries

Theory-Based and Tailor-Made:

Motivational Messages for
Behavior Change Technology

Roelof Anne Jelle de Vries

Ph.D. Dissertation Committee:

Chairman and Secretary:	
Prof. dr. J.N. Kok	University of Twente
Supervisor:	
Prof. dr. V. Evers	University of Twente
Co-Supervisor:	
Dr. K.P. Truong	University of Twente
Members:	
Dr. C.H.C. Drossaert	University of Twente
Prof. dr. D.K.J. Heylen	University of Twente
Prof. dr. M.C. Kaptein	Tilburg University
Prof. dr. M.A. Neerincx	TU Delft
Prof. dr. H. Oinas-Kukkonen	University of Oulu

Paranymphs:
K. Koopman
B.R. Schamhart

The research reported in this dissertation was carried out at the Human Media Interaction group of the University of Twente.

DSI Ph.D. Thesis Series ISSN: 2589-7721, No. 18-018
Digital Society Institute
P.O. Box 217, 7500 AE Enschede, The Netherlands

The research reported in this dissertation was supported by the Dutch national program COMMIT/.

SIKS Dissertation Series No. 2018-26
The research reported in this dissertation was carried out under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems.

© 2018 Roelof de Vries, Enschede, the Netherlands
Cover design by Koen Koopman and cover doodle by Cristina Zaga.
Typeset with LaTeX. Printed by Ipskamp

ISBN: 978-90-365-4649-2
DOI: 10.3990/1.9789036546492

All rights reserved. No part of this work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without prior permission from the copyright owner.

THEORY-BASED AND TAILOR-MADE:
MOTIVATIONAL MESSAGES FOR
BEHAVIOR CHANGE TECHNOLOGY

DISSERTATION

to obtain
the degree of doctor at the University of Twente,
on the authority of the rector magnificus
Prof.dr. T.T.M. Palstra
on account of the decision of the graduation committee,
to be publicly defended
on Wednesday November 14, 2018, at 16:45.

by

Roelof Anne Jelle de Vries
Born May 21, 1985
in Amsterdam, the Netherlands

This dissertation has been approved by:
Supervisor: prof. dr. V. Evers
Co-Supervisor: dr. K.P. Truong

Developing technology that effectively supports long-term behavior change is a challenge. This dissertation investigates how we can motivate people to inherently change their physical activity behavior through theory-based and tailor-made interventions in the form of motivational text messages.

The cover, designed by Koen Koopman, reflects the importance of text messages for this dissertation, as motivational text messages form the basis of this illustration. The cover also reflects one of the more surprising findings of this dissertation. Motivational messages designed by experts (back cover), we find, are perceived as more motivating by people in the earlier stages of physical activity behavior change, while motivational messages designed by peers (front cover) are perceived as more motivating by people in the later stages of physical activity behavior change. The five stages of change are represented on the bottom of the cover through doodles designed by Cristina Zaga. The doodles are inspired by the animated series La Linea from the Italian cartoonist Osvaldo Cavandoli.

Summary

Changing behavior is an intricate and difficult process, which requires people to make a conscious decision to change, and stick with it for a long period of time. To really be able to change we need awareness and motivation. Luckily, most people now own technology that can be used for communication, and therefore also for support and motivation, if we can harness the technology to our advantage. Using technologies to motivate people to change their behaviors is increasingly explored. However, developing motivational technology that effectively supports long-term behavior change is a challenge. Solutions offered in the field are: (1) basing motivational strategies on existing behavior change theory and (2) tailoring the strategies to characteristics of the user. But how can we operationalize these theory-based strategies so that we can use them in motivational technology? And what characteristics should one tailor these strategies to?

In this dissertation, we investigate how we can motivate people to inherently change their physical activity behavior through theory-based and tailored interventions in the form of motivational text messages. This involves exploring, evaluating, and documenting ways in which we can operationalize theory-based strategies, researching and evaluating characteristics of the user that can be used to tailor these theory-based strategies to, and finally, testing and comparing the theory-based strategies in a field study.

Samenvatting

Je gedrag of gewoontes veranderen is een ingewikkeld en moeilijk proces, dat vraagt om een bewuste keuze van mensen om hun gedrag aan te passen en zich daar langdurig aan te houden. Om gedrag echt te kunnen veranderen hebben we bewustwording en motivatie nodig. Gelukkig zijn de meeste mensen tegenwoordig eigenaar van een stukje technologie dat kan gebruikt worden voor communicatie en daarmee ook voor ondersteuning en motivatie, mits we die technologie in ons voordeel inzetten. Er wordt meer en meer gekeken naar hoe technologie kan worden ingezet om mensen te motiveren voor het veranderen van hun gedrag. Technologie ontwikkelen die ondersteuning biedt voor lange termijn gedragsverandering is een uitdaging. Oplossingen vanuit het onderzoeksveld zijn: (1) strategieën voor gedragsverandering baseren op theorie over gedragsverandering en (2) strategieën aanpassen op karakteristieken van de gebruiker. Maar hoe kunnen we deze op theorie gebaseerde strategieën zo operationaliseren dat we ze kunnen inzetten in motivationele technologie? En op welke karakteristieken van de gebruikers moeten we de strategieën aanpassen?

In deze dissertatie onderzoeken we hoe we met het inzetten van op theorie gebaseerde en aangepaste interventies in de vorm van motivationele tekstberichten mensen kunnen motiveren om hun beweeggedrag langdurig te veranderen. Dit houdt in: het verkennen, evalueren, en vastleggen van manieren waarop we op theorie gebaseerde strategieën kunnen operationaliseren, als ook het onderzoeken en evalueren van karakteristieken van de gebruiker die gebruikt kunnen worden voor het aanpassen van de strategieën, en tot slot, het testen en vergelijken van de op theorie gebaseerde strategieën in een veldonderzoek.

Acknowledgements

> First iteration of acknowledgements:
> Nothing too verbose, I acknowledge those, and those who I was supposed to, I suppose... also those.

Like basically everyone's story, this story begins with my dad. However, contrary to what you think I might mean, this story began when my dad suggested that I approach my current supervisor to ask if she would supervise my master thesis. I approached her saying something along the lines of: *"I want to investigate the addictive effects of 3G versus 2G phones"*, and she said something like: *"Great! That sounds super interesting! But here is what you are actually gonna do. Also, come join our weekly research meetings and informal lunch meetings to meet everybody!"*
Fast-forward almost 9 years later, and here I am defending my PhD dissertation! (assuming you are reading this right now in the room where the defense is taking place.)
As with every PhD trajectory, there are people to acknowledge that have kept you (relatively) sane throughout the years.
First of all, I want to acknowledge my friends and the people – past and present – at Human Media Interaction (special nod to the secretaries). Your continued contribution to my sanity, big or small, is and was much appreciated.
Second, I want to acknowledge my unofficial supervisor, for helping me navigate the psychological side of my research and helping me find and supervise great students.
Third, I want to acknowledge said students, for contributing to crucial parts of my research.
Fourth, I want to acknowledge my daily supervisor, for helping me through this whole ordeal and for putting up with (amongst other things) my overly casual communication and tendencies to procrastinate on deadlines.
Fifth, I want to acknowledge my master and PhD supervisor, for whom my overly casual communication and tendencies to procrastinate on deadlines seemed tailor-made. But more importantly, without whom I would have never even considered starting a PhD and without whom I could not have finished it.
Sixth, I want to acknowledge my family; my parents, my stubborn brother and wife-for-all-intents-and-purposes, and, most importantly, their amazing offspring who is already burdened with the task of surpassing the family in level or number of degrees.
Most importantly, I want to acknowledge my super significant other, partner-in-science, and cuore del mio cuore. You are my sanity.

I hope I can defend this dissertation well, wish me luck! (assuming you are reading this right now in the room where the defense is taking place.)

Contents

1 | Introduction

This dissertation investigates how people can be motivated to inherently change their physical activity behavior through theory-based and tailored interventions in the form of motivational text messages delivered by technology. Section 1.1 explains what reasoning has lead to choosing theory-based and tailored interventions to motivate people to change their physical activity behavior. Section 1.2 presents the research questions that are addressed. Section 1.3 explains the approach taken to answer these research questions. Lastly, section 1.4 provides an outline for the following chapters.

1.1 The difficulty in changing people's behavior

Changing habits or behaviors is hard. Whether it is quitting smoking, changing your diet, or exercising more, changing a deep-seated behavior does not happen over night. Behavior change is a long, difficult road, paved with challenges. To really speak of 'changed behavior', you need to have been doing the new behavior (e.g., not smoking, maintaining a new diet or regular exercise) for at least six months [Prochaska and Velicer, 1997]. Even when there is the intention to change, lasting change is highly unlikely. For example, it is estimated that only 8% [Renfree et al., 2016] of the people that make a New Year's resolution to change their behavior actually follow through. But why is it important that we change certain behaviors? Because changing a behavior, like starting to exercise regularly, can have substantial health, but also medical and societal benefits [Blair et al., 1995; Blair and Brodney, 1999; Warburton et al., 2006].

People's physical activity needs are still determined by the genetic makeup of people from the end of the Paleolithic era (roughly 10.000 years ago), whose activities revolved mainly around the hunting and gathering of food. Translated to contemporary activities, this hunting and gathering would amount to about three to four hours of moderate to vigorous exercise (e.g., brisk walking) *a day* [Fiuza-Luces et al., 2013]. However, people's occupational lifestyle — as a result of several revolutions (agricultural, industrial, digital) — has become ever more sedentary, and many people cannot reach the currently recommended amount of two and a half hours of physical activity *a week* in their leisure time. This lack of physical activity leads to an increased chance for numerous adverse health effects. According to the World Health Organization, as described in their Global Recommendations on Physical Activity for Health: "physical inactivity is now identified as the fourth leading risk factor for global mortality" [WHO, 2010, p. 7]. Moreover, "It has been shown that participation in regular physical activity reduces the risk of coronary heart disease and stroke, diabetes, hypertension, colon cancer, breast cancer and depression. Additionally, physical activity is a key determinant of energy expenditure, and thus is fundamental to energy balance and weight control ...” [WHO, 2010, p. 10]. To add, the ACSM’s Guidelines for Exercise Testing and Prescription (9th edition) state that: “Evidence to support the inverse relationship between physical activity and premature mortality, CVD/CAD, hypertension, stroke, osteoporosis, Type 2 diabetes mellitus, metabolic syndrome, obesity, colon cancer, breast cancer, depression, functional health, falls, and cognitive function continues to accumulate ...” [Pescatello and American College of Sports Medicine, 2014, p. 9]. From these descriptions, it is clear that regular exercise for the general population would be beneficial in reducing or preventing a legion of diseases and health conditions. It is safe to say that motivating people to change their physical activity behavior and keeping people physically active would be beneficial to our general health. So how can people be motivated to change or maintain their behavior, in particular their physical activity behavior?

Changing behavior is an intricate and difficult process, which requires people to make a conscious decision to change and stick with it for a long period of time. Some people stuck in a behavior that might need changing (e.g., smoking or sedentary behavior) are stuck because they are unwilling to change, others may be stuck because they are uninformed about the consequences of their behavior. Others still, may have tried to change a number of times and have become demoralized about their ability to change [Prochaska and Velicer, 1997]. To really be able to change we need awareness and motivation. However, not everybody has the motivation or the belief that they can change. As a result, people often look for some form of support or motivation elsewhere. Luckily, most people now own technology that can be used for communication, and therefore also for support and motivation, if we can harness the technology to our advantage. Using technologies to motivate people to change their behaviors is increasingly explored, as is shown by a steady increase in behavior change related research from 2003 to 2012 [Hekler et al., 2013].

On many occasions when looking for a way to support or motivate change, people end up choosing the way that makes the behavior easiest to do [Fogg, 2009]. Many applications on the smartphone nowadays promise you just that: an easy or fun way to change your behavior. These apps often resort to gamification-like approaches to behavioral change that leverage external factors and people’s extrinsic motivation (i.e., doing something because of external causes or sources, like receiving money or punishment), such as the use of praise, attention, rewards, levels, leaderboards, achievements, and external incentives [Lister et al., 2014]. Instead of tackling inner, individual motivational triggers or intrinsic motivation (i.e., doing something because it is enjoyable, interesting, or you are naturally driven to do it), people lean towards apps that use external or extrinsic motivations. Mobile applications like FitBit[1], for example, make extensive use of extrinsic rewarding strategies. Every time we hit a set number of steps, Fitbit sends a cute achievement badge to praise the success of the user. The fitness guru Kayla Itsines[2], developed a guide app — which generated more revenue than any other fitness app in 2016 — that mostly focuses on women

[1]https://www.fitbit.com/nl/home
[2]https://en.wikipedia.org/wiki/Kayla_Itsines

losing weight through extrinsic motivation by leveraging social media attention (i.e., Instagram posts) and community praises to support the journey to a 'bikini body'. The mobile application sensation Pokémon GO[3], caught the attention of many people and makes extensive use of gamification to invite users to go outside and move through the city. Pokémon GO makes its users more active. Unfortunately, just for as long as the game is interesting.

Although the long-term effects of Pokémon GO are not yet completely clear, physical activity interventions that use elements from games generally have poor long-term adherence [LeBlanc and Chaput, 2017]. Moreover, it is clear that the increase in physical activity for players of Pokémon GO does not extend to physical activity behavior in general [Gabbiadini et al., 2018]. In other words, players of Pokémon GO are only engaging in physical activity because it is a requirement of the game, and as such this physical activity will not continue when the players quit the game (for whatever reason). What all these applications and resources (like FitBit, the Bikini Body Guide, and Pokémon GO) have in common is that, to large extent, they are focused on using elements from games, external rewards, or extrinsic motivation, and are therefore usually only effective for a short while, or only as long as the application is used. Moreover, offering rewards while there is intrinsic motivation, can actually lead to a decrease in intrinsic motivation [Gneezy et al., 2011], provoking an overjustification effect [Biddle and Mutrie, 2007] where the intrinsic motivation is 'crowded out'. This leads to a decrease in performing the (previously intrinsically motivated) behavior after the behavior is extrinsically rewarded and the reward is subsequently discontinued. In general, providing extrinsic motivation or incentives, especially for behaviors that should be changed for a longer period of time, should be done with care because they can easily backfire [Gneezy et al., 2011].

In the psychology literature at large and in the behavior change literature in particular, there are a number of concepts that are related to 'motivation', some of which we have already come across, such as extrinsic and intrinsic motivation. The interventions that we present in this dissertation have the goal to change people's attitude and intention towards their behavior long term, convince people that engaging in physical activity is worthwhile for themselves, and instill the feeling that they are able to do it. This is captured in the concept of self-efficacy (however, other concepts could also fit, such as perceived behavioral control and (internal) locus of control, the subtle differences between these concepts are beyond this dissertation, but see [Ajzen, 2002] for a treatise on when, and when not, these concepts might be considered the same). The concept of self-efficacy is the basis of many behavior change theories and models and is considered a precursor to actual behavior change. Inspired by Prochaska and Velicer [1997], the definition that we use for self-efficacy is this: Self-efficacy is the confidence people have that they can do, or not do, the behavior that they want to change, regardless of the circumstances. So when we state that we want to 'motivate people' in this dissertation, we are referring to motivating people where we hope that this leads to higher self-efficacy, and in turn this higher self-efficacy leads to regularly engaging in physical activity long term. So how can we motivate people to change or maintain their physical activity?

[3] `https://www.pokemongo.com/`

To truly help people change their behavior, we believe that efforts should be focused on changing the behavior intrinsically and for a longer period of time, not as a side effect of using an application, but as a result of internalizing the behavior by focusing on changing people's attitudes towards and perceptions of a behavior. To avoid the pitfalls of using external factors or extrinsic motivation erroneously (such as the overjustification effect) and because we believe that efforts should be focused on changing the attitudes towards a behavior for a longer period of time, we turn towards theories on how people change their behavior. Behavior change theories are attempts to understand, describe, and explain the concept of change, and how people change behavior, whether it is learning (e.g., incorporating structured physical activity into weekly routine) or unlearning (e.g., quitting smoking) a behavior. As defined by Michie et al. [2017, p. 502]: "Theories of behaviour change [...] summarise what is known about constructs in the process of change, attempt to explain and predict when, why and how behaviour (change) occurs or does not occur, in addition to proposing both mechanisms of action and moderators of change along various causal pathways." There are many theories on changing and influencing behavior. These range from more practical theories, such as Persuasive Design [Fogg, 2003] (influencing people's behavior through the design of technology) to less practical theories, such as the Social Ecological Model [Stokols, 1996] (explaining behavior through the social, institutional, and cultural contexts of people's relations with their environment). Some have their roots in therapy, such as Motivational Interviewing [Miller and Rollnick, 2002] (a method to stimulate intrinsic motivation and change through the conscious and disciplined use of specific communication principles and strategies) or in work-psychology, such as the Goal-Setting Theory [Locke and Latham, 2002] (a theory on change involving setting effective goals). The use of theory or models, has been advocated in designing strategies or interventions to change behavior (e.g., Michie et al. [2008]; Cole-Lewis and Kershaw [2010]) because theory and models can guide the evaluation of interventions, ground the design of strategies, and offer explanations when the results show that interventions work or do not work. Hence, a theoretical foundation will help in understanding and targeting determinants of behavior.

Therefore, we aim to develop behavior change theory-based interventions (which we will refer to as theory-based interventions in this dissertation). However, just applying or using behavior change theory or theory-based interventions will not work on everybody the same way. As explained by Hekler et al. [2013, p. 7]: "Most behavioral theories traditionally explain, at best, only 20-30% of the total variance in a given health behavior, particularly when the behavior is tested in an intervention [...] In other words, approximately 75% of the variance is not accounted for by behavioral theory and thus can be attributed to unmeasured and unknown factors." A successful approach that is used to increase the effectiveness of behavior change strategies, is to tailor the strategies to characteristics of the user [Noar et al., 2007]. Therefore, we also aim to tailor the theory-based interventions to relevant user characteristics. Tailoring involves optimizing the impact of the message, or in other words: "How can we create and deliver messages to the public that are relevant, interesting, informative, and ultimately have the greatest chance of being persuasive?" [Noar et al., 2007, p. 674]. Optimizing the impact of the message means that we have to go beyond a one-size-fits-all message. Tailoring as a term has been used for different forms of 'going beyond a one-size-fits-all message', such as personalization or targeted communication. Some define tailoring to be on the personalized but generic level (for example, as a motivational message not just "You are doing great!", but "Roelof, you are doing great!"). Kreuter et al. [2000] define tailoring as "any combination of strategies and information intended to reach one specific person, based on characteristics that are unique to that person, related to the outcome of interest, and derived from an individual assessment". Following this definition would mean that the tailored communication has to be unique on the individual level (every person receives unique messages). Inspired by the definition that Noar et al. [2007] use for targeted communication, the definition that we use for tailoring is this: Adjusting to some specific characteristic of interest, such as personality traits, gender or stage of change, which is expected to increase the effectiveness (however defined) of the strategy (for example, if you would know Roelof is still in the stage of change where he is inactive, it might not make sense to say "You are doing great!"). However, in what form should the strategies or interventions be communicated through technology? And, how do we design the content for the strategies or interventions in this form? Or in other words, how do we operationalize the strategies? Using tailored text-based messages (which we will refer to as text messages in this dissertation) in combination with a behavior change theory or model can be effective to enhance motivation to attend to and process health information [Rimer and Kreuter, 2006], and to influence someone's physical activity behavior [Mutsuddi and Connelly, 2012]. Unfortunately, studies describing the development of such technology do not yet explain in detail *how* the researchers designed the motivational messages used [Latimer et al., 2010]. Moreover, there is little guidance on how to apply theory to the design of intervention strategies [Michie et al., 2008]. There are no best practices available to construct these messages.

We investigate how we can motivate people to inherently change their physical activity behavior through theory-based and tailored interventions in the form of motivational text messages. This involves exploring, evaluating, and documenting ways in which we can operationalize theory-based strategies as motivational text messages, researching and evaluating characteristics of the user that can be used to tailor these theory-based strategies to, and finally, testing and comparing the theory-based interventions in a 3-month field study.

1.2 Research questions

The research presented in this dissertation aims to investigate how to truly help people in changing their behavior through technology. To truly help people change behavior, we believe that efforts should not only be focused on external factors that can influence people, but efforts should also be focused on changing people's attitudes towards and perceptions of a behavior over a longer period of time, and this in turn will lead to lasting inherent behavior change. The main overarching research question for this dissertation is therefore:

How can people be motivated to inherently change their physical activity behavior using technology?

Behavior change theories focus on explaining how people change and typically incorporate strategies that are intended to increase people's motivation or change people's attitudes towards a behavior over a longer period of time. We expect that efforts to motivate people to change that are built on theories of behavior change can lead to lasting change. Literature discussed in Chapters 2 and 3 shows that to motivate people to change their physical activity behavior using technology, using theory-based behavior change strategies can indeed be effective, because theory-based strategies are well-founded and focus on intrinsic change over a longer period of time. However, using technology to deliver these strategies or interventions poses a major challenge: how to translate theory-based motivational strategies into real-world interventions that can be delivered by technology. For instance, a specific modality for these strategies has to be decided (e.g., voice, text, vibrations, colors), and the strategies need to be designed in this modality. The first research question is therefore:

RQ1: How can theory-based strategies be translated into a real-world technology-based intervention?

Traditionally, theoretically-grounded behavior change strategies are designed by experts. However, the content (e.g. the text or voice messages) for these strategies is not readily available and unfortunately, as pointed out by Latimer et al. [2010], studies describing the development of technology using theory-based strategies do not yet explain in detail on what basis or through which method the researchers designed the motivational messages used. Another option is to have peers design the interventions. Peer-designed interventions can be more engaging and more relevant to the user than expert-designed interventions, as is shown by Coley et al. [2013]. We investigate whether the content for theory-based strategies can be created by peers by asking the 'crowd' (known as crowdsourcing: employing a large number of people to contribute to a specific task) to design motivational text messages that can be matched to the theory-based behavior change strategies. By using the crowd, a wide selection of messages is readily available that can serve as effective theory-based behavior change interventions. However, theory-based interventions will not influence everybody the same way. A successful approach that can used to optimize the impact of the interventions is to tailor the interventions to certain characteristics of the user [Noar et al., 2007]. Tailoring is the focus of the second research question:

RQ2: How does tailoring the intervention to individual differences influence people's motivation for physical activity?

Literature discussed in Chapters 2 and 3 shows that to increase the impact of interventions, tailoring interventions to individual differences can be an effective approach. To optimize the effectiveness of the intervention messages that represent the theory-based behavior change strategies, we investigate whether there are individual differences in how people evaluate the messages. These differences form the basis for tailoring, adjusting which strategies to use based on the stage of change, personality or gender of the person, hopefully increasing the impact of the strategy. Another factor that may play a role in optimizing the impact of the theory-based strategies as real-world technology-based interventions, is the person who designs the intervention. As mentioned earlier, theory-based behavior change strategies are traditionally designed by experts of the specific context. Peer-designed messages were initially collected because they can be more relevant to the user, however, are these peer-designed messages as motivating as expert-designed messages could be? That is what the third research question is about:

RQ3: To what extent does the expertise of the designer of the intervention's motivational content influence how motivating the intervention is perceived?

It is reasonable to assume that experts have more expertise on how people change behavior. However, to what extent does this expertise matter in how motivating the content of the intervention messages is perceived? This research question, together with the other research questions aim to contribute to the answer on how people can be motivated to inherently change their physical activity behavior using technology.

1.3 Approach

This dissertation focuses on the design and evaluation of theory-based and tailored strategies in the form of motivational text messages. In the literature review chapters, we argue for the use of theory-based and tailored strategies to motivate people to change their physical activity behavior. In the design part, ways in which theory-based strategies can be translated to interventions in the form of motivational text messages are explored. In the evaluation part, ways in which the intervention can be tailored to individual differences in order to increase the effectiveness of the intervention are explored. Furthermore, we also evaluate how the motivational text messages can be delivered by technology in an in-the-wild experiment where participants receive motivational text messages on their smartphone. Overall, because of the **interdisciplinary nature** of the work, methods from social sciences as well as computer science are used throughout this dissertation. In the approach of this dissertation, one aspect in particular stands out: the use of **crowdsourcing** in the design and operationalization of theory-based interventions.

To date, the content of interventions based on theory has hardly been discussed or evaluated extensively. However, this begs the question whether the content is really representative of the theory-based strategy it is supposed to represent. Furthermore, the inexplicitness of the method and the content used to shape the theory-based interventions complicates efforts to reproduce the interventions or the results. As a side-effect, there is also no certified way to translate theory-based strategies to practical interventions. We use the method of **crowdsourcing** in an innovative way: to translate theory-based strategies to real-world interventions. Crowdsourcing involves using a large number of people (i.e., a crowd) to help do a specific task. Usually, this involves a small task that is reasonably simple for a human, but harder for a computer (e.g., identifying if there is a car in the picture shown) or a task that is simple to do for a large group of people, but hard for one person (e.g., 500 people thinking of a few motivational text messages each compared to one person thinking of a few thousand motivational text messages). Specifically, we use crowdsourcing for something called macrotasking [Cheng et al., 2015], which refers to a slightly more complex, creative and time-intensive task that is done by the crowd. In other words, it is a qualitative task, done in a quantitative way. In our case, this is the design of motivational **text messages** for specific scenarios.

1.4 Outline

Chapter 1 introduces the context of this dissertation and presents the research questions. **Chapter 2**, presents more context with an overview of technology designed to motivate people to engage in physical activity and a reflection on how this informs the research of this dissertation. **Chapter 3** provides a short overview of research describing the factors and differences that influence how people change their physical activity behavior, presents research that harnesses these differences to more effectively motivate people in changing their behavior, and discusses theoretical background on how people change physical activity behavior. **Chapter 4**, presents the first study, which is set up to see if we can use crowdsourcing to operationalize theory-based behavior change strategies (RQ1). Moreover, we present the analysis of the designed messages through coding. In **Chapter 5**, we present the second study, which is carried out to find out how people evaluate the theory-based strategies based on their stage of change, personality and gender (RQ2). **Chapter 6**, presents the third study, set up to have experts operationalize theory-based behavior change strategies and the fourth study set up to find out how people evaluate these strategies so that we can compare the previously peer-designed messages to these new expert-designed messages (RQ3), also on a linguistic dimension. **Chapter 7**, presents the fifth and final study, set up with two conditions, and carried out to see if an intervention sending people messages from the strategies matching their stage of change proves to be more effective than an intervention sending random messages from the strategies (main overarching research question). We discuss the implications of the studies and reflect on the results in **Chapter 8**, and we end with concluding remarks and our perspective on future work in **Chapter 9**.

2 | Exploring motivational technology for physical activity

Although the goal of this dissertation is to motivate people to inherently change their physical activity behavior, this chapter first considers a bigger picture and provides a review of technology that is designed to motivate people to do more physical activity, without this necessarily leading to behavior change. Technology that is designed to motivate people to do more physical activity can be approached from different perspectives. In this review we address literature from several perspectives, but this is by no means an exhaustive list. Based on this review two approaches emerge that are worthwhile pursuing to effectively motivate people to change their behavior: using theory-based interventions, and tailoring those interventions to relevant user characteristics. These two approaches will be discussed in Chapter 3.

Although there are many ways to stimulate and promote physical activity through the use of technology, in this dissertation the focus is mainly on discerning the ways in which motivating people through technology could lead to lasting physical activity behavior change. Or in other words, how can we motivate people in a way that they might eventually change their behavior to incorporate physical activity using technology? As a first step in answering this question, we look at the bigger picture and review literature that is focused on motivating people to do more physical activity using technology. In this review we describe the different forms that technology can be presented in, like using applied interfaces, standalone devices or mobile applications, with or without additional sensors, but mostly which motivational techniques were used to encourage physical activity.

The goal of this review chapter is to provide an idea of what applications are out there, how technology has been used to motivate people to do more physical activity, and what motivational strategies can used. From this we determine what motivational strategies are worthwhile to pursue in changing behavior long-term, which is the goal of this dissertation.

As of late, there have been more and more reviews published on using technology for the promotion of well-being of the public (e.g., on managing obesity [Hermawati and Lawson, 2014] or on general health interventions for mobile phones [Klasnja and Pratt, 2012]), but to our knowledge, none of these have focused solely on promoting physical activity.

2.1 Overview of technology promoting physical activity in HCI

As mentioned, there are many different fields in which researchers are working on the promotion of physical activity, and we will discuss the technologies and applications under the umbrella of specific fields. In many cases, this does not mean that this technology or application fits exclusively in the chosen category; a technology or an application can use strategies or ideas from multiple fields. Moreover, the research carried out in these fields has overlap to some degree in that they all try to influence or motivate a behavior, in this case physical activity behavior. The difference is in how influencing or motivating a behavior is approached. For example, exertion interfaces require physical activity to use a technology, ubiquitous technology influences people's physical activity behavior by using the ubiquity of technology to track and motivate people, gamification technology implements elements from games to improve the engagement with technology that in turn might require or promote physical activity behavior, persuasive technology focuses on influencing people's attitudes or behaviors through persuasive design or persuasive strategies, and behavior change technology focuses on changing people's behavior by using theory-based behavior change strategies delivered by technology.

2.1.1 Exertion interfaces

An interesting development for technology driven encouragement of physical activity are exertion interfaces. These interfaces are physically demanding and deliberately require physical effort. Although encouraging healthy living is not necessarily a goal for exertion interface research, the result of using a exertion interface is essentially the same (more physical activity). Exertion interfaces are a relatively new development for HCI and therefore have not received that much attention, but exertion combined with games (exergaming) is already well underway (e.g. Wii Sports on the Nintendo Wii).

An important line of research from this perspective is carried out by Mueller and others [Mueller et al., 2003, 2007; O'Brien and Mueller, 2007; Mueller et al., 2008, 2009, 2010; Graether and Mueller, 2012] in which they promote the design of interactive physical games and sports (see [Mueller et al., 2008] for an explanation on the classification of these new games). Mueller et al. [2003] discusses Break Out, a traditional interface focused on physical effort at both ends, modeling sport-like functionality. The exertion interface, where players had to shoot balls at a screen in a penalty-like situation, encouraged social bonding; users played better, had more fun and became better friends compared to the traditional interface counterpart. Mueller et al. [2009] experimented with a similar exertion interface game called Table Tennis for Three (see Figure 2.1), where three people can remotely play a gamified version of table tennis against each other. They found that the physical version facilitated more social play than their button-pressing version, which is suggested to be an important factor in the successfulness of games.

In [Mueller et al., 2007; O'Brien and Mueller, 2007] the researchers found through a small survey amongst joggers that more than half of their respondents run in groups. The main reasons to run together were: socializing, motivation to run faster, more

Figure 2.1: The Table Tennis for Three interface. The table is set upright and divided in two, with a opponent on each half.

fun and encouragement to show up. Based on this, the researchers tested an exertion interface (audio system) to connect remote joggers, with positive preliminary results. Mueller et al. [2010] continued this 'Jogging over a Distance' (see Figure 2.2) research, where they connect the spatial audio to the preferred heart rate of both joggers. This way, joggers of different capabilities can run together on the same 'effort' and still enjoy all the benefits of social running. In [Graether and Mueller, 2012] they also experimented with a flying robot as a companion during running (see Figure 2.2), creating some interesting possibilities as a personal trainer.

Figure 2.2: On the left, someone running with gear in both Melbourne, Australia and London, UK. On the right, the joggobot in action.

Although not necessarily the goal of exertion interfaces, when used, the result is always more physical activity by the users. Exertion interfaces are not necessarily designed to leverage any kind of influential strategy. Nevertheless, well designed interfaces do offer the possibility for the user to experience more immersion (e.g., socially or through fun) than its interface-less counterpart. It is this 'richer' experience that keeps users more engaged or motivated to keep going. Exertion interfaces also target the activity directly and not the technology that could possibly assist it (like an application on a smartphone). It seems that applying interfaces to these physical activities enhances the connectedness and social play experienced by the users. However, the focus of exertion interfaces mostly seems to be on supporting or enriching exercise behaviors that people already have by making it easier to maintain these behaviors, not on people actually changing their behavior inherently or long-term.

2.1.2 Ubiquitous technology

Ubiquitous computing ("computing" anytime and everywhere) and encouraging physical activity are in a sense an ideal (futuristic) match, because ubiquitous technology would facilitate encouragement to be at the moment it is needed the most (for example, when someone is feeling lazy on the couch). Although the focus of ubiquitous computing is not always on the health benefits of physical activity, the increasing pervasiveness and subsequent ubiquity of technology, like mobile phones but also wireless networks and sensors, is ideal for personalized and context-aware support of physical activity.

Gil-Castiñeira et al. [2011] describe the social sports application RunWithUs (see Figure 2.3), a tool integrated into the Finnish "Ubiquitous Oulu" city. The tool consists of three components: a component to keep track of statistics about workouts, a social network and a marketing component. Strategies that are used to influence people to do more physical activity include personal awareness, social comparison, competition and cooperation, all readily available in the ubiquitous environment of Oulu city.

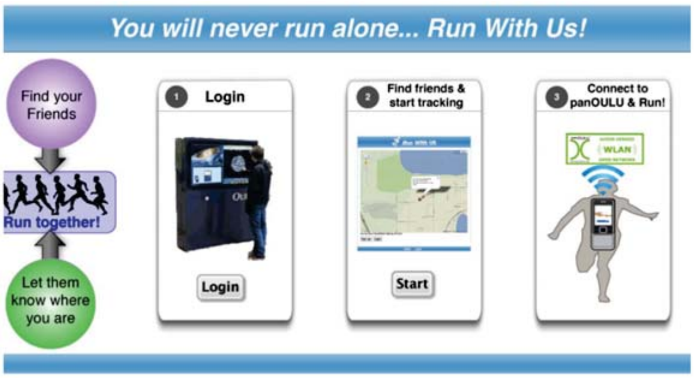

Figure 2.3: The environment of the ubiquitous Oulu city.

Feeding Yoshi by Bell et al. [2006] is a mobile multiplayer location-based team game where secured or open wireless networks (around a city) were used as pick up or drop off points in the game (see Figure 2.4). Although not discussed in their paper it is clear that many aspects of the game would be considered motivating, engaging or even addictive. As a short-term result of playing this game people increased their physical activity a lot (by walking around the city). In many aspects, this location-based game could be considered a precursor to Pokémon GO, players had to walk around the city, collect fruit, and possibly interact with other players through trading.

In iDetective from Kimura et al. [2011], users play a location-based game through the use of a mobile phone, GPS, compass and camera, and are challenged to find real life locations (see Figure 2.5). The system uses social comparison as motivational (persuasive) strategy and apply a strategy, namely goal-setting from a behavior

Figure 2.4: On the left: Nearby pick up points for fruit and people. Next: Interaction screen with another player. On the right: Fruit gamble screen.

change theory (i.e., Goal-Setting Theory [Locke and Latham, 2002]), to encourage people to set goals they want to achieve. In addition, the system uses a construct from another behavior change theory, the Transtheoretical Model's (TTM) [Prochaska and DiClemente, 1983] stages of change, for feedback decisions. The stages of change classify people in different stages of their behavior change (see also Table 3.1), making it possible to adapt feedback based on the stage of the user.

Figure 2.5: On the left: search mode. Next: Status screen. Next to that: The virtual agent feedback. On the right: The mission list.

The advancement of ubiquitous computing can, ideally, work very well to encourage physical activity. Any of the other perspectives can potentially benefit from the ubiquity of technology. Moreover, the ubiquity of technology makes it easier for users to be social, but also to be tracked, and be supported or motivated. However, the ubiquity of technology in and of itself does not seem enough to successfully support people in changing their behavior over a longer period of time, and additional strategies are usually applied.

2.1.3 Gamification technology

The previously discussed line of research by Mueller and others (section 2.1.1) focuses on facilitating (intense) physical activity through exertion interfaces. They do this through introducing interfaces to regular physical activities making them more interactive and location independent, such as their penalty shooting, table tennis and running with or against each other in geographically different locations. There is also the possibility to turn this around and introduce physical activities to already existing 'interfaces', such as introducing physical activity to an adapted real-time version of chess by Stanley et al. [2008]. The chess game increased physical activity levels and the users found it engaging and fun. Or one can design both a new activity and game, as presented by Berkovsky et al. [2010]. They present the game design of PLAY, MATE! and the public application of this game design called Neverball, a serious game to increase physical activity through leveraging a user's game motivation. Neverball is a time and goal-based navigation game, in which players collect sufficient coins in a limited period of time. In short, measured physical activity is converted to game time to motivate users to jump or step on the spot. Among other things, tailored goal setting and personalized rewards were used.

Fujiki et al. [2008] discuss the NEAT-o-Games (see Figure 2.6) where players' physical activity was logged, used, and reflected in a virtual race game based on four design principles: simple, informative, discreet and motivating. Winners were declared daily and collected more activity points. Feedback was given on a PDA were an avatar reflected satisfaction with the activity performed.

Figure 2.6: The four design principles in action: simple, informative, discreet and motivating

Campbell et al. [2008] analyzed game design principles and applied them to an everyday fitness game (see Figure 2.7). From prior work and literature, a set of game design principles were distilled which, they argue, should be carefully designed for to develop effective gamification technology: core game mechanic (the set of essential interactions a player repeats during play, usually easy to learn, difficult to master), representation (aesthetics and narrative), micro goals (to advance through the game, also macro goals), marginal challenge (no unachievable challenges, no easy challenges), free play (incorporate choices during play, mimicking a certain freedom of play), social play (incorporate in game design), and fair play (fairness of rules and equal opportunity to win).

Figure 2.7: On the left: the game environment of Kukini. On the right: social functions in Kukini

The research described all garnered positive preliminary feedback, and it seems that games and gamification principles can have the potential to motivate and sometimes even change behaviors. Moreover, gamification is an interesting principle to combine with a 'positive' behavior like structured physical activity, but when applied, it can be hard to untangle whether this 'positive' behavior actually gets ingrained or if it is just the game or game mechanics that are 'addictive' enough to leverage this positive behavior.

2.1.4 Persuasive technology

Persuasive technology is focused on influencing attitudes or behaviors. In a sense, every research or technology described has something to do with influencing attitudes or behaviors, and could therefore be considered persuasive technology. The concept of persuasion is based on psychological and sociological theories on behavior and technology.

B.J. Fogg in his seminal work [Fogg, 2003], discerns 42 strategies for persuasion. But ongoing work has culminated in 160+ [Rhoads, 2007] different strategies. How many strategies you can discern depends on how narrow and practical strategies are defined. For example, Cialdini in *his* seminal work [Cialdini, 2001] proposes (only) 6 strategies: (1) reciprocity, the obligation you feel to repay someone when they do something for you; (2) commitment and consistency, the urge you feel to be consistent with what you already agreed (or disagreed) with; (3) social proof, the safety that is in doing what others are also doing; (4) authority, the obligation we feel towards (ostensible) authority; (5) liking, the tendency to be more quickly convinced or persuaded by someone we like; (6) scarcity, the desire to get things which are 'special' or limited or running out. The strategies of Fogg and Cialdini are the inspiration for many of the strategies used in persuasive technology research.

Flowie from Albaina et al. [2009] (see Figure 2.8), is a prototype glanceable health coach application focused on getting elderly to walk. The application focuses on leveraging strategies from Fogg's persuasive technology theory to encourage walking. A pedometer was used to collect data and an in-house touchscreen with a virtual flower was used to display emotions modeling the progress. After a user panel, the following four strategies were used: goal-setting (having a challenging but realistic goal, set by an ostensible professional), self-monitoring (creating awareness of progress, through a touch screen monitor), consistency (the need to deliver on the promise of the goal, through goal-setting and self-monitoring), and intrinsic motivation (leveraged through empathizing with the flower and the other strategies, also based on classic learning theory). A preliminary user study concluded that the system was appreciated and leveraged motivation, but no real conclusions about increased physical activity could be drawn.

Figure 2.8: On the left: the Flowie application displayed on a tablet. On the right: the three views of the application: general overview, day overview and week overview

Toscos et al. [2006] discuss Chick Clique, a health application aimed at influencing teenage girls by their social desire to stay connected (social validation), also using the tools of goal setting, self-monitoring, positive reinforcement and social support. An interesting finding was that the addition of a shared 'group performance' (seeing the progress of everyone in the group) in the design was found a powerful method of changing behavior in a post-study questionnaire indicating that social support and social validation are powerful strategies to influence behavior.

Sohn and Lee [2007] present Up Health, an Instant Messaging (IM) system designed to explore the potential of IM as a tool to persuade or encourage users to change. During a well received preliminary study they implemented four persuasive techniques: personal awareness, social cooperation/competition/comparison, fun and enjoyable interaction (to leverage long-term engagement), and unobtrusive and intuitive notifications.

From this section it is clear that there are plenty of strategies and possibilities to influence a user's activity pattern. What also becomes clear however, is that due to the abundance of strategies, most strategies lack guidance on how to use and implement them. Moreover, due to the lack of a framework or model, it is unclear what results to expect when using these strategies. Recent research by Oinas-Kukkonen and Harjumaa [2008] and Oinas-Kukkonen [2010] is mitigating this problem by providing guidance for the use and implementation of persuasive strategies and by providing a model to use and interpret these strategies in. It is important for a scientific design to not overdo the design with too much strategies which are hard to measure or separate. Because even if the design would result in an application which leads to successful (long-term) change, it could be unclear why exactly this change happened or which strategies proved crucial in this success. In the next section we will look at technology borrowing constructs from behavior change theories. Behavior change theories provide an idea or a framework on how the use of these constructs should affect people.

2.1.5 Behavior change technology

Behavior change technology focuses on changing people's behavior by using theory-based behavior change strategies delivered by technology. Considering this definition, behavior change technology is closely related to persuasive technology. However, behavior change technology is designed to not only influence attitudes or behaviors, but to change attitudes or behaviors long-term. An important line of behavior change technology research in HCI is carried out by Consolvo and colleagues [Consolvo et al., 2006, 2008a,b, 2009b,a,c; Klasnja et al., 2009, 2011; Munson and Consolvo, 2012].

In [Consolvo et al., 2006] Houston is discussed, a prototype journal-sharing application to encourage physical activity focusing on step count, where a sharing journal plus goals and progress report version versus a non-sharing version were compared and it was found that sharing was a successful strategy to increase goal achievement. The stages of change from the TTM were used to assess which stages of behavior change people were in. The application focused on people in the contemplation (thinking about changing), preparation (preparing to change), action (taking the first steps to change) and maintenance (maintaining the changed behavior) stages. Also, four design requirements were identified for technologies like this: give users proper credit for activities, provide personal awareness of the activity level, support social influence, and consider the constraints of users' lifestyles.

In [Consolvo et al., 2008a,b] some of these requirements are followed up and the UbiFit Garden is discussed (see Figure 2.9), which provides easy personal awareness of activity levels, support of multiple activities with a glanceable display, a non-literal representation of physical activity and goal attainment to motivate behavior change (targeting the stages contemplation, preparation and action). For the non-literal representation of physical activity and goal attainment, flowers and butterflies in a garden are used to represent activities and achievements. A 'normal' interactive application is incorporated to see registered performance, to keep a journal, and correct activities. This version was tested versus a non-glanceable display (without abstract "wallpaper" garden representation, but with interactive application) and it was found that the glanceable version was more encouraging than the counterpart without glanceable display.

Figure 2.9: The glanceable UbiFit Garden on a mobile phone with flowers for activities, butterflies for goal attainment and large butterflies for weekly goals

Based on the experiments with the UbiFit Garden and several behavioral and social psychological theories (Goal-Setting Theory, TTM, Presentation of Self in Everyday Life and Cognitive Dissonance Theory) Consolvo et al. [2009c] present eight (not mutually exclusive) design strategies for technology used in everyday lifestyle changing. They suggest that applications should be designed such that they are: (1) abstract & reflective (use data abstractions to help the user reflect on goals and achievements); (2) unobtrusive (avoid unnecessary interruptive messages); (3) public (the technology used should not make people uncomfortable or feel ashamed); (4) aesthetic (the technology used must be comfortable and attractive); (5) positive (use positive reinforcement, avoid negative reinforcement); (6) controllable (permit the user to manipulate the data); (7) trending / historical (provide relevant historical data); and (8) comprehensive (account for a range of behaviors). The UbiFit Garden implements them as follows: (1, 2, 3 and 4) an abstract background animated garden on a mobile phone as a metaphor to represent physical activity and goal attainment; (5) rewards (addition of flowers and butterflies) to encourage behavior and no negative consequences if goals are not met; (6) the data is editable; (7) the garden with butterflies gives weekly and monthly reflections; (8) the platform used can infer multiple types of physical activity (e.g. walking, running, cycling).

Also based on research with the UbiFit system and additional interviews, Consolvo et al. [2009a] explore the importance of goal-setting (based on the Goal-Setting Theory) divided into two aspects, namely goal sources (who sets the goal) and goal time frames (what period of time is set for the goal). It is found that most participants would like to either set their own goals or work with a fitness expert (but preferably no medical guidelines or medical advisers) and to have a time frame of a week beginning on Monday or Sunday (compared to a rolling seven-day window, moving forward one day at a time) and thus resetting at the end of the week (which would be either on Sunday or Saturday).

Klasnja et al. [2009] discuss four lessons learned from both Houston and the UbiFit Garden with respect to designing for behavior change. These are: support the persistent activation of health goals (the glanceable display as a persistent goal reminder proved to be beneficial), encourage an extensive range of healthy behaviors (if the application focus is increasing physical activity, provide ways to add unregistered physical activity to prevent frustration and incongruity), focus on long-term patterns of activity (help users reflect on long-term activity and also design for periods of possible inactivity), and facilitate but not depend on social support (not everyone is driven by social comparison and competition).

Lin et al. [2006] developed Fish'n'Steps (see Figure 2.10), an engaging application for the computer where a user's physical activity is linked to the growth and emotional state of a virtual fish in a fish tank. Important motivational tools were cultivation of a strong internal locus of control (a concept similar to self-efficacy, which are both important conceps in many behavior change theories) through pet care and the incorporation of social influences. The strategies of cooperation versus competition were compared but no significant differences were found. Also, the use of negative reinforcement through the fish when the physical activity was disappointing had mixed results. Behavior change was measured in terms of the increase of steps and the stages of change from the Transtheoretical Model and most participants showed continued increase in activity levels in either or both.

Another example is Shakra [Maitland et al., 2006; Anderson et al., 2007], a non-intrusive tracking and activity sharing application using a standard mobile phone and the fluctuation in signal strength to estimate activity through the use of machine

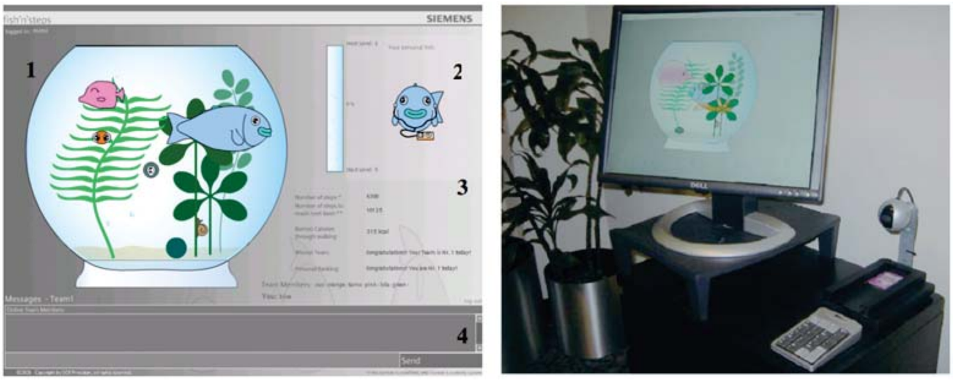

Figure 2.10: On the left: the personal display of the user. On the right: The fish tank the user is in with other fish (teammates). This fish tank was compared to other fish tanks (competition)

learning. Shakra focused on people not currently achieving minimum recommended daily activity level. Shakra employed strategies adapted from Transtheoretical Model and Social Cognitive Model such as self-monitoring, sharing and comparison, which gathered positively preliminary responses.

Although this section addressed technology that used strategies from theories of behavior change, no implementations of a complete theory have been undertaken in these technologies. However, using and implementing a theory in its entirety can be beneficial because this theory can then provide a framework in which to interpret and explain results. If only part of a theory is used, these interpretations and explanations might not be valid. In the last section we will discuss our findings and decide on practices worthwhile for our own research.

2.2 Discussion and conclusion

Through reviewing these technologies that all in some way or another try to motivate people to do more physical activity, we can compile of list of best practices and recommendations. These recommendations do not necessarily follow from demonstrable effectiveness. To really prove a strategy applied in technology is effective (long-term) is usually also not feasible, possible or even preferable for technology promoting physical activity in HCI [Klasnja et al., 2011]. Therefore, the recommendations compiled include future work and suggestions discussed in the reviewed materials.

Incorporating a social strategy (sharing, bonding, play, support, cooperation, competition, comparison, network) seems to be a very valuable motivational technique, whether this is explicit, for example, explicitly cooperating or competing as part of the design, game or leader board, or if it is implicit, because a social function is just incorporated in the game or design and competition or cooperation emerges. Another recommendation could be to include a goal setting option in your technology. Design should try to incorporate the possibility for users to define a lot of the characteristics of the goals they want to set. Other recommendations for the design of an application encouraging physical activity could be: (1) self-monitoring needs to be possible. This is best combined with creating personal awareness, a glanceable design, goal-setting and progress tracking; (2) a form of rewarding and/or achievements, preferably for short-term rewards as well as long-term achievements; (3) the incorporation of unobtrusive reminders; (4) information on long-term statistics; (5) the possibility to take a break from the application (e.g. for holidays); (6) proper privacy handling; (7) an adaptive application for when users progress in their activities (short-term and long-term); (8) an engaging, aesthetically pleasing and appealing design (or at least not unappealing); and (9) overall positive framing of displayed information and the use of positive (and in some situations negative) reinforcement. Based on these recommendations it should be possible to develop motivational technology. However, when reviewing these studies certain other things also stand out. First, although a lot of strategies can be discerned and have proven to be successful for short-term compliance, not many of the strategies have been tested for long-term adherence. This begs the question of how these strategies will affect long-term use of an application. Second, a lot of motivational and persuasive strategies are design suggestions. This makes it harder to get an idea of whether there is any intrinsic change, or whether this effect is to tied to the design of strategy. Third, not many of these discussed (prototype) applications have actually been followed through to a full-fledged application. Although this is a common practice of HCI and design research, this adds to the uncertainty of the effects of the strategies in 'real life' and long-term use. Fourth, although some papers discuss psychological motivation and behavioral change theories like the Goal-Setting Theory or the Transtheoretical Model they seem to use only some concepts of the theories, such as only the stages of changes from the Transtheoretical Model. This is understandable, because most of the theories are not easily implementable, but this does complicate the interpretation of the results. Fifth, practically all the discussed research aims to encourage physical activity or even measure and accomplish behavior change, but (almost) none of them seem to focus on people not willing to exercise, instead they focus on people who are already exercising, or who are rewarded or obligated through the experiment to exercise, which seems like the group who needs it the least. This is also understandable, because people who are not willing to change are probably also not willing to participate. But finding results for increased activity from people already motivated (or just paid enough) to change does not necessarily mean that these strategies will also work on people less (or not) motivated to change. Lastly, although some papers mention personalization or tailoring and tailored feedback, actual implementation is still very hard and requires more research and more knowledge on a user. There is still a lot of room for improvement in this direction.

When considering these points and the mixed results achieved in the research discussed, it seems beneficial to dive deeper into how people change their behavior (also those unwilling), how to design for behavior change, and how to tailor our potential strategies to the user by accounting for the individual differences in people. Which theory or theory-based strategies should be used? What characteristics should one tailor theory-based motivational strategies to? In the next chapter we aim to answer these questions for our goal of motivating people to inherently change their physical activity behavior through theory-based and tailored interventions in the form of motivational text messages delivered by technology.

3 Exploring determinants, strategies, and theories to motivate physical activity behavior change

Based on the review of technology that is designed to motivate people to do more physical activity presented in Chapter 2, two directions emerged that are worthwhile to explore further: using motivational strategies based on a theoretical background, and accounting for the individual differences in people to tailor the strategies to their liking. This chapter provides an overview of research describing factors and differences that influence how people change their physical activity behavior, and that could potentially be used for tailoring. Furthermore, this chapter provides theoretical background on how people change physical activity behavior and explores how a combination of theory and tailoring could work.

There is a substantial amount of work on identifying determinants of physical activity. In section 3.1 we review the most important determinants of physical activity. In section 3.2 we look at work on tailoring to individual differences. Section 3.3 discusses behavior change theory and specifically the model we chose to use for this dissertation, the Transtheoretical Model (TTM), which, we argue, is well suited for tailoring. In section 3.4 we conclude the chapter and discuss the next step of getting behavior change strategies operationalized.

3.1 Determinants of physical activity behavior and behavior change

Dishman and Sallis [1994] discuss the (then) state of the art scientific literature on known determinants of physical activity. They also discuss requirements for effective exercise interventions which will require: "... that both abstract (for example, beliefs) and concrete (for example, disability) determinants be addressed in complementary ways to (a) diminish or compensate for psychological and physical or environmental barriers to activity; (b) provide knowledge, skills, and reinforcements that augment the willingness and ability to be active; and (c) permit selection of appropriate forms and intensities of activity." This still seems appropriate today. Furthermore, Dishman and Sallis [1994] review determinants of physical activity which are compiled into a list of 38 determinants, divided into personal (e.g., age and education), environmental and activity characteristics. Sherwood and Jeffery [2000] also review the state of the art regarding the determinants of exercise behavior. They classify the determinants into two categories, namely personal characteristics (e.g., stage of change and motivation) and environmental characteristics. Bauman et al. [2002] discuss the factors related to physical activity in a more fine-grained manner. They distinguish between determinants, correlates, mediators, moderators and confounder variables. These variables differ in their statistical relation to physical activity. However, the exact statistical implication of these variables is beyond this dissertation. They also classify characteristics in different categories: demographic and biological factors (e.g., age, education, gender (male only) and socioeconomic status), psychological, cognitive and emotional factors (e.g., personality), behavioral attributes and skills, social and cultural factors, physical environment factors and physical activity characteristics. Overall, it is interesting to see that more factors are distinguished. Bauman et al. [2012] discuss correlates of physical activity. The trend to discuss factors involved in physical activity has since been more focused on the ecological models, i.e. taking a broad view on health behavior and also considering as factors, for example, urban planning and transportation systems. Their classification is: demographic and biological variables (e.g., age, education, gender (male) and socioeconomic status), psychosocial variables (e.g., stage of change), behavioral variables, social and cultural variables and environmental variables. Specifically for the purpose of this dissertation, it is worthwhile to note that in these reviews the stage of change, personality and gender are mentioned as factors associated with overall physical activity, because we intend to use these for tailoring our theory-based strategies to.

An interesting factor that influences physical activity is personality. Personality is a way to describe long-lasting individual characteristics (similarities and differences) between people. In the psychology literature, a lot of different personality classifications can be found. The most important one is the Big Five model [Goldberg, 1992], also known by its acronym for the personality traits; OCEAN. This framework classifies people in five dimensions: Openness to experience, (**O**), Conscientiousness (**C**), Extraversion (**E**), Agreeableness (**A**) and Neuroticism (**N**). While personality has been found to be a important determinant in behavior [Ajzen, 2005], it is not widely adopted to tailor interventions for physical activity to; possibly because most effects found were small to medium correlations [Rhodes and Smith, 2006]. Nevertheless, we are interested in personality as a characteristic to tailor interventions to because it is an important determinant and personality has been found to be, arguably, a relatively stable personal characteristic in adults [Rhodes et al., 2004]. Assuming personality is a stable characteristic and assuming personality varies across different people, we could tailor motivational strategies to a person's personality profile. Because personality usually remains constant over longer periods of time, this would allow for supporting longer-term tailoring to. In the context of changing or determining physical activity behavior, some direct relations have been found between physical activity and certain personality traits. Overall, Extraversion and Conscientiousness were found to be positively correlated with physical activity and exercise behavior [Gallagher et al., 2013; Hoyt et al., 2009; Rhodes and Smith, 2006] while Neuroticism appears to be negatively correlated to physical activity [Rhodes and Smith, 2006].

Courneya and Hellsten [1998] showed that exercise behavior, motives, barriers, and preferences were correlated to the personality traits of the Big Five. For instance, it was found that Openness was positively correlated to health and stress relief, Extraversion was positively correlated to exercising with other people, Extraversion and Openness were negatively correlated to supervised exercise, and Agreeableness was negatively correlated to competitiveness. These correlations suggest that people with different personalities have different considerations when it comes to exercising. This is consistent with Ingledew and Markland [2008], where they found that different personalities have different motivations for change in exercise participation, for instance Openness positively correlated with a health/fitness motivation. These studies indicate that people with different personalities should be motivated in different ways to participate in exercise and affirm the idea of personality tailored interventions for physical activity (e.g., Rhodes et al. [2004]; Rhodes and Smith [2006]).

3.2 Tailoring to individual differences

There has been some recent research into the role of personality when designing tailored persuasive strategies. For example, Kaptein et al. [2012] used Cialdini's six persuasive strategies [Cialdini, 2001] and developed a questionnaire to measure the user's susceptibility to those six persuasive strategies. They tested a setup where they tailored to the user's susceptibility versus a contra-tailored setup in the context of snacking and found a greater decrease in the tailored version. Halko and Kientz [2010] explored the relationship between personality (i.e., Big Five) and strategies in the context of health-promotion with mobile applications. Their results showed that people with different personality types had distinct preferences for (mobile) persuasive messages (for healthy living). Hirsh et al. [2012] let participants rate the level of persuasiveness of the messages for advertisements that were tailored to each personality type. For example, people with the Extraversion personality type would receive messages like "With XPhone, you'll always be where the excitement is" [Hirsh et al., 2012, p. 579] because extraverts are especially sensitive to rewards and social attention. The results showed a clear benefit of tailoring messages to personality type. Similar results were obtained in a study by Adnan et al. [2012] where an application was developed to persuade users to study more using persuasion strategies that were tailored to users' personalities: different personalities indeed preferred different (persuasive) study behaviors. Similar results were also obtained by Alkış and Temizel [2015], where they showed significant relations between personality traits and persuasive strategies. Finally, there have been some studies that suggest that future research should tailor messages that promote physical activity to people's personalities [Latimer et al., 2010], and that "Individuals with certain personality traits are more likely to be perceptive toward the idea of physical activities" [Arteaga et al., 2010, p. 8]. All these studies indicate that personality is a decisive factor in explaining the individual nature of people and their motivations and barriers for physical activity participation.

Another factor that has been researched in the context of tailoring and is also discussed as a determinant for physical activity is gender. Kristan and Suffoletto [2015] looked at how men and women rated feedback messages to reduce hazardous alcohol consumption and found that overall women responded more favorably to all feedback messages. Yan et al. [2015] found that women were not driven by competitive messages but men were. Busch et al. [2016] personalized persuasive strategies based on gender (measured on a fine-grained scale from masculine to feminine) and found that gender is a reliable variable for personalization, with femininity being significantly related to seven of their ten strategies. Like personality, we are interested in gender as a characteristic to tailor interventions to because it is an important determinant and gender is a stable personal characteristic, thus allowing longer-term tailoring to.

Overall, these studies indicate the potential effectiveness of tailoring strategies to personality or gender, and therefore of systems that can tailor to individual differences. Unfortunately, research is scarce on how systems should tailor motivational strategies or interventions for physical activity to preferences emerging from factors such as gender and personality. Moreover, research is scarce on how to do this tailoring with strategies from behavior change theory.

3.3 Behavior change theory for physical activity

As mentioned in Chapter 1, there are many theories and models on changing and influencing behavior, ranging from more practical, such as Persuasive Design [Fogg, 2003] to more theoretical ones, such as the Social Ecological Model [Stokols, 1996]. The use of theories or models in designing interventions to change behavior has been advocated (see [Cole-Lewis and Kershaw, 2010; Michie et al., 2008]) because it will help evaluate the model and the use of this theory or model can then help in explaining why the designed intervention strategy does or does not work. Hence, a theoretical foundation will help in understanding and targeting determinants of behavior, such as the stages of change. However, it is also noted that there is little guidance on how to use theories or models in designing concrete interventions [Michie et al., 2008].

For choosing the theory or the theory-based strategies that we want to use, we follow the line of [Glanz, 2015, p. 34]: "The choice of a suitable theory or theories should begin with identifying the problem, goal, and units of practice, not with selecting a theoretical framework because it is intriguing, familiar, or in vogue. One should start with a logic model of the problem and work backwards to identify potential solutions." The problem we are addressing is that of too little physical activity. Overall, people do not do enough physical activity, and our goal therefore is to motivate people in changing or maintaining their own physical activity behavior. Our unit of practice are individuals, specifically we want to tailor to individuals. Considering this, we chose the Transtheoretical Model (TTM) because the TTM fits our problem, our goal, and our unit of practice. In addition, another important consideration is of a practical nature. To apply the behavior change strategies in technology, they need to be operationalized. Many behavior change theories are not very specific in how their strategies could be practically used, the TTM however: *"provides a framework for both the conceptualization and the measurement of behavior change, as well as facilitating promotion strategies that are individualized and easily adapted"* [Nigg et al., 2011, p. 7]. In other words, the TTM provides strategies that can easily be adapted, i.e., made practical (for example, in the form of text messages), and provides a framework in which behavior change is formalized into measurable concepts, such as self-efficacy

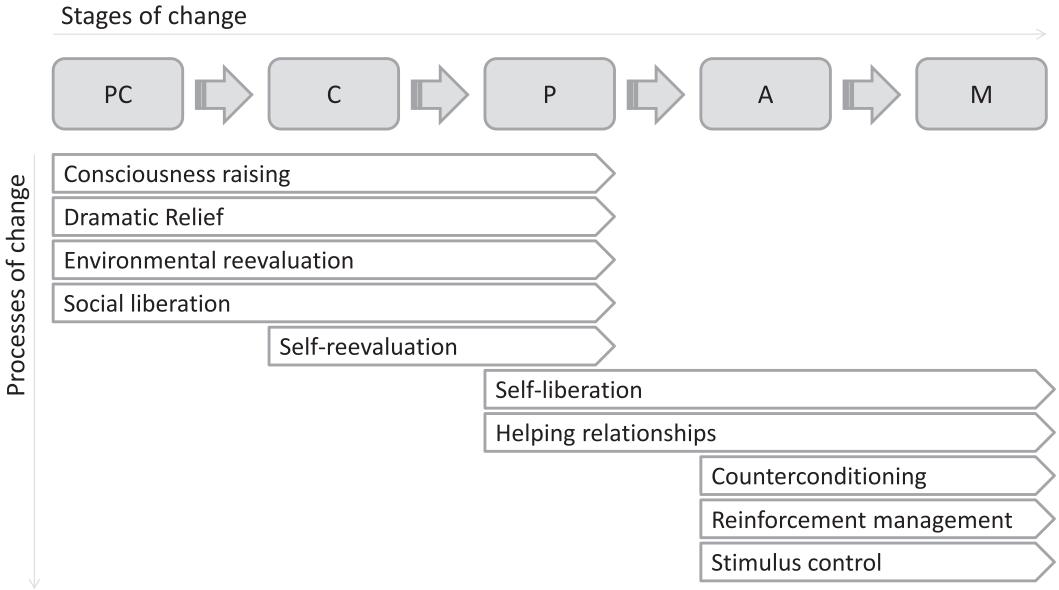

Figure 3.1: A general linear representation of the association of the stages with the accompanying processes. Depending on context, population and target behavior, this representation can change. PC = Precontemplation, C = Contemplation, P = Preparation, A = Action, and M = Maintenance. The length of the bar indicates in and over which stages the processes would most effectively be used.

and the stage of change. This more practical model of behavior change also allows operationalizations with higher fidelity, i.e., operationalizations more accurately resembling what is intended by the TTM.

The TTM from Prochaska et al. [Prochaska and DiClemente, 1983; Prochaska et al., 1993; Prochaska and Velicer, 1997] is a dynamic, integrative model focused on the individual and can be applied practically, especially the construct of the stages of change [Nigg et al., 2011], which classifies people into (not necessarily linearly) progressing stages of changing behaviors: Precontemplation (PC), Contemplation (C), Preparation (P), Action (A) and Maintenance (M). While the stages of change (see Table 3.1) are useful in explaining when changes in cognition, emotion, and behavior take place, the processes of change from the TTM help to explain how and why the progression through these stages occur. Ten covert and overt processes will usually be experienced when successfully progressing through the stages of change and achieving the desired behavioral change. The ten processes (see Table 3.2) can be divided into two groups: Experiential processes and Behavioral processes. Experiential processes are focused on changing people's ideas while Behavioral processes are focused on changing people's actions. The ten processes are Consciousness raising (CR), Dramatic relief (DR), Environmental reevaluation (ER), Social liberation (SOL), Self-reevaluation (SR) — Experiential, Self-liberation (SEL), Helping relationships (HR), Counterconditioning (CC), Reinforcement management (RM) and Stimulus control (SC) — Behavioral. The effectiveness of the processes of change depends on their associated stages of change (see Figure 3.1 for a simplified example). The processes are strongly associated with certain stage transitions, but they are not completely fixed.

Stages of change
Precontemplation (PC)
The individual is not willing to change in the foreseeable future (measured as the next 6 months). Individuals in this stage are mostly uninformed or demoralized.
Contemplation (C)
The individual is willing to change in the next 6 months. Individuals in this stage are aware of some pros of behavior change, but are still more inclined to value the cons.
Preparation (P)
The individual is willing to change in the foreseeable future (measured as the next month) and has already taken some small steps towards change (in the past year). Individuals in this stage usually have some plan on how to tackle this inactiveness.
Action (A)
The individual has changed, but not longer than 6 months. Individuals in this stage have not reached the duration which exemplifies real behavior change.
Maintenance (M)
The individual has changed, longer than 6 months. Individuals in this stage have changed and are working not to relapse.

Table 3.1: The stages of change with a short description.

Research has shown that stage-based interventions can be more effective than neutral interventions according to Marcus et al. [1998a]. In their study, they mailed (at baseline, 1, 3 and 6 months) intervention materials. A tailored intervention (tailored to the participant's stage of change, associated processes and more) versus a neutral intervention was tested and they found that both increased physical activity levels, but the tailored version increased physical activity levels most. This is a good example of the success of tailoring, but also the success of applying the TTM. In fact, when tailoring to the stages of change, it is more important than anything else to combine the stages with the processes of change [Spencer et al., 2006], because they are codependent. The importance of combining both is clear from several studies and as Spencer et al. conclude in their review of TTM literature *"Ensuring that participants use the appropriate processes of change as they move through the stages is essential for their success."* [Spencer et al., 2006, p. 438]. Moreover, it has been suggested that the TTM and/or stages of change can benefit from personalization [Ferron and Massa, 2013]. Accounting for individual preferences, for example those caused by differences in personality [Alkış and Temizel, 2015], could increase the effectiveness even further. In this sense, the TTM fits perfectly with our vision of using strategies that can be operationalized and can be tailored to individual differences like personality and gender, and, already included in the TTM, tailored to stages of physical activity behavior change (which we will refer to as stages of change in this dissertation).

Experiential Processes
Consciousness raising (CR): The individual seeks increased knowledge about the causes, consequences and cures for their problem behavior.
Dramatic relief (DR): The individual's emotions about the problem behavior and possible solutions are evoked.
Environmental reevaluation (ER): The impact that the individual's problem behavior has on their environment is reevaluated.
Social liberation (SOL): Attempts are made to increase alternatives for the individual's former problem behavior.
Self-reevaluation (SR): Cognitions and emotions regarding the individual with respect to their problem behavior are reevaluated.
Behavioral Processes
Self-liberation (SEL): The individual has the belief that he can change and commits to it by choosing a course of action.
Helping relationships (HR): The individual seeks trust and open discussion about the problem behavior as well as support for the healthy behavior change.
Counterconditioning (CC): The individual substitutes positive behaviors for the individual's problem behavior.
Reinforcement management (RM): Steps or changes made by the individual are rewarded when in a positive direction or punished when in a negative direction.
Stimulus control (SC): Stimuli that may cue a lapse back to the problem behavior are avoided and prompts for more healthier alternatives are inserted.

Table 3.2: The processes of change divided in experiential and behavioral processes with a short description.

3.4 Conclusion

More factors need to be considered when tailoring behavior change strategies to people. Based on the discussed literature, we expect that personality and gender will influence how motivating theory-based strategies are perceived, and that therefore these factors can be used to tailor the strategies of the TTM to. We argue that using theory-based strategies should be combined with tailoring. However, the content for these theory-based strategies is not readily available and unfortunately, there is little guidance on how to use theories or models in designing concrete interventions [Michie et al., 2008]. Studies describing the development of such technology do not often explain in detail *how* the researchers designed the motivational messages they used [Latimer et al., 2010]. How, then, should such strategies or interventions be operationalized? In Chapter 4 we present our approach to operationalizing the strategies of the TTM.

4 | Eliciting and categorizing peer-designed motivational messages for physical activity behavior change

In the previous chapter we argued that the Transtheoretical Model of behavior change is a suitable and appealing model to use to effectively support people in changing their physical activity behavior using technology. In this chapter we introduce our approach to operationalizing the processes of change from the Transtheoretical Model as motivational text messages. We explain how the data collection is set up to elicit motivational text messages and we explain our approach to coding these messages as processes of change categories. We present results on how the crowdsourced messages fit into categories based on the processes of change, and on the relation between personality and the stages, and between personality and the stages, and the processes of change.

This chapter[1] is based on two publications:

de Vries, R. A. J., Truong, K. P., & Evers, V. Crowd-Designed Motivation: Combining Personality and the Transtheoretical Model. *In International Conference on Persuasive Technology*. (pp. 41-52). Springer International Publishing. [de Vries et al., 2016a]

de Vries, R. A. J., Truong, K. P., Kwint, S., Drossaert, C. H. C., & Evers, V. Crowd-Designed Motivation: Motivational Messages for Exercise Adherence Based on Behavior Change Theory. *In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. (pp. 297-308). ACM. [de Vries et al., 2016b]

4.1 Introduction

Strategies operationalized and presented by means of tailored text messages to influence someone's behavior has proven effective in various contexts. For example, Mutsuddi and Connelly [2012] showed that tailored text messages in combination with a behavior change theory or model to influence someone's behavior can be effective for physical activity. Rimer and Kreuter [2006] showed that tailoring text messages according to a theory or model can enhance motivation to attend to and process health

[1]We thank Sigrid Kwint for her efforts in developing the codebook and her efforts in coding the peer-designed messsages.

information. This dissertation will focus on text messages as the modality to deliver the strategies. Unfortunately, as pointed out by Latimer et al. [2010], most studies describing the development of technology using text messages to shape motivational strategies do not, up until now, explain in detail *how* the researchers designed the text messages used. Therefore, there are no best practices available to construct these messages. To overcome this limitation, research dedicated to discerning how to translate theories and models of behavior change (such as the TTM) into practically applicable interventions (such as text messages) and evaluating the effect of those interventions on the user, is increasing.

For example, Redfern et al. [2014] designed (based on behavior change techniques) 137 text messages tailored to the user's name. Patrick et al. [2009] developed 3000 SMS and MMS messages for tailoring to the user's preferences on timing and frequency of the messages. Kaptein et al. [2012] had two researchers design 42 text messages for six strategies tailored to the user's susceptibility to persuasion. For most studies, it is the authors or other experts who designed the messages. The (implicit) assumption is that authors or experts, given their arguably intimate domain knowledge, will always design the most effective text messages. However, crowdsourced peer-designed text messages can be more engaging and more relevant to the user in comparison to expert-designed text messages, as is shown by Coley et al. [2013]. Therefore, we decided to use crowdsourcing to collect peer-designed motivational messages.

Crowdsourcing involves employing a large number of people to contribute to a specific task. Over the last few years, researchers have been using online crowdsourcing platforms such as Amazon Mechanical Turk (AMT) for a growing variety of tasks; for example, for user-studies [Kittur et al., 2008], graphical perception tasks [Heer and Bostock, 2010], parallel prototyping tasks [Dow et al., 2010], evaluating user interfaces [Toomim et al., 2011], but also for natural language processing tasks [Callison-Burch and Dredze, 2010].

Crowdsourcing written transcriptions, translations or annotations (e.g., [Hsueh et al., 2009; Marge et al., 2010; Zaidan and Callison-Burch, 2011]) is a relatively common natural language processing task, but we were aware of only two related studies on collecting motivational text messages. Coley et al. [2013] crowdsourced peer-written text messages to encourage users to quit smoking and compared these text messages to expert-written text messages. They found the *peer-written* text messages to be more engaging and relevant to the user. They also found that the peer-written messages reflected the same key theoretical concepts also addressed in the expert-written messages. In the context of reducing alcohol consumption among young adults, Kristan and Suffoletto [2015] also looked at peer-written text messages and evaluated expert-written text messages. They found a difference in what messages (informational, motivational, or strategy facilitating) experts come up with compared to peers, where the motivational expert-designed messages were rated the lowest. However, when looking at the peer-designed messages, the highest proportion of messages were motivational, which could suggest according to Kristan and Suffoletto [2015, p. 50] that: "young adults may perceive someone else trying to motivate them as manipulative, but when they create messages, they are more authentic".

Moreover, they found that there were no universal positive attitudes towards any of their messages, suggesting the need for tailoring. By using crowdsourcing, we can study the message content and effectiveness for a large number of messages that are designed to motivate people to change their behavior.

In this chapter, we will explain the steps we took to both operationalize behavior change strategies (the processes of change) from the Transtheoretical Model as practical motivational messages, as well investigate how we could tailor the processes of change, measured through a questionnaire, to the stages of change and personality of the respondent.

We make the following contributions: 1) we present the novel approach we use to operationalize the processes of change as motivational messages; 2) we show that, through coding and the use of a codebook, motivational messages can be aligned to the processes of change; 3) we show that crowdsourcing is a feasible way to collect theory-based motivational messages; 4) we provide the codebook used for coding; 5) we find that personality is correlated to the stage of change; and 6) we find that personality and stage of change are correlated to the experience of certain processes of change.

The chapter is outlined as follows: first we discuss our hypotheses (section 4.2), and our method (section 4.3). Next we discuss our data analysis and the approach we followed for coding the motivational messages in our dataset (section 4.4). We then continue with the results of the analyses we did on how personality and the stages relate, how the processes of change, measured through a questionnaire, related to the participant's stage of change and personality, and on what the distribution of the motivational messages is over the different processes of change and across the stages of change of the designer of the messages (i.e. the participant) (section 4.5). We end the chapter with a discussion (section 4.6) and a conclusion (section 4.7).

4.2 Hypotheses

In the previous chapters we looked into already existing technology that tries to motivate and/or support people (Chapter 2) and we looked into theories, models and individual differences in the context of physical activity and changing physical activity behavior (Chapter 3). We found that the Transtheoretical Model of behavior change seems a suitable and appealing model for our goal of effectively supporting people in changing their physical activity behavior using technology as it provides, as Nigg et al. [2011, p. 7] conclude: *"a framework for both the conceptualization and the measurement of behavior change, as well as facilitating promotion strategies that are individualized and easily adapted"*. We found several individual differences that influence physical activity that we decided are appealing to use for tailoring, like the stages of change, personality, and gender. Moreover, it has been suggested that the stages of change can benefit from personalization [Ferron and Massa, 2013]. The next step in getting our conceptual framework (the Transtheoretical Model combined with tailoring the strategies of the TTM to individual differences) into technology is to operationalize the strategies. That is what our first research question for this dissertation (see section 1.2) focuses on.

RQ1: How can theory-based strategies be translated into a real-world technology-based intervention?

We present our scenario-based crowdsourcing method for collecting peer-designed motivational messages and we evaluate whether these messages can align to the processes of change from the TTM. Because previous research [Coley et al., 2013] showed that peer-written messages reflected the same key theoretical concepts as expert-written messages, we expect that:

H1: The stage described in the scenarios that participants have to design motivational messages for, influences the type of motivational messages that participants come up with.

Specifically, the stage described in the scenario will more often trigger the type of message that, after categorization as process of change, fits the stage as defined by the TTM (see Figure 3.1). Messages reflecting Experiential processes (see Table 3.2) will be more prevalent in the earlier stage-scenarios, while messages reflecting Behavioral processes will be more prevalent in the later stage-scenarios. Although not the main focus of the research question for this chapter, this study also provides the opportunity to explore the relation between personality and the stages, and between personality and the stages, and the processes of change. This can provide invaluable insight for our next study and the second research question. For the first relation, we expect that:

H2: People's personality relates to how physically active they are.

Specifically, given that Extraversion and Conscientiousness have been found to correlate positively to fitness and health [Rhodes and Smith, 2006; Hoyt et al., 2009; Gallagher et al., 2013], and Neuroticism was found to correlate negatively to physical activity [Rhodes and Smith, 2006], we expect to find that:

H2a: People who are more Conscientious will be more physically active.
H2b: People who are more Extraverted will be more physically active.
H2c: People who are more Neurotic will be less physically active.

For the second relation between personality and the stages, and the processes of change, we expect that:

H3: People's personality and stage of change relate to how they experience the ten processes of change.

However, since this seems to be the first effort to investigate personality traits in the context of the stages and processes, we have no expectation specifically about how traits, stages and processes are related and it should therefore be considered as an exploratory hypothesis. Moreover, we will not be able to determine the causality of these relations because of the cross-sectional nature of the study.

4.3 Data collection: peer-designed motivational messages

We designed an online survey study with language-elicitation tasks to collect motivational messages and questionnaires to measure participant's personality, stage of change and experiences with the processes of change (see also subsection 4.3.5). The participants were gathered through the crowdsourcing platform Amazon Mechanical Turk[2] (AMT), with a link to SurveyMonkey[3] where the survey study was hosted.

4.3.1 Participants

The initial sample consisted of 500 people. Data from 19 respondents was excluded because their questionnaires were incomplete. The final sample included 481 respondents (250 male and 231 female). The study was conducted in English. All but 5 respondents were native English speakers, their data was not anomalous and was kept.

The minimum age was 18 and the maximum was 68. The average age was 31.09 (SD = 9.22) and the median 29. With respect to education, 201 respondents received some college education, 183 had a college degree, 46 completed a master degree, 42 completed high school, 5 obtained a PhD, and 4 had other types of qualifications. The AMT requirements for the respondents were that they should have already completed >1000 tasks on AMT, that >98% of these tasks were approved successfully and that the respondents were located in the United States. These requirements ensured that respondents were already familiar with surveys (our survey was quite extensive), that they were serious about filling in the survey (only 19 were not, which is low for online anonymous surveys), and that they had some proficiency in English. In fact, 473 reported 'very good' as self-assessed proficiency, 7 'good' and 1 'average', but none 'bad' or 'very bad'.

Although our sample consisted only of AMT workers, which could misrepresent the general adult population of a Western society, AMT can be considered to deliver an acceptable representation, especially by online survey standards [Mason and Suri, 2012]. However, we are aware that using a sample of the AMT population has limitations in terms of generalizability.

4.3.2 First scenario-based language-elicitation task

To obtain motivational messages, we developed a language-elicitation task where participants received one of five versions of a scenario, each version about a different stage of change. The participant was given one short paragraph (the scenario) in which a person in a specific stage of change (for instance, Precontemplation) was described and for whom participants had to design messages. The version of the scenario that the participant received was randomized. We asked the participants to

[2] https://requester.mturk.com/
[3] https://www.surveymonkey.com/

imagine that they had to motivate this person to exercise. In total, participants were asked to come up with six messages, three messages that would ostensibly be used over a longer period of time (for example 1 year) being provided to the recipient every other week, and three messages that would ostensibly be used over a short period of time (for example a month) being provided three times a week. Examples of scenarios and actual participant responses are shown in Table 4.1.

Stage of change scenario and **participant-designed text message**
Precontemplation: "Consider a middle-aged person, with a steady personal life and solid friend foundation. This person lacks regular exercise in his/her daily life and is unwilling to consider starting with this, at least not within the next 6 months."
• **Self-reevaluation**: "You would really feel better if you started exercising regularly"
• **Dramatic relief**: "Do you really like being a muffin top?"
• **Helping relationships**: "Maybe you can get one of your friends to work out with you"
Maintenance: "Consider a middle-aged person, with a steady personal life and solid friend foundation. This person participates in regular exercise in his/her daily life, and has been doing so for an extended period of time. This person has been active for more than 6 months."
• **Self-liberation**: "Keep it up!"
• **Reinforcement management**: "You have really lost a lot of weight! Keep up the good work on your exercise!"
• **Counterconditioning**: "You can spend your time on more exercises that will remove all your stress"

Table 4.1: Examples of two participants' responses coded as processes of change for a scenario-described person who is in a certain stage of change.

4.3.3 Second scenario-based language-elicitation task

To obtain motivational messages that could be used right before, during, and right after actual exercise, we developed another scenario-based language-elicitation task with three scenario versions. In this case, the participant received a scenario about a fictional character during, after and before a run in three short paragraphs. These paragraphs describe either a character that has had a good run, a run with too low intensity, or a run with too high intensity. For all the phases in the scenario (during, after, before), the participant was asked to design messages that could motivate someone in that situation. The data from this second language-elicitation task has not been used in this dissertation.

4.3.4 Voice data

To obtain the text-based motivational messages that the participants designed also as speech data, we asked participants to record them if they had access to a microphone. We presented participants with the 15 text-based motivational messages (6 for the first language-elicitation task and 9 for the second language-elicitation task) they had designed and asked them to record it in the way they meant the message to be heard. The participants could re-record the messages if they were not satisfied with the result. 19 participants (of the 481) did not have access to a microphone. The final sample of the voice data includes 464 participants (246 male). The data from this speech-recording task has not been used in this dissertation.

4.3.5 Measures

At the start of the survey, we asked participants about their gender, age, native language, understanding of the English language, education level, maternal education level (as an indication of socioeconomic status [Green, 1970]) and main field of work. Then, we presented participants with the scenario tasks of the survey and the recording page for the voice data. At the end of the survey, we measured participants current stage of change[4] [Norman et al., 1998], the frequency of weekly leisure-time physical activities (Godin Leisure-Time Exercise Questionnaire [Godin and Shephard, 1997]), perceived experiences with processes of change[4] [Nigg et al., 1999], self-efficacy[4] [Benisovich et al., 1998], and decisional balance[4] [Nigg et al., 1998] for exercise. To measure participants' personality we used the 50-item IPIP representation of the revised version of Costa and McCrae's [Costa and McCrae, 1992] NEO Personality Inventory[5] which posed 50 statements (e.g., "Make plans and stick to them."). Participants were asked to answer how descriptive they found these statements of themselves (on a 5-point Likert scale, 1 being "very inaccurate" and 5 being "very accurate"). We ended the survey with a yes/no question asking if the participant was already familiar with the Transtheoretical Model.

4.3.6 Procedure

Participants were recruited through AMT. They were informed of their compensation, the goal of the survey and the estimated time to complete the survey. Participants could then decide to accept or decline and proceed to the SurveyMonkey website where the survey was hosted. There, the goal of the survey was summarized and participants were asked to complete a consent form. The participants were then presented with the demographic questions, the first language-elicitation task, the second language-elicitation task and the voice-recording task. Next, participants were presented the questionnaires as detailed in section 4.3.5. Afterwards, participants were debriefed about the goals of the survey and given a completion code to fill in on AMT to receive payment. On average, the survey took about 45 minutes to complete. Participants were compensated with 3 US dollars for their participation.

4.4 Data analysis

The survey was distributed to 500 respondents. However, data from 19 respondents was excluded because their questionnaires were incomplete. This is due to the possibility to finish the survey on AMT, while not having completed the survey on Survey-

[4] `http://web.uri.edu/cprc/measures/`
[5] `http://ipip.ori.org/`

Monkey. This also resulted in these respondents not filling in the completion code, and therefore their data was excluded. The final sample we worked with included 481 respondents (250 male and 231 female) with no missing data except for the voice data of 19 respondents (see section 4.3.4). However, recording voice data was not a requirement to complete the survey.

For the data analysis, we are interested in the relation between personality and the stages (H2), and between personality and the stages, and the processes of change (H3). For the first relation, based on literature we expected that certain personality traits scores significantly relate to self-assessed stage of change. An ordinal logistic regression was carried out to determine the effect of the traits (OCEAN) on the self-assessed (ordinal) stages of change. For the second relation, we expected that the stages of change and different personalities would relate to different self-assessed processes of change in relation to exercise, but we had no expectation specifically, about how traits, stages and processes related. We were interested in understanding the relations between the continuous personality trait variables (O, C, E, A and N) of the compound variable personality and the continuous variables of the processes of change, which we can assess with regression coefficients. Hence we carried out multiple regression analyses.

To analyze whether the stage described in the scenarios that participants had to design motivational messages for, influenced the type of motivational messages that participants came up with (H1), we first need to code the messages.

To see how the content of the 2886 (481 x 6) crowdsourced peer-designed messages reflected the processes of change across the stage-scenarios (H1), we translated the processes of change from the TTM to coding categories. We used a similar process to MacQueen et al. [1998]. To use the coding categories, we operationalized the fairly fluid processes of change descriptions into more fixed definitions that considered the perspective of 'text messages sent by a peer'. The procedure started with two coders (coder 1 was the main investigator and author of this dissertation) separately coming up with operationalized definitions of the processes of change for a first version of the codebook.

The codebook was developed following the guidelines of MacQueen et al. [1998] and Guest and MacQueen [2007] who advise structuring the codebook with (at least) six parameters: the code, brief definition, full definition, when to use, when not to use, and examples. We added a seventh parameter, namely the 'perspective', which we defined as an alternative definition of what the process could look like in terms of a text message. Each category (a process of change) for coding was described in the same manner. Because the majority of the messages were only a few words long, it was decided that only one code per message would be used. The context for the definitions of the processes of change was the TTM in general, but also the TTM specifically in the context of exercising. A unified codebook was defined before coding started.

The messages were coded iteratively, independently, and without scenario information by the two coders and afterward the coders resolved any mismatches. After an iteration, the codebook would be updated to reflect these resolved mismatches, and then the next round of coding would start. The first round of coding started with

60 messages and this increased in steps to a maximum of 300 for the final round of coding (approximately 10% of the data). The final round of coding was declared when the coders felt they had reached a saturation point at which achieving a higher agreement would not be feasible. The final agreement was a Cohen's kappa of 0.72, which is substantial [Landis and Koch, 1977].

For the next planned study (described in Chapter 5), we needed only the most representative messages for each of the process categories. To this end, we selected only the best representations of the processes through a follow-up coding, designed to complement the last round of coding. The follow-up coding used a 'certainty' measure along with the existing coding. This certainty measure meant that both coders would add a certainty code to their 'normal' coding of the messages if they were 99% sure about the message belonging to that particular coding category (i.e., one of the ten processes) and that the other coder would agree with this. The messages that were coded by coder 1 scored a Cohen's kappa of 0.86 for the process-categories when selecting only the messages with the certainty measure.

The remaining messages (2586 messages) were coded by one coder (coder 1), yielding a dataset containing 2886 coded messages in total and a subset of 1433 (49.7%) messages coded with the certainty measure (examples of coded messages are given in Table 4.1 and a snippet of the codebook in Table 4.2, for the full codebook see Appendix A).

The motivational messages from 481 participants were used in the coding process. It is important to note that the experimental stages of change scenarios were not completely equally distributed: 91 participants designed motivational messages for the Precontemplation scenario (PC-S), 93 for the Contemplation scenario (C-S), 87 for the Preparation scenario (P-S), 102 for the Action scenario (A-S), and 108 participants for the Maintenance scenario (M-S). This means that each scenario did not receive exactly the same amount of peer-designed messages.

To analyze whether the stage described in the scenarios that participants had to design motivational messages for, influenced the type of motivational messages that participants came up with (H1), a Chi-square test was carried out to see if the condition (i.e., the stages) had an effect on the counts within the higher-order (i.e. the Experiential and Behavioral) processes. At a lower level, we also looked at the distribution for the ten separate processes. A Chi-square test was again performed to see if the condition (i.e., the stages) had an effect on the counts of the processes.

4.5 Results

In this section, we present the results of the online survey and how the crowdsourced messages fit into categories based on the processes of change. Our hypothesis was that the stage described in the scenarios that participants have to design motivational messages for, influences the type of motivational messages that participants come up with (H1). We looked at what kinds of messages (process categories) and how many of each kind people would come up with for each of the stage of change scenarios, and whether the distribution of messages more or less aligned to our prediction: the messages reflecting Experiential processes were expected to be more prevalent (have

Process – Consciousness Raising/Increasing Knowledge	
Coding – CR	
Brief definition – Increased awareness of causes, consequences and cures for not being physical active.	
Full definition – CR is a process that involves increased awareness of causes, consequences and cures for not being physical active. The intention is increasing the knowledge of unaware individuals with objective information.	
Practical Definition – Encourage subject to read and think about physical activity (cognitive process)	
Perspectives – To start/trigger or advance the process: messages that give (objective) information about the benefits or disadvantages of not exercising.	
Inclusion – Arguments with information (mostly) facts about benefits or disadvantages for health; (objective) confrontations with diseases; prevention for diseases.	
Exclusion – Subjective arguments why people should exercising; benefits for appearance; a proposal.	
Example inclusion – "It can prevent all types of diseases like Diabetes and cancer" – "Exercise can help you live longer" – "Exercise will help you to be healthy and fitt."	
Example exclusion – "You worked hard for everything, why not also for your health?" - This would fit better with SR; "You will look much better; You will feel better when you exercise." - This would fit better with SR	

Table 4.2: An example of one of the processes of change, Consciousness Raising, translated to a coding category.

more counts) in the earlier stages compared to the later stages, while the messages reflecting Behavioral processes were expected to be more prevalent (have more counts) in the later stages compared to the earlier stages. We also looked at whether the same distribution of counts for the messages was shown for each of the separate process categories. Moreover, we present the results of the relation between personality and the stages, and between personality and the stages, and the processes of change. We hypothesized that there is a relation between people's personality and the stages of change they are in (H2) and that there is relation between personality traits, the stages of change and the ten processes of change (H3). Data from 481 participants was analyzed. Important to note is that the self-assessed stages of change measure was not equally distributed: 175 participants rated themselves to be in the *Maintenance stage* (M), 114 in the *Preparation stage* (P), 91 in the *Action stage* (A), 68 in the *Contemplation stage* (C), and 33 participants rated themselves in the *Precontemplation stage* (PC).

4.5.1 Distribution of the messages over the processes (H1)

To see whether the distribution of coded messages follows our prediction, we first looked into the higher-order Experiential and Behavioral processes. The results are shown in Table 4.3. Of the total of 2886 messages, 800 (27.7%) are coded as Experiential processes, and 2086 as Behavioral (72.3%) processes. Table 4.3 shows that the distribution of the higher-order processes over the stages resembles our expectation: Experiential processes are more prevalent in the earlier stages and Behavioral processes are prevalent in the later stages (see the "count" row of Table 4.3). The results show that there is a significant association between the stages and higher-order processes ($\chi^2(4) = 223.179, p < .001$). The values of the standardized residuals are used to further interpret the results of the Chi-square test. The standardized residuals represent the error between the observed frequency (i.e., what the data actually shows) and expected frequency (i.e., what the model predicts) if no relation would exist between the stages and the coded messages (null hypothesis). A positive value indicates an overrepresentation and a negative value points to an underrepresentation. A z-value higher than 1.96 or lower than -1.96 for either the over or underrepresentation is considered to be significant at $p < 0.05$ [Field, 2013].

The residuals show that for the Experiential processes in the Precontemplation stage, there is a significant overrepresentation of the processes ($z = 8.9$), and in the Action and Maintenance stage there is a significant underrepresentation of the processes ($z = -6.0$ and $z = -6.5$). For the Behavioral processes this is reversed, in the Precontemplation stage, there is a significant underrepresentation of the processes ($z = -5.5$), and in the Action and Maintenance stage there is a significant overrepresentation of the processes ($z = 3.7$ and $z = 4.0$). Overall, this means that there are more Experiential messages in the earlier stages and fewer in the later stages than the Chi-square model predicts. Also, this means there are fewer Behavioral messages in the earlier stages and more in the later stages than the Chi-square model predicts.

At a lower level, we also looked at the distribution for the ten separate processes. Table 4.4 shows that the distribution of the processes over the stages somewhat resembles our expectation for some, but not all, of the processes. The results show that there is a significant association between the stages and processes ($\chi^2(36) = 437.851, p < .001$). The values of the standardized residuals are used to further interpret the results of the Chi-square test.

For the Experiential processes, Conscientiousness Raising, Dramatic Relief, Environmental Reevaluation and Self-reevaluation show a significant trend of more counts in the earlier stages and fewer in the later stages. The Experiential process of Social Liberation has too few counts to interpret any results.

For the Behavioral processes, only Reinforcement Management shows a significant trend of fewer counts in the earlier stages and more in the later. Self-liberation and Counterconditioning both show some inclination toward this trend near the Action stage, but decline for the Maintenance stage. Helping Relationships is more or less equal throughout the stages and Stimulus Control displays the opposite (Experiential) trend significantly.

Overall, the distribution over the stages for both higher-order processes is in line with our first expectation, but when looking at the distribution of the separate processes, the same trend is not present for all ten processes.

Categories/Stage scenarios		**PC-S**	**C-S**	**P-S**	**A-S**	**M-S**	Total
Experiential	Count	299	177	160	92	72	800
	Expected	179.6	154.7	144.7	169.6	151.4	800
	Std. residual	**8.9**^{3}	1.8	1.3	**-6.0**^{3}	**-6.5**^{3}	
Behavioral	Count	349	381	362	520	474	2086
	Expected	468.4	403.3	377.3	442.4	394.6	2086
	Std. residual	**-5.5**^{3}	-1.1	-0.8	**3.7**^{3}	**4.0**^{3}	
Total	Count	648	558	522	612	546	2886

Table 4.3: The distribution of all the codes over the 2 higher-order process categories and 5 stages of change scenarios. $^1p < .05$, $^2p < .01$, $^3p < .001$

4.5.2 Relation between personality and stages of change (H2)

An ordinal logistic regression showed that the general model (OCEAN) statistically significantly predicted the stages of change over and above the intercept-only model, $\chi^2(5) = 66.526$, $p < .001$. Concerning the contributing factors, an increase in Extraversion was associated with an increase in stage of change, with an odds ratio of 1.051 (95% CI, 1.028 to 1.075), $\chi^2(1) = 18.578$, $p < .001$. Moreover, a decrease in Neuroticism was associated with an increase in stage of change, with an odds ratio of 0.971 (95% CI, 0.946 to 0.996), $\chi^2(1) = 5.091$, $p < .024$. The other personality traits were not significantly ($p < 0.05$) related to stage of change (see Table 4.5).

4.5.3 Relation between personality and stages and processes (H3)

In Table 4.6 the standardized regression coefficient (β) scores from our multiple regression analyses are reported for the predictor variables stages of change and personality traits (OCEAN), and the outcome variables of the processes of change (10 times). The regression coefficient results suggest that different personality traits scores relate differently to processes of change. All processes are significantly correlated to at least one personality trait, and all personality traits are significantly related to at least one process. It should be noted that although there are significant personality-trait-to-process relations, the stages of change are usually a much larger predictor (this can be seen from the standardized β reported in Table 4.6). The personality trait results could be considered ‘nuances’ to the already existing relation between stages and processes.

4.6 Discussion

Using theory in practice is not always easy and effective. Although there is a general consensus on the value of most behavior change theories and more specifically the Transtheoretical Model, there is also still plenty of room to increase the effectiveness and salience of such theories by identifying more determinants (e.g., personality) for

Categories/Scenarios		**PC-S**	**C-S**	**P-S**	**A-S**	**M-S**	Total
	Count	67	18	36	10	7	138
CR	Expected	31.0	26.7	25.0	29.3	26.1	138
	Std. residual	**6.5**^{3}	-1.7	**2.2**^{1}	**-3,6**^{3}	**-3,7**^{3}	
	Count	33	8	13	4	1	59
DR	Expected	13.2	11.4	10.7	12.5	11.2	59
	Std. residual	**5.4**^{3}	-1.0	0.7	**-2.4**^{1}	**-3.0**^{2}	
	Count	33	13	9	14	10	79
ER	Expected	17.7	15.3	14.3	16.8	14.9	79
	Std. residual	**3.6**^{3}	-0.6	-1.4	-0.7	-1.3	
	Count	2	6	1	0	2	11
SOL	Expected	2.5	2.1	2.0	2.3	2.1	11
	Std. residual	-0.3	**2.7**^{2}	-0.7	-1.5	-0.1	
	Count	164	132	101	64	52	513
SR	Expected	115.2	99.2	92.8	108.8	97.1	513
	Std. residual	**4.5**^{3}	**3.3**^{3}	0.9	**-4.3**^{3}	**-4.6**^{3}	
	Count	178	168	190	221	182	939
SEL	Expected	210.8	181.6	169.8	199.1	177.6	939
	Std. residual	**-2.3**^{1}	-1.0	1.5	1.6	0.3	
	Count	44	53	31	59	51	238
HR	Expected	53.4	46.0	43.0	50.5	45.0	238
	Std. residual	-1.3	1.0	-1.8	1.2	0.9	
	Count	49	65	49	83	43	289
CC	Expected	64.9	55.9	52.3	61.3	54.7	289
	Std. residual	**-2.0**^{1}	1.2	-0.5	**2.8**^{2}	-1.6	
	Count	40	67	72	146	187	512
RM	Expected	115.0	99.0	92.6	108.6	96.9	512
	Std. residual	**-7.0**^{3}	**-3.2**^{2}	**-2.1**^{1}	**3.6**^{3}	**9.2**^{3}	
	Count	38	28	20	11	11	108
SC	Expected	24.2	20.9	19.5	22.9	20.4	108
	Std. residual	**2.8**^{2}	1.6	0.1	**-2.5**^{1}	**-2.1**^{1}	
Total	Count	648	558	522	612	546	2886

Table 4.4: The distribution of the coding over the ten process categories and stages of change scenarios with their respective counts. $^{1}p < .05$, $^{2}p < .01$, $^{3}p < .001$

Trait	M	SD	α	$ratio$	sig	CI
Openness to exp.	38.94	6.65	.77	1.015	.252	$[0.989 - 1.042]$
Conscientiousness	37.75	7.43	.89	1.025	.074	$[0.998 - 1.055]$
Extraversion	31.02	8.91	.90	1.050	**.000**	$[1.028 - 1.075]$
Agreeableness	38.24	6.14	.80	0.969	.057	$[0.939 - 1.001]$
Neuroticism	24.61	8.91	.90	0.971	**.024**	$[0.946 - 0.996]$

Table 4.5: Averages (M), standard deviations (SD), and Cronbach's alpha's (α) for all the construed scales. Scales are added, instead of averaged to keep origin clear. Personality scales are 10 items with scoring from 1 to 5 added up (possible scores from 10 to 50). Ordinal regression with stage of change as dependent variable and the personality traits (OCEAN) as independent variables. (N = 481)

PoC	M	SD	α	R^2	SoC β	O β	C β	E β	A β	N β
CR	8.81	3.38	.89	.246	**0.398**^{3}	**0.102**^{1}	0.014	**0.151**^{2}	0.005	0.018
DR	9.10	3.10	.75	.112	**0.278**^{3}	0.080	**0.127**^{1}	0.072	-0.068	0.115
ER	10.65	2.88	.75	.071	0.087	**0.160**^{2}	**0.114**^{1}	0.072	0.011	0.047
SOL	10.64	2.57	.63	.121	0.087	0.064	**0.147**^{2}	**0.171**^{2}	**0.171**^{2}	**0.143**^{1}
SR	12.34	2.75	.86	.260	**0.389**^{3}	**0.225**^{3}	0.014	-0.010	**0.119**^{1}	0.007
SEL	10.98	3.07	.82	.476	**0.619**^{3}	**0.083**^{1}	**0.112**^{2}	0.045	0.006	-0.001
HR	7.81	3.75	.90	.194	**0.297**^{3}	-0.077	0.035	**0.203**^{3}	0.027	-0.046
CC	8.08	3.29	.85	.462	**0.578**^{3}	-0.005	**0.146**^{2}	**0.111**^{2}	0.008	-0.011
RM	11.21	3.11	.84	.310	**0.441**^{3}	**0.164**^{3}	0.062	0.058	0.074	0.001
SC	8.30	3.45	.77	.335	**0.471**^{3}	0.053	**0.115**^{1}	**0.126**^{2}	-0.033	-0.015

Table 4.6: Averages (M), standard deviations (SD), and Cronbach's alpha's (α) for all the construed scales. Scales are added, instead of averaged to keep origin clear. Processes of change are 3 items with scoring from 1 to 5 added up (possible scores from 3 to 15). Standardized regression coefficients of personality traits, stages of change and the processes of change are reported. (N = 481) $^{1}p < 0.05$, $^{2}p < 0.01$, $^{3}p < 0.001$.

specific situations (e.g., the exercise domain) and by revealing new dependencies between them. As a first step towards combining personality and behavior change theory to motivate people to exercise, we explored the possibility of personality-based tailoring of the processes of change through crowdsourcing and self-assessment measures. We conclude that personality traits (E and N) relate to the stages of change and personality traits and the stages of change relate to preferences for certain processes of change. In the next chapter (Chapter 5) we will explore this further in the context of text messages representing the processes of change. The research question we tried to answer with this study is: how can theory-based strategies be translated into a real-world technology-based intervention? To this end, we explored the approach of using crowdsourcing to collect peer-designed messages. To evaluate whether these crowdsourced peer-designed messages fit theoretically-grounded behavior change strategies, we had to code the motivational messages according to the processes of change (the behavior change strategies). Concerning the coded messages, we found that the messages that people design for different stage of change scenarios *can* be reliably coded into categories of processes of change. In this section we discuss the implications of the results separately for each hypothesis.

4.6.1 Distribution of the messages over the processes (H1)

The results of the study, after coding the crowdsourced peer-designed text messages, showed that the stage described in the scenarios that participants had to design motivational messages for, influenced the type of motivational messages that participants came up with (H1).

The kinds of messages (process categories) people designed are different between the stages in the same way as predicted: Experiential processes are more prevalent in the earlier stages while Behavioral processes are more prevalent in the later stages. Although the distribution *within* the higher order processes over the stages aligned quite well with our predicted distribution, we did not have an prediction for the distribution of the number of messages *between* the higher order processes themselves. It is important to note that the distribution between the higher-order processes was skewed (2086 of 2886 were coded as Behavioral processes). Meaning that people came up with messages fitting mostly with Behavioral processes ($72,3\%$). This could mean that, in our study setup, people found it much easier to think of more 'action-oriented' Behavioral messages than more 'thinking-oriented' Experiential messages, or that in general people find it easier to think of more 'action-oriented' messages.

For the separate processes of change distributions over the stages, quite a few processes aligned to our expectation, although also several processes did not. Interestingly, most of the Experiential processes followed the predicted distribution, but not many of the Behavioral processes. The counts for the Behavioral processes (except Reinforcement Management) are reasonably stable across all stages. One interpretation could be that when someone is motivated to *learn* a behavior it is more natural to start earlier with Behavioral messages as well as Experiential messages. Also interesting is that the significant results are mostly for the Precontemplation and Maintenance stages: an interpretation could be that this is where the behaviors are stable, and the differences in the processes that are used largest.

We used scenarios based on the stages to collect a broad range of messages. The scenarios included general context (middle-aged, steady personal life, solid friend foundation) to make it realistic, which could have biased the participants, to mitigate this bias, the text was kept the same for all scenarios and concerned less text than the actual stage description (see Table 4.1). The coding was carried out without the stage information of the messages. Although a Chi-square test is not optimal to test the predicted distribution of messages, together with the message counts, this test does show that the scenarios influence the message content. There is no ratio for distribution over the different stages yet, so we could not do a rate comparison. We hope this research is a first step in dealing with a ratio for the distribution of messages.

4.6.2 Relation between personality and stages of change (H2)

The results of the study show that people's personalities are related to their progression through the different stages of change for exercise behavior (H2), specifically that the Extraversion trait was positively correlated with people progressing through the stages (H2b), while the Neuroticism trait was negatively correlated with progressing through the stages (H2c), but no (positive) relation was found for the Conscientiousness trait (H2a). In other words, people scoring higher on Extraversion are more likely to be progressed further into stages of behavior change, while people scoring higher on Neuroticism are more likely to be in the earlier stages of behavior change. No significant relation was found between Conscientiousness and the stages of change. Although the correlations of personality traits to the stages of change are relatively small, the results are in line with previous research on personality and physical activity [Rhodes and Smith, 2006] and considered still important for the health context. For Extraversion and Neuroticism, which are significantly correlated to the stages, one tentative explanation could be that people change their personality when changing their behavior. But, because it is believed that personality is relatively temporally stable [Rhodes et al., 2004], a more likely explanation is that people with low Extraversion and high Neuroticism scores need different triggers and see different barriers when trying to change their behavior (compared to high Extraversion, low Neuroticism scoring people) then those addressed in current motivational technology and programs and therefore these people have more difficulties in changing their behavior. This could be addressed by examining the relation between personality, the stages, and the processes of change.

4.6.3 Relation between personality and stages and processes (H3)

Our study also shows that there are relations between different personality traits and different processes of change they find important in relation to exercise (H3). Conscientiousness is related to six processes, Openness to Experience and Extraversion are related to five, Agreeableness to two, and Neuroticism to one. Interesting to see is that Neuroticism, which correlated negatively to the stages of change, does not (significantly) relate to many processes. This could support our previous interpretation that the processes believed to help people through the stages are not very appealing to people scoring high on Neuroticism and therefore they also do not progress through the stages. Similar results with health-promoting strategies for people scoring high on Neuroticism were found in previous work [Halko and Kientz, 2010]. Conscientiousness, which we expected to relate to the stages of change did not, but in turn correlated to the most processes. Previous work also expected a relation between Conscientiousness and the stages of change, but found that this was fully mediated by the relation between Conscientiousness and certain processes [Bogg, 2008]. In any case, the results show a relation between different processes of change and personality traits, which serves as an indication that the tailoring of processes to personality trait preferences could be very helpful in making the messages more salient for behavior change. For example, referring back to the previous section, people with low Extraversion and high Neuroticism do not seem to like the Social Liberation process (while high Extraversion, low Neuroticism scoring people do). However, appealing to the social dimension, like social comparison or competition (as we discussed in Chapter 2), is a popular technique in current motivational technology. It is important to note that we did not correct for the alpha-inflation of our multiple (10) tests, so the barely significant relations should be interpreted with caution.

Overall, crowd-designed TTM-informed motivational messages seems to be a promising approach to designing motivational messages that fit theoretically-grounded behavior change strategies. The messages reflected the processes of change. However, we do not yet know whether people in a specific stage of change will indeed find the TTM-tailored messages motivating as we expected.

4.7 Conclusion

As a first attempt to design motivational messages grounded in behavior change theory (the TTM's stages and processes of change), we carried out an elicitation survey that assessed the possibility of using crowdsourcing as a method to design the motivational messages. We collected peer-designed motivational text messages and manually coded these messages into categories based on the processes of change. We conclude that 1) people design motivational messages that reflect the processes of change; 2) these messages relate to the stages of change like the processes of change do (Experiential processes prevalent in earlier stages, Behavioral processes prevalent in later stages); and 3) we identified new dependencies between the different processes, stages and personality traits in the context of the exercise domain.

Overall, the findings in this paper can help inform researchers designing motivational technology for long-term behavior change who: 1) look for a method to translate theoretical insights to practical text messages; and who 2) want to go beyond a one-size-fits-all strategy and design effective motivational technology that tailors to the stage of change a user is in, but also to more individual factors, like personality.

Although there is a general consensus on the value of most behavior change theories and more specifically the TTM, there is still plenty of room to increase the effectiveness and the application of such theories by designing practical implementations of whole theories and testing these in various contexts, such as the exercise domain.

As part of a larger series of studies (described in this dissertation), we sought to leverage HCI practices, to come to useful and practical insights about how to translate theoretical constructs (i.e., the processes of change) to text messages and to explore theory on behavior change from psychological research to come to useful and practical insights about how to further adapt the processes of change to (robust) user characteristics (e.g., personality traits). We showed a method by which theoretical constructs of a behavior change theory or model (the TTM) can be translated to crowd-designed motivational text messages. Although these results are context-dependent, it could prove to be a valuable method for other theories, models or contexts. The motivational text messages will be used in motivational technology, specifically in a smartphone application, where users can receive these messages to get motivated to exercise or to be reminded to exercise regularly (see Chapter 7). To further optimize the effectiveness of these messages that represent the theoretically-grounded behavior change strategies, the next step (see Chapter 5) is to investigate whether there are differences in how people evaluate these messages. These differences can then form the basis for tailoring, by adjusting which strategies to use based on differences in their evaluation based on the stage of change, personality or gender of the people.

5 | Evaluating peer-designed motivational messages for physical activity behavior change

In the previous chapter we introduced our approach to operationalizing and coding the processes of change from the Transtheoretical Model as text messages. In this chapter we evaluate a selection of five of these text messages for each of the ten processes of change categories. Furthermore, we investigate whether there are individual differences in how people evaluate these messages that can be used to tailor to, like their stage of change, personality and gender.

This chapter is based on two publications:

de Vries, R. A. J., Truong, K. P., Kwint, S., Drossaert, C. H. C., & Evers, V. Crowd-Designed Motivation: Motivational Messages for Exercise Adherence Based on Behavior Change Theory. *In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. (pp. 297-308). ACM. [de Vries et al., 2016b]

de Vries, R. A. J., Truong, K. P., Zaga, C., Li, J., & Evers, V. A Word of Advice: How to Tailor Motivational Text Messages Based on Behavior Change Theory to Personality and Gender. *Theme Issue: Supporting a Healthier Lifestyle with e-Coaching systems of the journal Personal and Ubiquitous Computing*. (pp. 675-687). Springer International Publishing. [de Vries et al., 2017a]

5.1 Introduction

In the previous chapter we took a step towards translating behavior change theories and models into concrete interventions by crowdsourcing motivational text messages and coding these text messages into motivational strategies based on existing behavior change theory. The preferred next step would be to evaluate these messages in terms of their effectiveness. This is usually determined through a long-term in-the-wild study. However, long-term in-the-wild studies are usually resource expensive to set up and carry out. Therefore, a more feasible and practical approach was chosen to get a first impression of how the messages are perceived: an evaluation survey study. This approach gives us the opportunity to evaluate, with a large participant sample, a selection of messages (and the strategies that these messages represent) on how the messages are perceived, without the drawbacks of doing an in-the-wild study. Moreover, this approach also allows us to investigate the relation between our messages and the characteristics of interest that we would want to tailor to. Therefore, the next question is, what characteristics should one tailor an exercise application's theory-based motivational strategies to?

Also in the previous chapter, we explored the relationship between the processes of change and personality, both measured through a questionnaire. Moreover, in Chapter 3 we discussed how both personality and gender are individual differences found to relate to physical activity. In this chapter, we will explore how motivating people perceive the processes-of-change-aligned text messages to be in context of the stage of change they are in, but also how other user characteristics influence how motivating the processes-of-change-aligned text messages are perceived.

We make the following contributions: 1) we find that how motivating people perceive the processes-of-change-aligned text messages to be in context of the stage of change they are in, is different from how the processes are theoretically aligned per stage; 2) we find that the personality of a person (operationalized by their personality traits) influences how motivating the different processes-of-change-aligned text messages are perceived; and 3) we find that the gender of a person also influences how motivating the different processes-of-change-aligned text messages are perceived.

The chapter is outlined like this: first we discuss the hypotheses (section 5.2), and the method (section 5.3). Next we discuss the data analysis (section 5.4) and we then continue with the results of the analyses on how motivational messages are perceived (section 5.5). We end the chapter with a discussion (section 5.6) and a conclusion (section 5.7).

5.2 Hypotheses

In previous chapters we have discussed that designing and developing effective motivational technology to empower people to change or maintain their behavior is a challenge. To address this challenge, we decided to base the motivational strategies we want to use on existing behavior change theory and to tailor these strategies to characteristics of the user to increase the effectiveness of the strategy. In the previous chapter (Chapter 4) we operationalized and coded the processes of change from the Transtheoretical Model as text messages. To increase the impact of motivational strategies, we argue that tailoring is important and that more factors need to be considered when tailoring behavior change strategies to users. The next step is to see whether certain individual differences can be used to tailor theory-based motivational strategies to. That is what the second research question for this dissertation (see section 1.2) focuses on.

RQ2: How does tailoring the intervention to individual differences influence people's motivation for physical activity?

In this chapter, we present our evaluation of a selection of the text messages representing the ten processes of change (the behavior change strategies), and their relation to several user characteristics, namely the stage of change, personality and gender of the participant. For the relation between the text messages representing the ten processes of change and the participants stage of change, we expect that:

H1: The stage of change participants are in relates to how motivating they find the text messages that are aligned to the processes of change.

Specifically, given that we use the TTM in the context of exercise, we expect, based on the findings of Marcus et al. [1992], that the Experiential processes (see Table 3.2) are rated as more motivating (than the Behavioral processes) in the earlier stages of change, peak in the Action stage and decline in the Maintenance stage, while the Behavioral processes peak in the Action stage, but do not decline in the Maintenance stage. This means that, for example, a person in the Precontemplation stage would find the processes-of-change-aligned text messages of Consciousness Raising, Dramatic Relief, Environmental Reevaluation, Social Liberation and Self-reevaluation (all Experiential processes) more motivating than other processes, while a person in the Maintenance stage would find the processes-of-change-aligned text messages of Self-liberation, Helping Relationships, Counterconditioning, Reinforcement Management, and Stimulus Control (all Behavioral processes) more motivating than other processes. For the other individual differences, based on the discussed literature in Chapter 3, we expect that both factors, personality and gender, will influence how motivating the text messages representing the certain processes are rated. Therefore, we expect that:

H2: People's personality influences how motivating they find the text messages that are aligned to the processes of change.

H3: People's gender influences how motivating they find the text messages that are aligned to the processes of change.

However, because there is no existing work on evaluating these text messages representing the processes in relation to personality or gender, we have no specific expectations as to how personality and gender will influence the evaluation of the text messages and the text message categories.

5.3 Survey: evaluating motivational messages

To investigate how motivating the messages representing the processes of change would be rated, we designed an online survey study. The survey was carried out through Amazon Mechanical Turk[1] (AMT) on SurveyMonkey[2].

[1] https://requester.mturk.com/
[2] https://www.surveymonkey.com/

5.3.1 Participants

The sample consisted of 350 respondents. The study was conducted in English. All but 5 respondents were native English speakers, their data was not found anomalous and was kept.

The minimum age was 20 and the maximum was 71. The average age was 36.53 (SD $= 11.83$) and the median 34. With respect to education, 106 respondents received some college education, 142 had a college degree, 35 completed a masters, 59 completed high school, 5 obtained a PhD and 3 had other types of qualifications. To ensure consistency and a high quality of responses, the AMT requirements for the respondents were that they should have already completed $>$1000 tasks on AMT, that $>$98% of these tasks were approved successfully and that the participants were located in the United States. These requirements ensured that participants were already familiar with online surveys, that they were serious about filling in the survey and that they had some proficiency in English. In fact, 342 reported 'very good' as self-assessed proficiency and 8 'good', and none 'average', 'bad' or 'very bad'.

5.3.2 Task

In the survey, people evaluated 50 motivational text messages (five messages for each of the ten strategies, see Table B.1 in Appendix). The 50 motivational text messages were selected at the author's discretion, and selections were made based on how unambiguously the messages represented the process, yet also did not overlap in content too much between each other. We asked them to rate each message according to how motivating they thought the message was for them ("Please rate how motivating or demotivating you find the following messages for yourself.") with context to consider ("The intended context for the messages is that you would receive them in day to day life, the sender of the messages is not specific, but could be considered someone who wants the best for you as a person, like a friend or personal coach."). All the 50 messages were presented together on one page and the order of the messages was randomized for each participant. The messages were rated on a scale from 1 ("Very demotivating") to 5 ("Very motivating") with 3 as the neutral position ("Neither demotivating nor motivating").

5.3.3 Measures

At the start of the survey, we asked the participants about their gender, age, native language, understanding of the English language, education level, maternal education level (as an indication of socioeconomic status [Green, 1970]) and main field of work. After, we presented participants with the specific task of this survey: rating the 50 text messages. At the end of the survey, we asked participants information about when they would prefer to receive messages such as these, how many of such messages they would prefer to receive weekly, and whether they had experience with smartphones and exercise apps. Moreover, we measured people's susceptibility to persuasion [Kaptein et al., 2012], and their current stage of change[3] [Norman et al., 1998], perceived experiences with processes of change[3] [Nigg et al., 1999], self-efficacy[3] [Benisovich et al., 1998], and decisional balance[3] [Nigg et al., 1998] for exercise. To measure participants' personality we used the 50-item IPIP representation of the revised version of Costa and McCrae's [Costa and McCrae, 1992] NEO Personality Inventory[4] which posed 50 statements (e.g., "Make plans and stick to them."). Participants were asked to answer how descriptive they found these statements of themselves (on a 5-point Likert scale, 1 being "very inaccurate" and 5 being "very accurate").

5.3.4 Procedure

Participants were recruited through AMT. They were informed of their compensation (3 US dollars), the goal of the study (finding out which text messages are motivating), and the estimated time to complete the survey (35 minutes). Participants could then decide to accept or decline and proceed to the SurveyMonkey website were the survey was hosted. On the first page, the goal of this study was repeated and participants were asked to complete a consent form. On the second page, the participants were asked to fill in demographic information. On the third page, information about the messages and instructions on how to rate the messages was given. On the fourth page, context for rating the messages was provided and they were given 50 motivational messages to rate. On the pages five through sixteen, they were given multiple questionnaires. At the seventeenth and last page, participants were debriefed and given a completion code to fill in on AMT to receive payment.

5.4 Data analysis

We looked at whether the selection of our coded message categories were evaluated as internally consistent and if our measure of personality was reliable. Assuming that the coded messages we selected represented the processes well, and the coded messages we selected for each process (five for each process) were evaluated as internally consistent, this would mean that the coded messages we selected for each process were representative of our developed categories, which is shown in Table 5.2 through the reliability of the coded message categories. The reliability of the measures was overall very good. The only disputable measure was that of Counterconditioning, with a Cronbach's alpha of $.68$ which we still found acceptable (and also comparable to other relevant work [Marcus et al., 1992] where they had this score for Social Liberation). Otherwise, the reliability scores were between $.72$ and $.88$. The internal reliability of the personality traits was overall very good (see Table 5.1).

We also looked at whether the motivational messages were evaluated in the same way that the processes relate to the stages of change (in the context of exercise): the Experiential processes were expected to be rated as more motivating in the earlier stages of change, peak in the Action stage and decline in the Maintenance stage, while the ratings of the Behavioral processes were expected to also peak in the Action stage, but not decline in the Maintenance stage (H1). We carried out separate univariate

[3] http://web.uri.edu/cprc/measures/
[4] http://ipip.ori.org/

Trait	M	SD	α
Openness to experience	3.77	0.77	.84
Conscientiousness	3.83	0.77	.91
Extraversion	2.97	0.92	.92
Agreeableness	3.82	0.68	.86
Neuroticism	2.43	0.92	.92

Table 5.1: Averages (M), standard deviations (SD), and Cronbach's alpha's (α) for all the scales. Personality scales are 10 averaged items with scoring from 1 to 5. (N = 350)

Category	M	SD	α
Consciousness Raising	3.75	0.66	.76
Dramatic Relief	3.02	0.94	.82
Environmental Reevaluation	3.13	0.97	.88
Social Liberation	3.15	0.69	.73
Self-Reevaluation	3.67	0.69	.73
Self-Liberation	3.74	0.59	.72
Helping Relationships	3.69	0.67	.74
Counterconditioning	3.51	0.61	.68
Reinforcement Management	3.86	0.70	.85
Stimulus Control	3.21	0.65	.73

Table 5.2: Averages (M), standard deviations (SD), and Cronbach's alpha's (α) for all the scales. Scales are averaged. Process-categories scores are based on 5 messages with scoring from 1 to 5 averaged.

analyses of variance (ANOVA) with the stages as predictor variable and the process categories as separate outcome variables (i.e., CR, DR, ER, SOL, SR, SEL, HR, CC, RM and SC) to see if the process categories differed over the stages. To see between *which* stages these processes of change differ significantly, we performed *post hoc* tests with Bonferroni corrections.

To test whether personality traits; gender; the text message categories; the interaction between personality traits and the message categories; and the interaction between gender and the message categories influenced the ratings of the text messages (H2 and H3), we ran a linear mixed-effects model analysis in R [Team, 2016] with the lme4 package [Bates et al., 2015] and the lmerTest package [Kuznetsova et al., 2016] to calculate the significance of the differences. The contribution of the variables were assessed using the AIC and BIC scores, where a smaller number indicates a better fit. An analysis of variance comparison of the incrementally built models is reported in Table 5.3 (built with xtable package [Dahl, 2016]). We followed the AIC score because we selected a fair number of parameters for the model, and the AIC

	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)
0	4	45613.35	45644.43	-22802.67	45605.35			
..1	13	45577.51	45678.52	-22775.75	45551.51	53.84	9	**0.0000**
..2	18	45566.25	45706.11	-22765.13	45530.25	21.25	5	**0.0007**
..3	19	45558.90	45706.53	-22760.45	45520.90	9.35	1	**0.0022**
..4	64	45452.53	45949.80	-22662.26	45324.53	196.37	45	**0.0000**
..5	73	45143.27	45710.48	-22498.63	44997.27	327.26	9	**0.0000**

Table 5.3: An analysis of variance comparison of incrementally built models. Starting from the baseline model (0), consisting of Rating $\sim$ (1|RespondentID) + (1|Message), added effects are described in the table: Category (1), O C E A N (separate traits) (2), Gender (3), O C E A N:Category (interaction effect, 4), and Gender:Category (interaction effect, 5). Model fit reported as scores of AIC and BIC, smaller scores means better fit.

does not penalize the number of parameters as strongly as BIC [Burnham and Anderson, 2004]. Our final model with the smallest AIC score included the motivational rating of the text messages as the outcome, and the categories of the text messages (processes of change), personality traits, gender, the interaction between personality traits and the categories and the interaction between gender and the categories as fixed effects, and participants and messages as random effects.

5.5 Results

In this section, we present the results of the online evaluation survey for a selection of the text messages representing the ten processes of change (our behavior change strategies), and their relation to several user characteristics, namely the stage of change, personality and gender of the participant. Our hypothesis was that the stage the participant is in influences how motivating they find certain processes-of-change-aligned text messages (H1). We looked at how motivating people perceived the text messages across stages of change. We also hypothesized that there is a relation between people's personality (H2) and gender (H3) and processes-of-change-aligned text messages they find motivating. It is important to note that for this study, we measured the *self-assessed* stages of change, which were not equally distributed: 120 participants rated themselves to be in the Maintenance stage (M), 94 in the Preparation stage (P), 73 in the Contemplation stage (C), 46 in the Action stage (A), and 17 participants rated themselves in the Precontemplation stage (PC).

5.5.1 Relation between stages of change and processes-of-change message categories (H1)

The coded messages somewhat followed the expectation of the processes aligning to the stages in a similar way to Marcus et al. [1992] (H1), as is shown by the ratings given to the process categories for each stage of change (see Figure 5.1). We found that there are peaks for all Experiential and Behavioral processes in the Action stage

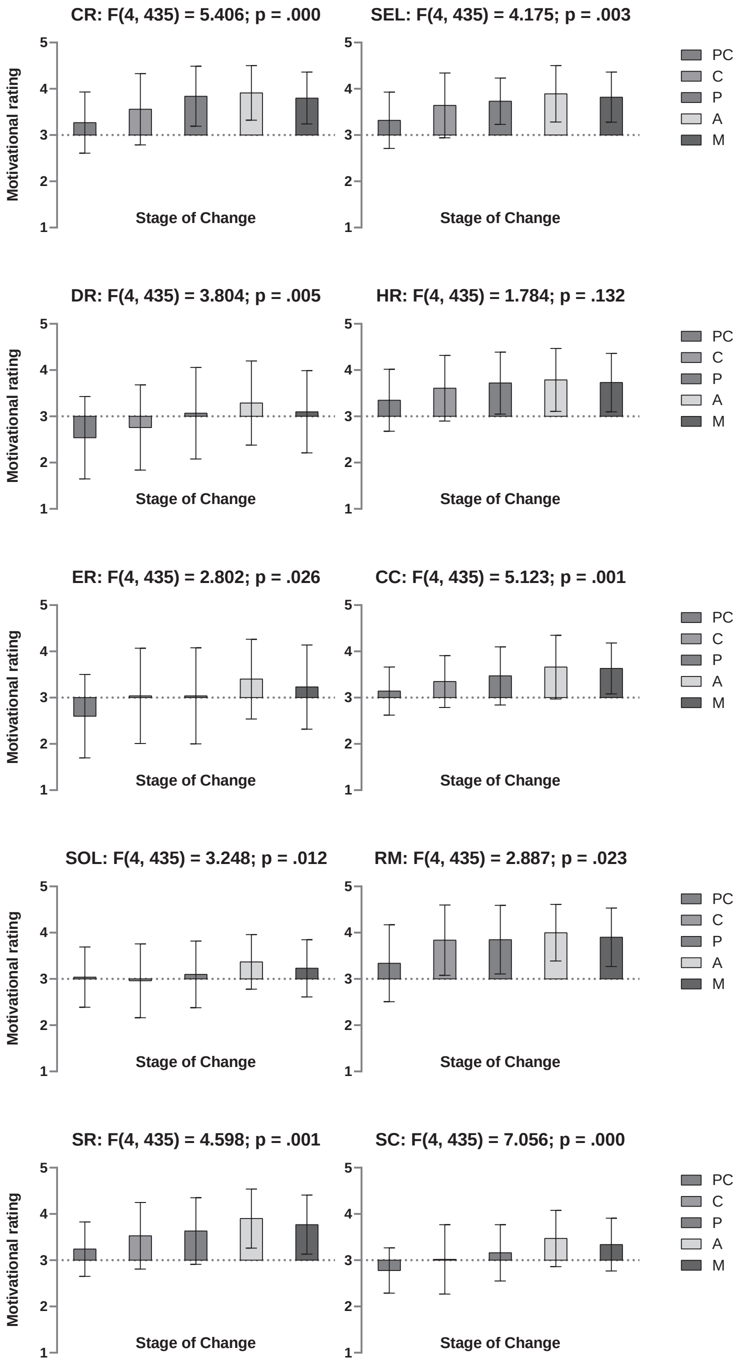

Figure 5.1: Bar charts with the averages (M), standard deviations (SD), F-statistics and p-values for all coded message categories. On the left, all Experiential process categories; on the right, all Behavioral process categories. Messages rated on a scale from 1 (“Very demotivating”) to 5 (“Very motivating”) with a 3 as neutral (“Neither demotivating nor motivating”). N = 350, PC = Precontemplation (N = 17), C = Contemplation (N = 73), P = Preparation (N = 94), A = Action (N = 46) and M = Maintenance (N = 120)

	Compared stages					
Cat.	$PC-P$	$PC-A$	$PC-M$	$C-P$	$C-A$	$C-M$
CR	$\rightarrow$	$\rightarrow$	$\Rightarrow$	$\Rightarrow$	$\Rightarrow$	
DR		$\Rightarrow$			$\Rightarrow$	
ER		$\Rightarrow$				
SOL					$\Rightarrow$	
SR		$\rightarrow$	$\Rightarrow$		$\Rightarrow$	
SEL		$\rightarrow$	$\rightarrow$			
CC		$\Rightarrow$	$\Rightarrow$		$\Rightarrow$	$\Rightarrow$
RM	$\Rightarrow$	$\rightarrow$	$\Rightarrow$			
SC		$\rightarrow$	$\rightarrow$		$\rightarrow$	$\rightarrow$

Table 5.4: *Post hoc* analyses for the coded messages categories and the stages. Only significant stage-to-stage *post hoc* test results are displayed. Arrows represent significant mean differences between stage $X - Y$, meaning there is a significant increase in average rating from stage X to Y. $\rightarrow$ represents $p < 0.05$, $\Rightarrow$ represents $p < 0.01$

and that the Behavioral processes are rated as more motivating in the later stages. But, the Experiential processes are not rated as more motivating in the earlier stages (Dramatic Relief and Environmental Reevaluation are even rated as demotivating in the earlier stages) and although there is a decline for Experiential processes in the Maintenance stage, the same decline is also found for the Behavioral processes.

As can be seen in Figure 5.1, all process categories differ significantly along the stages of change (all nine $p < .05$) except for the Helping Relationships category. This indicates that for these process categories, the scores differ significantly when compared between the stages of change. In Figure 5.1 we can observe the direction of the relations in the reported means for all the stages. The *post hoc* test results for the significant process categories (all but the Helping Relationships category) are shown in Table 5.4. As can be seen, it is mostly the Precontemplation stage that differs from the Action and Maintenance stage and the Contemplation stage that differs from the Action stage.

Overall, when looking at the average ratings over the stages it is clear that: the peak for all processes is in the Action stage, which is what we expected. There are no higher ratings for the Experiential processes than for the Behavioral processes in the earlier stages; in fact, two Experiential processes (Dramatic Relief and Environmental Reevaluation) have negative ratings in the Precontemplation stage, and the decline in the Maintenance stage is there for Experiential but also for Behavioral processes, which is both not what we had expected.

5.5.2 Results of the linear mixed-effects model analysis

For the linear mixed-effects model analysis, we report on a summary of the model (see Table 5.5, 5.6, and 5.7) with the Consciousness Raising (CR) category as reference

Random effects	variance	SD
RespondentID (Intercept)	0.235	0.485
Text Message (Intercept)	0.055	0.234
Residual	0.720	0.848

Table 5.5: The variance and standard deviation of the random effects of the model: participants and text messages. Number of observations: 17500, number of respondents: 350, number of text messages: 50.

level (benchmark to which to compare the scores of the other categories relatively). To make the intercept more interpretable, we centered the personality traits by subtracting the mean score of each trait for all scores in each trait. This means that the intercept score is now the score on the reference level (CR) while all other factors are zero, meaning a gender of zero (male) and personality scores of zero (meaning sample average).

5.5.2.1 Main effects

For the main effects, participants who scored higher on Openness scored higher on the motivational rating for the CR category (intercept reference level). Participants who scored higher on Agreeableness scored higher on the motivational rating for CR, and females scored significantly lower on the motivational rating for CR category (Table 5.6).

5.5.2.2 Interaction effects

For the interaction effects, we see, for example, that there was no statistical evidence that participants who scored higher or lower on Openness also scored higher or lower on the motivational rating for CC messages, compared to CR messages (reference level), but the participants who scored higher on Openness did score significantly lower on the motivational rating for all the other message categories compared to CR. Also, there was no statistical evidence that participants who scored higher or lower on Conscientiousness also scored higher or lower on the motivational rating of CR messages (from Table 5.6), but participants who scored higher on Conscientiousness did score significantly lower on the motivational rating for DR messages compared to CR messages (see Table 5.7 for these and more significant interaction effects).

5.5.3 Main effects with different reference levels

However, we are interested in the direct relation between all the categories and the personality traits, and all the categories and gender, not in the relations *in comparison to* the reference level category CR. The summary of the model does report on the relationship between categories and the personality traits and gender through the interaction estimates, because the estimate for a personality trait and a selected category other than the reference category is equal to the estimate of that trait for the reference category plus the estimate for the interaction between that trait and the

Fixed effects	Estimate	Std. Error	t-value
(Intercept)	**3.841**^{3}	0.113	33.925
DR	**-0.604**^{3}	0.153	-3.961
ER	**-0.434**^{2}	0.153	-2.846
SOL	**-0.655**^{3}	0.153	-4.293
SR	-0.061	0.153	-0.401
SEL	-0.152	0.153	-0.999
HR	-0.120	0.153	-0.786
CC	**-0.319**^{1}	0.153	-2.094
RM	-0.024	0.153	-0.157
SC	**-0.603**^{3}	0.153	-3.951
Openness	**0.149**^{2}	0.046	3.260
Conscientiousness	0.012	0.057	0.216
Extraversion	-0.027	0.046	-0.586
Agreeableness	**0.262**^{3}	0.063	4.127
Neuroticism	0.085	0.053	1.612
Gender (female)	**-0.219**^{2}	0.070	-3.141

Table 5.6: The estimates and standard error of the fixed effects of the model: process-categories, personality traits, and gender. Significant effects reported for $^{1}p < 0.05$, $^{2}p < 0.01$, $^{3}p < 0.001$.

selected category. However, this does not incorporate the significance of the relation between a trait and a category, only the significance of the relation between a trait and a category compared to the reference level category. To illustrate this: as stated in section 5.5.2.2, there was no statistical evidence that participants who scored higher or lower on Conscientiousness also scored higher or lower on the motivational rating of CR messages (Table 5.6), but participants who scored higher on Conscientiousness did score significantly lower on the motivational rating for DR messages compared to CR messages (Table 5.7). This could perpetrate the idea that there is a significant relation between scoring higher on Conscientiousness and scoring lower on DR messages. However, based on "post-hoc" tests (i.e., changing the reference level), participants who scored higher on Conscientiousness did not score significantly lower on the motivational rating of the DR message category (see column C, row DR in Table 5.8). Therefore, to see how significant the main effects of personality and gender were on the other nine categories (H2 and H3), we ran the same model again with changed reference levels (see Table 5.8). Note that the first row of Table 5.8 reports identical estimates to estimates reported for the personality traits and gender in Table 5.6, where we also report the standard error.

Fixed eff.	Estimate	Std. Error	t-value	Fixed eff.	Estimate	Std. Error	t-value
(Intercept)	**3.841**^{3}	0.113	33.925				
DR:O	**-0.181**^{3}	0.040	-4.542	DR:C	**-0.101**^{1}	0.050	-2.026
ER:O	**-0.253**^{3}	0.040	-6.358	ER:C	0.018	0.050	0.359
SOL:O	**-0.128**^{2}	0.040	-3.224	SOL:C	-0.087	0.050	-1.741
SR:O	**-0.137**^{3}	0.040	-3.453	SR:C	-0.014	0.050	-0.280
SEL:O	**-0.138**^{3}	0.040	-3.460	SEL:C	-0.043	0.050	-0.858
HR:O	**-0.128**^{2}	0.040	-3.233	HR:C	0.038	0.050	-0.766
CC:O	-0.052	0.040	-1.316	CC:C	-0.001	0.050	-0.024
RM:O	**-0.087**^{1}	0.040	-2.188	RM:C	-0.054	0.050	1.080
SC:O	**-0.174**^{3}	0.040	-4.375	SC:C	-0.072	0.050	-1.442
DR:E	0.034	0.040	0.847	DR:A	**-0.184**^{3}	0.055	-3.324
ER:E	0.043	0.040	1.083	ER:A	0.007	0.055	0.134
SOL:E	**0.155**^{3}	0.040	3.891	SOL:A	**-0.153**^{2}	0.055	-2.762
SR:E	0.073	0.040	-0.401	SR:A	-0.066	0.055	-1.195
SEL:E	**0.115**^{2}	0.040	2.882	SEL:A	**-0.111**^{1}	0.055	-2.002
HR:E	0.001	0.040	0.031	HR:A	-0.106	0.055	-1.906
CC:E	**0.082**^{1}	0.040	2.065	CC:A	**-0.184**^{3}	0.055	-3.316
RM:E	**0.101**^{1}	0.040	2.548	RM:A	**-0.140**^{1}	0.055	-2.530
SC:E	**0.149**^{3}	0.040	3.731	SC:A	**-0.115**^{1}	0.055	-2.068
DR:N	**-0.148**^{2}	0.046	-3.230	DR:G (♀)	**-0.319**^{3}	0.061	-5.260
ER:N	**-0.165**^{3}	0.046	-3.588	ER:G (♀)	**-0.452**^{3}	0.061	-7.447
SOL:N	**-0.108**^{1}	0.046	-2.365	SOL:G (♀)	**-0.121**^{1}	0.061	1.996
SR:N	-0.082	0.046	-1.791	SR:G (♀)	-0.045	0.061	-0.735
SEL:N	-0.073	0.046	-1.593	SEL:G (♀)	**0.354**^{3}	0.061	5.833
HR:N	**-0.105**^{1}	0.046	-2.296	HR:G (♀)	**0.144**^{1}	0.061	2.381
CC:N	-0.066	0.046	-1.446	CC:G (♀)	**0.183**^{2}	0.061	3.024
RM:N	-0.076	0.046	-1.656	RM:G (♀)	**0.315**^{3}	0.061	5.196
SC:N	**-0.114**^{1}	0.046	-2.478	SC:G (♀)	**0.161**^{2}	0.061	2.652

Table 5.7: The estimates and standard error of the fixed interaction effects of the model: process-categories, personality traits, and gender. Significant effects reported for $^{1}p < 0.05$, $^{2}p < 0.01$, $^{3}p < 0.001$.

Cat.	O est.	C est.	E est.	A est.	N est.	G est.
CR	**0.149**[2]	0.012	-0.027	**0.262**[3]	0.085	**-0.219**[2]
DR	-0.032	-0.088	0.007	0.078	-0.063	**-0.537**[3]
ER	**-0.104**[1]	0.030	0.016	**0.270**[3]	-0.080	**-0.670**[3]
SOL	0.021	-0.074	**0.128**[2]	0.109	-0.024	-0.098
SR	0.011	-0.002	0.046	**0.196**[2]	0.003	**-0.263**[3]
SEL	0.011	-0.030	0.088	**0.151**[1]	0.012	0.135
HR	0.020	0.050	-0.026	**0.157**[1]	-0.020	-0.074
CC	**0.096**[1]	0.011	0.055	0.079	0.019	-0.035
RM	0.062	0.066	0.075	0.122	0.009	0.096
SC	-0.025	-0.059	**0.121**[2]	**0.147**[1]	-0.029	-0.058

Table 5.8: The estimates of the main effects of the model with category levels as reference levels and personality (O, C, E, A, and N), and gender (G). Significant effects reported for $^1p < 0.05$, $^2p < 0.01$, $^3p < 0.001$.

5.5.3.1 Relation between personality and processes-of-change message categories (H2)

We investigated the relations between the personality traits variables (O, C, E, A, and N) and the process-categories on the motivational rating of the text messages (H2). We conducted “post-hoc” tests (i.e., changing the reference level) because it is not possible to determine from the basic lmer test summary whether significant relations corresponding to interaction effects are simply null effects or effects in the opposite direction to the main effect. In Table 5.8, the estimate scores are reported for the personality traits (OCEAN) and the process-categories (10 times as reference category). The p-values suggest that the personality traits O, E and A significantly relate to eight of the process-categories. There was a significant positive relation between Openness and the rating of CR and CC, and a significant negative relation between Openness and ER. There was a significant positive relation between Extraversion and the rating of SOL and SC. And, there was a significant positive relation between Agreeableness and the rating of CR, ER, SR, SEL, HR, and SC. This shows that the use of these eight process-categories (CR, ER, SOL, SR, SEL, HR, CC, and SC) could be adapted to how people score on the personality traits of O, E and A.

5.5.3.2 Relation between gender and processes-of-change message categories (H3)

We investigated differences in ratings of text message categories between males and females (H3) (see Table 5.8). The reference category for gender is male. CR, DR, ER, and SR messages were more motivating to male participants than female participants, because the estimate is significantly negative. This shows that the use of these four Experiential process-categories (CR, DR, ER, and SR) could be adapted to which gender people identify themselves as (male or female).

5.6 Discussion

In this section, we discuss how people evaluated motivational messages with the results of the analyses, considerations for behavior change technology design, and the limitations of our current work. The selection of coded messages to represent the process categories showed very good reliability. Therefore, the five messages we selected for each category were a good fit for the process of change they represented. Overall, using crowdsourcing to design motivational messages for behavior change seems practical, easy, and sufficient.

5.6.1 Relation between stages of change and processes-of-change message categories (H1)

We also found that the motivational rating of coded messages of the participants did not entirely match our expectation of processes over stages (H1). Specifically, that the messages representing the Experiential processes were *not* rated as more motivating in the earlier stages of change, although they did peak in the Action stage. And although the messages representing the Behavioral processes were rated as more motivating in the later stages of change and peaked in the Action stage, they did decline in the Maintenance stage. To investigate further we looked at where the significant differences between the stages were for all the processes. It is important to note that we did not correct for the alpha-inflation of our multiple (10) tests, so the barely significant relations (such as ER, SOL, and RM) should be interpreted with caution.

Even though the Experiential processes were *not* rated more motivating than the Behavioral processes in the earlier stages of change and the Behavioral processes *did* decline in the Maintenance stage, the broader trend for Experiential and Behavioral processes *was* found (both peak in the Action stage). This general trend is also shown in the *post hoc* test results (see Table 5.4), where four of the five Experiential processes (not Environmental Reevaluation) show significant differences between the Action stage and the Precontemplation stage, and between the Action stage and the Contemplation stage. Moreover, the *post hoc* test results on Experiential processes showed twelve significant mean differences and eight of those were in relation to the Action stage. The Behavioral peak was also found in the Action stage. The *post hoc* test results for the four significant Behavioral processes (not Helping Relationships), showed differences between the Action stage and the Precontemplation and Contemplation stage, but also between the Maintenance stage and the Precontemplation and Contemplation stage. The latter could be an indication that, even though there is a decline in the Maintenance stage for the Behavioral processes, overall, the Behavioral processes are still relevant in the Maintenance stage.

It is important to note that for this study, it was possible for people to rate messages on the demotivating end of the spectrum. We did this on purpose, to see if there might be messages we definitely should avoid for certain stages. The results for Dramatic Relief, Environmental Reevaluation and Stimulus Control actually show that they were rated as demotivating in the Precontemplation stage, and Dramatic Relief also in the Contemplation stage (as can be seen in Figure 5.1). In fact, none of the processes are rated as highly motivating for the early stages. One reason may be that unmotivated people do not find the processes they should go through to change motivating, even though it is still required for behavior change. Another reason may be that they are unmotivated *because* they do not easily find something motivating.

5.6.2 Relation between personality and processes-of-change message categories (H2)

We found a number of significant relations between the ratings of the text message process-categories and personality traits (H2), and in this section we discuss how to interpret and use these results. Specifically, we found significant relations between process-categories CR, ER (inverse relation), and CC and the personality trait Openness, significant relations between process-categories SOL and SC and the personality trait Extraversion, and significant relations between process-categories CR, ER, SR, SEL, HR, and SC and the personality trait Agreeableness. These significant relations indicate that personality plays a role in how motivational the text messages are perceived. In other words, people scoring high on Openness found CR text messages to be more motivating, ER text messages to be less motivating and CC text messages to be more motivating. People scoring high on Extraversion found SOL and SC text messages to be more motivating, and people scoring high on Agreeableness found CR, ER, SR, SEL, HR, and SC text messages to be more motivating. These personality trait findings indicate that, when choosing motivational strategies or interventions for behavior change technology for physical activity, it is important to take into account the personality of the user, because strategies can be perceived differently than expected, potentially leading to counterintuitive results.

Overall, Agreeableness was related to the most processes (six in total). This is somewhat similar to the results of Alkış and Temizel [2015], where they found Agreeableness to be the most susceptible to their six persuasion strategies (i.e., authority, reciprocation, scarcity, liking, commitment and consensus). On the other hand, they found Openness to Experience to be the least susceptible to these strategies, which is not reflected in our results. Their strategies are also not identical to our processes of change text messages strategies. To the best of our knowledge, this is the first effort to translate the processes of change to operationalized text messages, therefore exact comparison is not possible.

It is interesting to note that these results are different from our previous chapter (Chapter 4), where we investigated the relation between personality and the processes of change *through a questionnaire measuring the everyday experiences with the processes*, in contrast to our current work where we investigate the relation between personality and the processes of change *through rating text messages representing the processes on how motivating the messages are*. Although the factors we incorporated are not exactly equal, making comparison difficult, in our previous work, we found mostly different relations between the traits and the processes, and Conscientiousness was related to the most processes (six in total).

An explanation could be that our text messages representing the processes are not good representations of the complex processes of change and that is why the results do not match up to the processes of change through a questionnaire. Representing the processes of change through 'simple' text messages may not cover the complexity of a process. But, with our rigorous coding following the guidelines of Guest and MacQueen Guest and MacQueen [2007]; MacQueen et al. [1998], who also used this to develop a codebook for the processes of change from the TTM [Guest and MacQueen, 2007, p. 120], and with decent reliability of the categories (see Table 5.2), we have tried to mitigate this.

Another explanation for the difference between our current results and our previous results could be that the processes of change questionnaire asks for the participants recent experience with this process (e.g., "I look for information related to exercise."; 1: Never – 5: Repeatedly). This alludes to how relevant the process is or could be for the participant (consciously or subconsciously). This relevance (or irrelevance) in turn could suggest the experimenter that this process could be a good way to try to stimulate the participant. On the other hand, our evaluation of the process of change messages asks the participant to rate how motivating they think this message is for their current situation (e.g., "Exercise will help clear your mind and reduce stress."; 1: Very demotivating – 5: Very motivating), which pushes them to think very consciously about how motivated they would be by this message if they received it. In that sense, it is interesting that in our previous results Conscientiousness (often associated with following norms and rules, planning, organizing [John et al., 2008]) played a greater role in terms of correlation to the experience of processes (through the questionnaire), while in this research it is Agreeableness (often associated with being more cooperative, compliant, trusting [John et al., 2008]) that plays a greater role in terms of correlation to how motivating the text messages processes are rated. The difference being that one does not have to be motivated by a process to experience it or consider it valuable to experience. You can schedule your exercise time because it is a valuable way to get yourself to exercise, but you do not have to find it motivating. Similarly, you can find messages on scheduling workouts motivating because they are valuable in getting (or reminding) yourself to exercise, while you are not (too) conscious about scheduling exercise yourself.

5.6.3 Relation between gender and processes-of-change message categories (H3)

Of the ten categories, we found four to be significantly related with gender (H3). All four of these significant differences indicated that male participants found the text messages of these process-categories more motivating (i.e., CR, DR, ER, and SR) than the female participants. It is interesting to note that these were all Experiential processes, meaning that these processes focus on influencing the experiences related to the physical activity behavior change. Overall, these gender findings indicate that when choosing motivational strategies or interventions for behavior change technology for physical activity, it is also important to take into account the gender of the user, because strategies can be perceived differently than expected, specifically strategies that try to motivate people by appealing or referring to experiences.

In comparison to related work, the results are not in line with previous research. For example, Kristan and Suffoletto [2015] found that overall women responded most favorably to all their feedback messages (divided into informational, motivational or strategy facilitating). Busch et al. [2016] found males to be more motivated by comparison and competition strategies, but unfortunately comparison and competition strategies do not relate directly to any of the Experiential strategies we found more motivating for males.

Overall, the differences found in our results suggest that it is important to take into account the gender of the person engaging in behavior change, especially in regard to the (Experiential) process-categories that usually are associated with the earlier stages of behavior change, but we should be careful in generalizing these results (also those related to the personality traits) further than our study, due to the lack of other research with similar findings. In terms of findings for our system, it means that the use of these four process-categories (CR, DR, ER, and SR) could be adapted to the gender people identify themselves with, especially for the process-categories of DR and ER, which are evaluated *demotivating* by females, but slightly *motivating* by males. The current results are obtained with the same messages, and the messages are in the same form, as how we intend to use them for our technology and therefore the relations we observed can be the foundation for tailoring system behavior.

5.6.4 System design considerations

We interpret the results as an indication of the need for tailoring (to personality and gender, and somewhat to stage), but, in the context of how long behavior change can take, also of the need for having *multiple* strategies to tailor to the user. From a different perspective, the results could also be interpreted as needing only a *few* strategies. If we take into account only how motivating people rated the strategies, we could choose (or *customize* Kaptein [2015]) a strategy for the user based solely on what scored the highest on rated motivation for the user's features (which implies the highest 'return' for a user's behavior change). However, considering the long-term aspect and the complexity of behavior change (in contrast to a one-time act of compliance), we do not think that the most optimal solution to achieve behavior change is to focus on one (highest rated) strategy. Instead, we would suggest a combination [Craig et al., 2008] of highest rated strategies or interventions (for the user's features), to avoid repetition over longer periods of time (in combination with plenty of variation of messages within a strategy), but also to maximize the effectiveness of the message that behavior change has benefits for that user [Dusseldorp et al., 2014]. This could still be modeled, as suggested by Kaptein [2015], by making the model more complex, even without focusing on only the highest rated strategies, but also on which combination of strategies results in the optimal outcome (the most 'permanent' behavior change). However, not all (combinations of) strategies or interventions might be or are appreciated by the user, which could result in the technology not being used and the intervention not taking place. For example, Horsch et al. [2016] found that the implemented reminder strategy, in the context of sleep behavior, did not work for people who had a negative attitude towards their intended activity. In that sense, the measurement of how motivating these messages are perceived is not only input for what strategies to definitely use, but also for what strategies to avoid for which users to increase the likelihood of this system being used over longer periods of time.

Another point that could be made about investigating all the different factors that might relate to the strategies a system designer wants to use to motivate a behavior, is to use an adaptive system. An adaptive system would not need any prior input on how the factors relate, and could find the most optimal model for the user by testing all the strategies. However, for the purpose of our system, we feel that this does not fit our research approach in two ways. First, in an adaptive system with a wide variety of possible strategies and no prior input, the users will, on average, be exposed to a number of strategies that might not be motivating or even demotivating them (one of the results of the evaluation), resulting in a higher chance of abandoning the technology. Second, starting from an adaptive system with no prior input might optimize the model in an more effective way compared to a predetermined static model, but it will not help explain or interpret why the model works. For example, we could find that a certain group of messages is highly effective for some user features, but if these messages do not explicitly group on a certain theme or underlying construct it will be hard to interpret and replicate the results. Nevertheless, we feel that eventually it would be preferable for every system to be, at least partially, adaptable, to better accommodate for changes in users that might not have been modeled beforehand (e.g., relapse). Especially when the adaptive system can be effectively combined with prior input from an already (moderately) effective model.

With this study we aimed to investigate how people evaluate motivational messages designed by the crowd based on their stage of change, personality and gender (RQ2). The findings allow us to inform the development of our own behavior change technology, and might help inform the design of motivational strategies, interventions, or behavior change technology in general. Our main result is that the stage of change the participant was in did not completely match our expectation of how this would influence the motivational rating of the process-categories, and that we found significant relations between process-categories and personality traits and between process-categories and gender. The significant relations indicate that based on different personalities, people prefer different motivational strategies. Hence, the use of these eight process-categories (all except DR and RM) could be tailored to how people score on the personality traits of Openness, Extraversion and Agreeableness. We also found four significant inverse relations between gender (female) and the process-categories CR, DR, ER and SR, meaning that men found these four process-categories to be more motivating than women. This suggests that the gender of the person being motivated to change their behavior plays a significant role, especially in regard to the process-categories that are usually associated with the earlier stages of behavior change. Taking these results into account could improve the effectiveness of our motivational strategies and motivational strategies in general.

5.6.5 Limitations of the current work

There were some limitations to the present study. We gathered our respondents through Amazon Mechanical Turk, which could mean a misrepresentation of the general adult population of a Western society, although studies have reported that AMT generally has quite a good representation for online survey standards [Mason and Suri, 2012]. However, we are aware that there can be limitations for how representative the sample is. Also, because we used a cross-sectional design our results do not provide evidence for causation, only correlation.

5.7 Conclusion

We evaluated a subset of the previously collected peer-designed processes-of-change-aligned motivational text messages to see whether they represent the intended process categories, whether people in different stages of change perceived messages tailored to their stage of change as more motivating, and whether people evaluated the messages differently based on their personality and gender. Our main results are that 1) the way people rate the messages (process categories) on how motivating they are does not always match the expectation of what processes should be most relevant for the stage of change they are in; and 2) that there exist significant relations between process-categories (CR, ER, SOL, SR, SEL, HR, CC, and SC) and personality traits (Openness, Extraversion, and Agreeableness) and between process-categories (CR, DR, ER, and SR) and gender.

Overall, the findings suggest that, when choosing strategies to use in motivational technology or coaching systems it is important to consider the individual differences of the users, like stage, personality and gender, how this influences their preferences for strategies, and to design systems that can tailor to these differences. From an empirical perspective, the findings provide a contribution with the relations found that indicate that personality as well as gender play a role in how motivational the text messages are perceived, especially in regard to the process-categories that are usually associated with the more early stages of behavior change. Moreover, we found that how motivating people perceive the processes-of-change-aligned text messages in context of the stage of change they are in, is different from how the processes are theoretically aligned per stage. From a theoretical perspective, our findings provide a contribution because they indicate that the saliency and impact of theory-based strategies could be improved by tailoring them to personal characteristics, like personality and gender. However, when looking at our results for the stages of change, it seems that all the messages were predominantly rated as motivating in the later stages of behavior change, but not in the earlier stages, including those strategies intended for the earlier stages. This could mean that which processes to use is not so dependent on the stages of change. Another explanation could be that the way the messages were designed did not appeal to people in the earlier stages of change. To see if we can find an explanation for these unexpected results, our next step (see Chapter 6) is to follow our approach again, but this time with experts.

6 | Eliciting, categorizing and evaluating expert-designed motivational messages and comparing peer- and expert-designed motivational messages

In the previous chapters we introduced our approach to operationalizing and coding the processes of change from the Transtheoretical Model as text messages by using crowdsourcing to have peers design these messages because they are expected to be able to design theoretically sound interventions. In this chapter we use the same approach, but this time with experts as the designer of the messages. Similar to Chapter 5, we evaluate a selection of five of these expert-designed text messages for each of the ten processes of change categories, and compare the results of the evaluation of these expert-designed messages to the evaluation of the peer-designed messages. Furthermore, we explore whether the selections of the expert-designed messages and peer-designed messages are different along several linguistic dimensions.

This chapter[1,2] is based on one publication:

de Vries, R. A. J., Zaga, C., Bayer, F., Drossaert, C. H. C., Truong, K. P., & Evers, V. Experts Get Me Started, Peers Keep Me Going: Comparing Crowd- versus Expert-Designed Motivational Text Messages for Exercise Behavior Change. *In Proceedings of the EAI International Conference on Pervasive Computing Technologies for Healthcare.* (pp. 155-162). ACM. [de Vries et al., 2017b]

6.1 Introduction

In previously discussed related works on tailored text messages to influence someone's behavior, it is usually the authors or (other) experts who translate the theory-based strategies into messages. In our previous chapters (Chapter 4 and 5), however, we presented studies involving crowdsourcing to design and evaluate motivational text messages for physical activity, and we showed that these peer-designed text messages

[1] We thank Franciszka Bayer for her efforts in collecting expert-designed motivational messages and her efforts in coding the expert-designed messsages.

[2] We thank Nicola Marcon for his efforts in doing an linguistic analysis and word count on the peer and expert-designed messages.

aligned to behavior change strategies from theory. To reiterate, we chose to have peers design the motivational messages because of e.g., the work of Coley et al. [2013] and Kristan and Suffoletto [2015], where they show that peer-designed text messages are more engaging and more relevant to the user compared to expert-designed text messages. Therefore, we expected that peer-designed messages would work well as motivational messages for people with all different physical activity levels. However, in our work, contrary to our expectations, the messages were predominantly rated as motivating by people in the later stages of behavior change, but not by people in the earlier stages, including those strategies intended for the earlier stages. We speculate that the peers —and therefore the messages they designed —did not have sufficient expertise to motivate people in earlier stages. It is reasonable to assume that experts will have this expertise. In this chapter, we present studies we designed to replicate the approach with experts.

We make the following contributions: 1) we show that, through coding and the use of a codebook, also these expert-designed motivational messages can be aligned to the processes of change; 2) from the elicitation survey, we find that the experts come up with more Experiential-themed messages compared to peers, while peers come up with more Behavioral-themed messages compared to experts; and 3) we find that how motivating people perceive the processes-of-change-aligned expert or peer-designed text messages in context of the stage of change they are in, is different from how we expected: two of the process-categories with expert-designed messages were found more motivating in the earliest stage, while for several of the strategies peer-designed messages were rated more motivating for later stages.

The chapter is outlined as follows: first we discuss the hypotheses (section 6.2), the method (section 6.3), data analysis (section 6.4), and the results of the analyses for the first survey (section 6.5). Next we discuss the method (section 6.6), data analysis (section 6.7), and results of the analyses for the second survey, and we compare the evaluation of the expert-designed messages with the evaluation of peer-designed messages (section 6.8). We end the chapter with a discussion (section 6.9) and a conclusion (section 6.10).

6.2 Hypothesis

In previous chapters we presented our approach to operationalizing theory-based motivational strategies by designing and coding the processes of change from the Transtheoretical Model as text messages. Through an evaluation survey we found that all these peer-designed messages were predominantly rated as motivating in the later stages of behavior change, but not in the earlier stages, including those strategies intended for the earlier stages. We argue that the peers from the crowdsourcing did not have the expertise needed to motivate people across all the five stage of change. We expect that experts will have the expertise needed to motivate people across all the five stages of change, and that this will result in the messages designed by experts being perceived as more motivating. Therefore, we decided to use the same approach and evaluation we previously used with peers, but now with experts. This is what the third research question for this dissertation (see section 1.2) focuses on.

RQ3: To what extent does the expertise of the designer of the intervention's motivational content influence how motivating the intervention is perceived?

We investigate whether text messages designed by experts will be perceived as more motivating, in the earlier (more difficult) stages of behavior change as well as in the later (relatively easier) stages. We evaluate how people rate the expert-designed motivational messages representing the ten processes, and compare this to the data from our previous study involving peers. We expect expert-designed messages to be perceived as more motivating than peer-designed messages across all the five stages of change. We hypothesize that:

H1: People perceive the expert-designed messages representing the processes of change as more motivating than peer-designed messages representing the processes of change, regardless of their stage of change.

The following section explains how we addressed operationalizing the theory-based motivational strategies using same approach and evaluation we previously used with peers, but now with experts.

6.3 Data collection: expert-designed motivational messages

We followed the approach of the previous chapters (Chapter 4 and 5) aimed at eliciting peer-designed motivational messages. Although peer-designed messages can be motivating, we expect that experts can design more motivating messages based on their expertise. This expertise is also what makes experts in a specific area relatively scarce, and therefore we plan to collect a smaller dataset of messages, but still sufficient to compare to the peer-designed messages. First we designed the (online and paper-based) elicitation survey.

Similar to the previous chapter (Chapter 5), the first study was set up to enable participants to the design the motivational text messages. However, in this version we asked people who are experts in the field of behavior change to design these motivational messages. The experts we approached are people working in the field of health care (such as fitness or behavioral coaches), people working in public health programs, and people teaching and conducting research in the field of health psychology.

6.3.1 Participants

In total 25 people responded, of which 17 were female and 8 were male. The age varied between 21 and 62 years with a average of 33.29 (SD = 11.22). Of these 25, 10 identified themselves as fitness coaches, 8 as behavioral coaches and 7 as researchers in health psychology. 16 were of Dutch nationality and 9 were of German nationality.

Stage of change scenario and three expert-designed text messages
Precontemplation: "Consider a middle-aged person, with a steady personal life and solid friend foundation. This person lacks regular exercise in his/her daily life and is unwilling to consider starting with this, at least not within the next 6 months."
Consciousness Raising: "Regular exercise will make you less vulnerable for diseases."
Dramatic relief: "The longer you push exercise aside, the longer you are denying yourself a better quality of life."
Social Liberation: "Look around in your neighbourhood if you can find any facilities to do some sports."

Table 6.1: Example of a scenario used in the elicitation survey and examples of experts' responses coded as processes of change.

6.3.2 Task

In the previous survey used for crowdsourcing peer-designed motivational messages, we asked each participant to design six motivational messages for one scenario out of a possible five scenarios (see Table 6.1 for an example of one of the five scenarios and for an example of designed messages), where each scenario is based on one of the five stages of change. In this version of the survey we kept all of the scenarios the same, but due to a lower number of participants, we asked participants to come up with motivational messages for three randomly selected scenarios. To avoid repetition of designed messages, we asked for at least one and at maximum six motivational messages per scenario.

6.3.3 Measures

At the start of the survey, we asked the participants about their gender, age, nationality, education level, main work field in Health Psychology, and how long they have been working in this field. After, we presented participants with three random scenarios, each of which they could design up to six motivational messages for. At the end of the survey, we asked participants if they were familiar with the TTM.

6.3.4 Procedure

Participants were recruited through email or approached in person. First, we introduced our research team, explained the goal of the survey and gave the estimated time to complete the survey. If the participant was willing to participate they could go to the SurveyMonkey website where the survey was hosted or fill in a paper version. At the start of the survey the participants were welcomed and a summary was given of the purpose of the study. The survey was written in English, but the participants were informed of the possibility to respond in either English, or Dutch or German if necessary. At that point, the participants were asked to complete the consent form. The participants were then presented with the demographic questions and the language-elicitation task. Afterwards, participants were debriefed about the goals of the survey. Participants were not compensated with anything of monetary value.

6.4 Data analysis: elicitation survey

In total, 377 messages were designed. To see how these messages designed by experts reflected the processes of change, the same approach and the same codebook was used as described in the previous chapter (Chapter 4). This codebook was developed following the guidelines of Guest and MacQueen [2007]. The codebook describes definitions and guidelines for ten categories, each representing one of the ten processes of change. Two coders coded the messages and each message could only be coded with one coding category. Next to the coding of one of the ten categories, an additional 'certainty' code was used to indicate that the coder that coded the message was 99% sure that the message belonged to that particular coding category and that the other coder would agree on this. Two coders iteratively coded the messages and reached a Cohen's kappa of 0.62, which was deemed a substantial agreement [Landis and Koch, 1977]. The messages that were coded by coder 1 (same coder as coder 1 in the previous study) scored a Cohen's kappa of 0.80 for the certainty measure. The remaining messages were coded by coder 1.

For the data analysis, a Chi-square test was carried out to see if the scenarios (describing the stages) had an effect on the counts within the higher-order processes. To evaluate how the distribution of 377 expert messages over the processes was proportionately different from the distribution of the 2886 peer messages over the processes from our previous work, we performed a Chi-square goodness-of-fit test.

6.5 Results: elicitation survey

The 377 messages that experts designed were divided between the higher-order experiential and behavioral processes and between the ten different processes. In total, experts came up with 147 experiential messages (39%) and 230 behavioral messages (61%). The results (see Table 6.2) show that there is a significant association between the stages and higher-order processes ($\chi^2(4) = 131.733, p < .001$). The values of the standardized residuals are used to further interpret the results.

A positive value indicates that there were more messages of that process than expected and a negative value points to less messages than expected ($p < .05$ for a z-value higher than 1.96 or lower than -1.96 [Field, 2013]). The residuals show that for the Experiential processes in the Precontemplation ($z = 5.4$) and Preparation stage ($z = 3.1$), there is a significant overrepresentation of the processes, and in the Action ($z = -5.3$) and Maintenance ($z = -3.5$) stage there is a significant underrepresentation of the processes. For the Behavioral processes this is reversed, in the Precontemplation ($z = -4.3$) and Preparation stage ($z = -2.5$), there is a significant underrepresentation of the processes, and in the Action ($z = 4.3$) and Maintenance stage ($z = 2.8$) there is a significant overrepresentation of the processes. This means that there are more Experiential messages in the earlier stages and fewer in the later stages than the Chi-square model predicts and there are fewer Behavioral messages in the earlier stages and more in the later stages than the Chi-square model predicts. Unfortunately, there were not enough expert-designed messages to also investigate the lower-order (the ten processes) distribution over the stages.

Categories/Stage scenarios		**PC-S**	**C-S**	**P-S**	**A-S**	**M-S**	Total
Experiential	Count	53	29	52	5	8	147
	Expected	25.7	23.8	33.9	37.8	25.7	147
	Std. residual	**5.4**^{3}	1.1	**3.1**^{2}	**-5.3**^{3}	**-3.5**^{3}	
Behavioral	Count	13	32	35	92	58	230
	Expected	40.3	37.2	53.1	59.2	40.3	230
	Std. residual	**-4.3**^{3}	-0.9	**-2.5**^{1}	**4.3**^{3}	**2.8**^{2}	
Total	Count	66	61	87	97	66	377

Table 6.2: The distribution of all the codes over the two higher-order process categories (experiential and behavioral) and five stages of change scenarios. $^{1}p < .05$, $^{2}p < .01$, $^{3}p < .001$

We also report how the distribution of 377 expert messages over the processes was proportionately different from the distribution of the 2886 peer messages over the processes from our previous work. The results (see Table 6.3) show that there is a significant difference ($\chi^2(9) = 143.026, p < .001$) between the proportional distributions of the expert messages and the peer messages. Again, the values of the standardized residuals are used to further interpret the results of the Chi-square test.

The residuals show that for the higher order construct of Experiential processes ($z = 4.16$) there are proportionately more expert messages than peer messages related to those processes, and for the higher order construct of Behavioral processes ($z = -2.57$) there are proportionately less expert messages than peer messages related to those processes (see Table 6.3). Four processes differ significantly, namely Consciousness Raising ($z = 3.27$), Social Liberation ($z = 10.21$), Self-reevaluation ($z = 2.66$) and Self-liberation ($z = -3.95$) (also Table 6.3) This means that there were proportionately more expert messages than peer messages related to the processes Consciousness Raising, Social Liberation and Self-reevaluation, and proportionately less expert messages than peer messages related to the process Self-liberation.

In summary, as can be seen in Table 6.3, experts provide proportionately more messages than peers for Experiential processes (CR, SOL, SR), which are related to the earlier stages, and significantly less for Behavioral processes (SEL), which are related to later stages of change.

6.6 Survey: evaluating motivational messages

Similar to the previous chapter (Chapter 5), the second study was set up to investigate how motivating the messages representing the processes of change would be rated. We designed an online survey study, but this time to evaluate expert-designed motivational messages. The survey was again carried out through Amazon Mechanical Turk[3] (AMT) on SurveyMonkey[4].

[3]https://requester.mturk.com/
[4]https://www.surveymonkey.com/

Processes of change	peers (n = 2886) n (obs)	%	experts (n = 377) n (exp)	n (obs)	%	std. res.
Exp.	800	27.7	104.5	147	39.0	**4.16**
CR	138	4.8	18.1	32	8.5	**3.27**
DR	59	2.0	7.6	7	1.9	-0.22
ER	79	2.7	10.2	5	1.3	-1.63
SOL	11	0.4	1.5	14	3.7	**10.21**
SR	513	17.8	67.2	89	23.6	**2.66**
Beh.	2086	72.3	272.5	230	61.0	**-2.57**
SEL	939	32.5	122.8	79	21.0	**-3.95**
HR	238	8.2	31.0	38	10.1	1.26
CC	289	10.0	37.8	32	8.5	-0.94
RM	512	17.7	66.9	63	16.7	-0.48
SC	108	3.7	14.0	18	4.8	1.07

Table 6.3: Observed frequency (n (obs)) and percentages are reported for peer and expert messages, and also expected frequency (n (exp)) for expert message categories. Reported standardized residuals represent the difference between the observed and the expected expert frequency. Numbers in bold are $p < .05$.

6.6.1 Participants

In total we had 341 respondents of which 171 were female and 170 were male. The age varied between 20 and 75 years with a average of 38.77 (SD = 12.16). To ensure consistency and a high quality of responses, the AMT requirements were the same as the previously reported AMT studies from Chapter 4 and 5, namely: the respondents should have already completed $>$1000 tasks on AMT, with $>$98% of these tasks approved successfully, and be located in the United States.

6.6.2 Task

We presented participants with fifty messages designed by experts (see Table C.1 in the Appendix), five representative messages for each of the ten processes of change categories. Similar to the peer-designed evaluation, the expert-designed messages were selected at the author's discretion, and selections were made based on how unambiguously the messages represented the process, yet also did not overlap in content too much between each other. We asked participants to rate each message according to how motivating they found the messages for their own situation ("Please rate how motivating or demotivating you find the following messages for yourself"). All 50 messages were presented on one page and the order of messages was randomized for each of the participants. Messages were rated on a scale from 1 ("Very demotivating") to 5 ("Very motivating") with a 3 as neutral ("Neither demotivating nor motivating").

6.6.3 Measures

At the start of the survey, we asked the participants about their gender, age, native language, understanding of the English language, education level, maternal education level (as an indication of socioeconomic status [Green, 1970]) and main field of work. After, we presented participants with the specific task of this survey: rating the 50 *expert-designed* text messages. At the end of the survey, we measured people's current stage of change[5] [Norman et al., 1998] and perceived experiences with processes of change[6] [Nigg et al., 1999]. To measure participants' personality we used the 50-item IPIP representation of the revised version of Costa and McCrae's [Costa and McCrae, 1992] NEO Personality Inventory[7] which posed 50 statements (e.g., "Make plans and stick to them."). Participants were asked to answer how descriptive they found these statements of themselves (on a 5-point Likert scale, 1 being "very inaccurate" and 5 being "very accurate").

6.6.4 Procedure

Participants were recruited through AMT. They were informed of their compensation (3 US dollars), the goal of the study (finding out which text messages are motivating), and the estimated time to complete the survey (35 minutes). Participants could then decide to accept or decline the survey and proceed to the SurveyMonkey website were the survey was hosted. On the first page, the goal of this study was repeated and participants were asked to complete a consent form. On the second page, the participants were asked to fill in demographic information. On the third page, instructions and context to rate the messages were given. On the fourth page they were presented 50 motivational messages to rate. On pages five and six, they were given questionnaires to fill in. At the seventh and last page, participants were debriefed and given a completion code to fill in on AMT to receive monetary compensation.

6.7 Data analysis: evaluation survey

For the expert-designed messages we selected, we looked at whether the selection of the coded messages were evaluated as representative of the developed process categories and whether they were evaluated similar to the results with the peer-designed messages. The coded messages we selected for each process (five for each process) were representative of the developed categories, as is shown in Table 6.4 through the reliability of the coded message categories. The reliability of the measures was overall very good and similar to the results with peer-designed messages. The reliability scores were between $.68$ and $.83$.

To investigate whether categories of expert-designed messages aligned to the processes were perceived as more motivating than the categories of peer-designed messages aligned to the processes across the stages (H1), we combined the dataset of rated expert-designed messages ($N = 341$) with the dataset of rated peer-designed

[5] `web.uri.edu/cprc/exercise-stages-of-change-short-form`
[6] `http://web.uri.edu/cprc/measures/`
[7] `http://ipip.ori.org/`

	peers			experts		
Coded category	M	SD	α	M	SD	α
Consciousness Raising	3.75	0.66	.76	3.67	0.61	.73
Dramatic Relief	3.02	0.94	.82	2.67	0.90	.78
Environmental Reevaluation	3.13	0.97	.88	3.13	0.69	.75
Social Liberation	3.15	0.69	.73	3.13	0.67	.82
Self-reevaluation	3.67	0.69	.73	3.68	0.66	.75
Self-liberation	3.74	0.59	.72	3.75	0.60	.69
Helping Relationships	3.69	0.67	.74	3.45	0.70	.72
Counterconditioning	3.51	0.61	.68	3.57	0.61	.68
Reinforcement Management	3.86	0.70	.85	3.96	0.71	.83
Stimulus Control	3.21	0.65	.73	3.17	0.66	.70

Table 6.4: Averages (M), standard deviations (SD), and Cronbach's alpha's (α) for all the evaluated motivational peer and expert-designed text message categories. Messages were rated on as scale from 1 ("Very demotivating") to 5 ("Very motivating") with a 3 as neutral ("Neither demotivating nor motivating"). (N = 350)

messages from Chapter 5 ($N = 350$). In Table 6.4 descriptive statistics are reported. To test whether there were differences in how motivating people perceived the messages between the designer of the messages (expert or peer) for the text message categories across the stages of change of the participant, we ran a linear mixed-effects model analysis in R [Team, 2016] with the lme4 package [Bates et al., 2015], and the lmerTest package [Kuznetsova et al., 2016] to calculate the significance of the differences. To accommodate for hierarchical completeness, the model included the motivational rating of the text messages as the outcome, and the categories of the text messages (processes of change), designer of the message (expert or peer), the stages of change of the participant, the interaction between the categories and designer of the message, the interaction between the categories and the stage of the participant, the interaction between the designer of the message and the stage of the participant, and the three-way interaction between category, designer, and stage as fixed effects. The participants and messages were included as random effects. An analysis of variance comparison of the incrementally built models is reported in Table 6.5 (built with xtable package [Dahl, 2016]).

We report on a summary of the model (see Table 6.6, 6.7, and in the Appendix Table D.1, D.2, D.3, and D.4) with the Consciousness Raising (CR) category as the reference level (benchmark to which to compare the scores of the other categories relatively). To make the intercept more interpretable, we recoded the stages from 0 to 4 (e.g., stage 0 = Precontemplation, stage 4 = Maintenance). This means that the intercept score is now the score on the reference level (CR) while all other factors are zero, meaning designer 0 (expert) and stage 0 (Precontemplation).

Looking at the main effects (see Table 6.7) of the model will only give an idea

	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)
0	4	90782.34	90816.14	-45387.17	90774.34			
..1	13	90716.82	90826.67	-45345.41	90690.82	83.53	9	**0.0000**
..2	17	90703.82	90847.47	-45334.91	90669.82	21.00	4	**0.0003**
..3	18	90702.86	90854.96	-45333.43	90666.86	2.97	1	0.0850
..4	54	90685.12	91141.43	-45288.56	90577.12	89.74	36	**0.0000**
..5	63	90514.41	91046.77	-45194.20	90388.41	188.71	9	**0.0000**
..6	67	90512.61	91078.77	-45189.31	90378.61	9.80	4	**0.0440**
..7	103	90527.83	91398.20	-45160.92	90321.83	56.78	36	**0.0151**

Table 6.5: An analysis of variance comparison of incrementally built models. Starting from the baseline model (0), where value is predicted from the intercept and random effects of respondent and messages, adding category (1), stage (2), designer (3), category:stage (4), category:designer (5), stage:designer (6), and category:stage:designer (7).

Random effects	variance	SD
RespondentID (Intercept)	0.226	0.475
Text Message (Intercept)	0.030	0.173
Residual	0.755	0.869

Table 6.6: The variance and standard deviation of the random effects of the model: participants and text messages. Number of observations: 34550, number of respondents: 691 (350 + 341), number of text messages: 50.

of how the expert or peer version of the messages does not make a difference in the rating of the CR category (intercept reference level) for stage 0 (Precontemplation). But, we are interested in all these main effects between the designer of the messages and all the categories across each of the stages. The summary of the model reports the significance of these relations only *in comparison to* the reference level category CR (see Appendix Table D.1, D.2, D.3, and D.4). The summary of the model does report on the relationship between categories and the designer of the message and the stages through the interaction estimates, because the estimate for the designer of the message, in relation to a selected stage and a selected category other than the reference category is equal to the estimate of the designer or the message for the reference category plus the estimate for the interaction between the designer and the selected category plus the estimate for the designer and the selected stage plus the estimate for the designer, the selected category and the selected stage. However, this does not give us the significance of the relation between the designer of the messages and a category for that stage, only the significance of those relations in comparison to the reference level category. Therefore, to see how significant the main effects of designer of the message was between the different stages on the other nine categories, we ran the same model with changed reference levels (see Table 6.8) for

Fixed effects	Estimate	Std. Error	t-value
(Intercept)	**3.436**^{3}	0.152	22.622
DR	**-1.027**^{3}	0.160	-6.415
ER	-0.255	0.160	-1.590
SOL	-0.245	0.160	-1.533
SR	-0.082	0.160	-0.511
SEL	0.164	0.160	1.022
HR	-0.164	0.160	-1.022
CC	-0.055	0.160	-0.341
RM	**0.455**^{2}	0.160	2.839
SC	**-0.345**^{1}	0.160	-2.157
Contemplation	0.183	0.156	1.173
Preparation	0.271	0.150	1.808
Action	**0.316**^{1}	0.149	2.117
Maintenance	0.227	0.142	1.598
Designer (peer)	-0.166	0.198	-0.837

Table 6.7: The estimates and standard error of the fixed effects of the model: process-categories, stage of change of the participant, and designer of the message. Significant effects reported for $^{1}p < 0.05$, $^{2}p < 0.01$, $^{3}p < 0.001$.

the categories and for the stages. Note that the cell in the first row and first column of Table 6.8 reports the identical estimate to the estimate reported for the designer of the message in Precontemplation stage (0 level) in Table 6.7, where we also report the standard error.

The self-assessed stages of change of the 341 participants were as follows: 22 rated themselves in Precontemplation, 53 in Contemplation, 71 in Preparation, 72 in Action and 123 in Maintenance.

To further interpret our comparative results we performed a linguistic inquiry and word count [Tausczik and Pennebaker, 2010] (Table 6.9, Table 6.10, and Table 6.11) with the LIWC program [Pennebaker et al., 2015a] to see if there were differences in the content of the messages between the expert-designed messages and peer-designed messages [Kim et al., 2011]. The LIWC program calculates the number of words that match each of the 93 LIWC dimensions [Pennebaker et al., 2015b], however, for our comparison we made a selection of 35 dimensions that could be relevant for our linguistic comparison. These include seven summary variables, ten linguistic variables, five grammar variables, eight psychological variables, four time orientation variables, and a punctuation variable. All variables except five of the seven summary variables are reported as percentages. The first non-percentage variable is word count, which is just a raw number of words in the text. The summary variables analytical thinking, clout, authenticity, and emotional tone also represent a different score, namely: "Each of the summary variables are algorithms made from various LIWC variables based on previous language research. The numbers are standardized scores that have been converted to percentiles (based on the area under a normal curve) ranging from 0 to 100."[8] In short, analytic thinking captures the degree to which people use words that suggest formal, logical, and hierarchical thinking patterns. Clout refers to the relative social status, confidence, or leadership that people display through their writing or talking. Authenticity refers to whether people reveal themselves in an authentic or honest way, in which they are more personal, humble, and vulnerable. Emotional tone represents a summary score of both the positive and negative emotion dimensions, the higher the score, the more positive the tone (scores below 50 suggest a more negative emotional tone).

6.8 Results: comparing expert-designed and peer-designed messages (H1)

A statistically significant difference was found between the designer of the message (expert versus peer) and the categories within four out of five stages. In Table 6.8, the estimate scores are reported for the designer of the message and the categories between the stages (changing reference levels). The p-values show that the expert-designed messages of the categories ER and RM are rated significantly more motivational than the peer-designed messages of ER and RM by people in the Precontemplation stage. The expert-designed messages of RM are also rated significantly more motivational than the peer-designed messages of RM by people in the Preparation stage. The peer-designed messages of DR and HR are rated significantly more motivational than the expert-designed messages of DR and HR by people in the Preparation, Action and Maintenance stages. Also, the peer-designed messages of SOL and SC are rated significantly more motivational than the expert-designed messages of SOL and SC by people in the Action stage, and for SC also for people in the Maintenance stage.

For the linguistic inquiry and word count analysis, the summary variables (Table 6.9) results show no large differences, except in the authentic dimension (expert messages: 57.35, peer messages: 42.03). Scoring higher in this dimension suggests that people reveal themselves in an authentic or honest way, while scoring lower suggests a more guarded, distanced form of discourse. The expert text messages scored as more personal, humble, and vulnerable compared to the peer messages. All the other linguistic categories do not reveal large differences between the expert and peer-designed messages (Table 6.10 and Table 6.11).

6.9 Discussion

Similar to the approach of our previous research (Chapter 5), the selection of coded messages representing the process categories had substantial reliability. This shows that these messages were a good fit for the processes of change they presented. We

[8] http://liwc.wpengine.com/interpreting-liwc-output/

Peer-Expert	Stages of change				
est. diff.	PC	C	P	A	M
CR	-0.166	-0.060	0.133	0.160	0.140
DR	0.132	0.095	**0.396**	**0.645**	**0.363**
ER	**-0.582**	-0.024	-0.036	0.207	0.097
SOL	-0.156	-0.098	-0.025	**0.292**	0.038
SR	-0.119	-0.139	-0.160	0.094	0.149
SEL	-0.282	-0.163	-0.070	0.050	0.152
HR	0.080	0.145	**0.217**	**0.254**	**0.320**
CC	-0.241	-0.179	-0.112	-0.022	0.085
RM	**-0.550**	-0.143	**-0.210**	-0.043	0.025
SC	-0.314	-0.124	0.012	**0.237**	**0.179**

Table 6.8: Estimates of peer- versus expert-designed message categories across all the stages. Reported scores are estimates of the peer process ratings in relation to the expert process ratings means. All negative numbers represent a higher estimate for experts in that cell, all positive numbers represent a higher estimate for peers in that cell. Numbers in bold are $p < .05$. $(N = Exp/Peer; PC : N = 17/22, C : N = 73/53, P = 94/71, A = 46/72, M = 120/123)$

investigated whether experts designed messages that fit the stage-scenarios (following the theoretical distribution) they were given in the elicitation survey. Looking at higher-order processes, the Experiential process messages were more prevalent in the earlier stages while Behavioral process messages were more prevalent in the later stages, as would be expected following the theoretical distribution. This strengthens our belief that designing messages for behavior change theory this way, both by experts or peers, is a valid approach.

The distribution of expert messages over the processes was proportionately different from the distribution of the peer messages over the processes. Experts provide proportionately more messages than peers for Experiential processes (CR, SOL, SR), which were related to the earlier stages, and significantly less for Behavioral processes (SEL), which were related to later stages of change. This is in line with work from Kristan and Suffoletto [2015] who found that 82% of their expert-designed messages were informational or strategy facilitating messages, which loosely correspond to our categories of Consciousness Raising and Social Liberation (both of which have significantly more messages in our results compared to the peer results of our previous results). Note that the result of Social Liberation needs to be interpreted with caution because the expected number of messages for this category was lower than five.

Summary Language Variables	Expert messages	Peer messages
Word count	692	677
Words per sentence ratio	10.33	11.47
Analytical thinking	49.12	51.18
Clout	97.57	98.16
Authentic	57.35	42.03
Emotional tone	98.42	91.05
Word $>$ 6 letters (%)	17.34	15.81
Dictionary words (%)	96.10	95.42

Table 6.9: The summary variables from the LIWC program. Word count is a raw count of all words used in our 50 expert and peer messages. Words per sentence is the ratio of used words per sentence. Analytic thinking, clout, authentic, and emotional tone are computed scores reported as standardized score from 0 to 100. Word $>$ 6 letters and Dictionary words are reported as percentages.

Linguistic Categories	Expert messages (%)	Peer messages (%)
Linguistic Dimensions		
Total function words	55.20	52.32
Total pronouns	16.91	17.43
Personal pronouns	11.99	12.41
Impersonal pronouns	4.91	5.02
Articles	5.06	3.84
Prepositions	13.87	15.21
Auxiliary verbs	10.98	11.67
Common adverbs	5.78	2.81
Conjunctions	4.34	5.61
Negations	1.73	1.62
Other Grammar Dimensions		
Common verbs	20.38	22.75
Common adjectives	7.95	7.98
Comparisons	3.32	4.28
Interrogatives	1.73	1.03
Quantifiers	3.90	3.25

Table 6.10: Part one of variables from several linguistic dimensions from the LIWC program. Reported scores are in percentages.

Linguistic Categories	Expert messages (%)	Peer messages (%)
Psychological Process Dimensions		
Affective words	7.66	7.53
Positive emotion	6.65	5.76
Negative emotion	1.01	1.77
Social words	13.87	14.48
Cognitive processes	13.01	12.11
Perceptual processes	1.45	2.36
Biological processes	6.21	6.65
Drives and needs	9.10	11.37
Time Orentation Dimensions		
Past focus	1.30	0.89
Present focus	15.90	18.32
Future focus	3.32	2.81
Relativity	18.21	15.81
Informal language	0.29	0.00
Punctuation Dimensions		
All punctuation	14.31	15.36

Table 6.11: Part two of variables from several linguistic dimensions from the LIWC program. Reported scores are in percentages.

6.9.1 Comparing expert-designed and peer-designed messages (H1)

The results indicate only partial acceptance for H1. We hypothesized that people would perceive expert-designed messages representing the strategies as more motivating than peer-designed messages representing the strategies, across the five stages of change. The results show that, *expert-designed* messages were perceived more motivating by people who are yet unwilling to change their exercise behavior, but unexpectedly, *peer-designed* messages were perceived more motivating by people who are either planning on changing, are already taking steps to change, or are maintaining their exercise behavior. Interestingly, Table 6.8 shows that, while messages from only three expert process categories are perceived as significantly more motivating than the peer categories, two of these process categories are in the first stage (Precontemplation), arguably the most difficult stage to motivate people in. On the other hand, messages from nine peer process categories are perceived to be more motivating than expert categories. This is in line with results from Coley et al. [2013], where they found peer messages to be more motivating, and where it could be argued that most of their participants were not in the Precontemplation stage (the exact distribution is not reported), since they participated in a behavior change program (therefore not strictly fitting the definition of Precontemplation: *not* willing to change their behavior).

The fifty messages for experts and the fifty messages for peers were compared between all the stages, but we did not find any differences for any of the categories across *all* the five stages. This indicates that the stage that people are in seems to play a role in how the expert- or peer-designed motivational messages are perceived.

To further interpret our comparative results and to get more insight for the third research question of this dissertation (To what extent does the expertise of the designer of the intervention's motivational content influence how motivating the intervention is perceived?), we performed a linguistic inquiry and word count. Overall, whether experts or peers designed the messages did not seem to have a huge impact on the linguistic categories, except in the authentic dimension. Expert messages scored a 57.35, peer messages scored a 42.03. An explanation of this score according to the program: "When people reveal themselves in an authentic or honest way, they are more personal, humble, and vulnerable. The algorithm for Authenticity was derived from a series of studies where people were induced to be honest or deceptive (Newman, Pennebaker, Berry, & Richards, 2003) as well as a summary of deception studies published in the years afterwards (Pennebaker, 2011)."[9] Scoring higher in this dimension suggests that people reveal themselves in an authentic or honest way, while scoring lower suggests a more guarded, distanced form of discourse. The expert text messages scored as more authentic. An interpretation of this could be that indeed the expert messages are more honest and direct in their advice in the motivational messages, or at least in using the words that are deemed to signal more honesty by the LIWC algorithm. However, no big differences were found on any other linguistic dimension. Interestingly, we did find differences in the evaluation of our expert and peer-designed messages. This indicates that the difference between the expert and peer-designed messages is not easily demonstrated in linguistic dimensions, and more (semantic) interpretation is needed.

6.9.2 Limitations of the current work

Our setup has some limitations, because we elicited the expert messages from Dutch and German experts, and tested them on Americans, while in our previous work, we elicited the peer messages from Americans and tested them on Americans. We mitigated this by not using all the messages for the evaluation, only a selected subset, and for consistency, the same trained coder (coder 1) selected both sets of messages. At this point, we do not know if these text messages will also be motivational when used in applications or other formats (e.g., a virtual agent) or modalities (e.g., audio) or other populations (e.g., non-AMT Americans). Moreover, we do not know yet if messages that are perceived motivating will actually result in behavior change. Nevertheless, we think that the comparison of expert-designed versus peer-designed text message categories across the stages of change might have a real impact. Motivational text messages have been shown to work before in a real context, and tailoring to stage of change has been shown to be good way to further increase effectiveness.

[9] http://liwc.wpengine.com/interpreting-liwc-output/

Therefore, we expect the findings to generalize to other contexts. Future work will need to test our theory-driven design approach in various contexts.

6.10 Conclusion

In this chapter, we described two surveys, one where we asked experts to come up with motivational text messages for a scenario, and one where we asked people to rate fifty of these expert-designed text messages, aligned to the behavior change strategies the ten processes of change, on how motivating they are. We compared these results to our previous study (see Chapter 5) where they rated peer-designed text messages aligned to the ten processes of change. Through this comparison, we evaluated whether there was a difference in how motivating expert or peer-designed text messages aligned to the ten processes of change were rated. We examined these differences across the five stages of change of the participants. The findings are relevant for researchers who want to design motivational text messages with a theory driven-approach and the findings contribute with two clear recommendations for the design of motivational strategies (text messages) for behavior change systems:
(i) People perceive messages tailored to relevant user characteristics as more motivating. For the perception of how motivating the different behavior change strategies are, the user's stage of change plays a role. In terms of system design, this means that the system should accommodate users to, not only fill in their regular demographic information, but also more behavior specific information, like their stage of change. Strategies should be selected that fit the stage of change of the user.
(ii) People perceive differences in how intervention strategies are designed. When identifying which strategies work on users in different stages of change, consider that the designer of that message for that strategy has an effect on how motivating the message itself is perceived. In terms of a system that uses text messages to motivate users, this means that the system should be able to select multiple sources of the same strategy, adapting to the situation of the user. Specifically, expert-designed messages should be used to get people exercising, peer-designed messages should be used to keep people exercising.

Overall, the findings are relevant for researchers who aim to design motivational text messages with a theory driven-approach and it can inform the operationalization of behavior change strategies from the perspective of the designer. To see whether these theory-based motivational text messages are as motivating as we evaluated them to be, the next step (see Chapter 7) is to send a selection of these messages through an application to people in their daily lives over a longer period of time.

7 | Designing behavior change technology and evaluating motivational messages for physical activity behavior change in-the-wild

In the previous chapters we introduced our approach to operationalizing and coding the processes of change from the TTM as text messages. We evaluated our selections of the messages to see how motivating they were perceived to be and whether the messages represented the processes of change. Furthermore, we evaluated whether there are individual differences in how people evaluate these messages that can be used to tailor to, like their stage of change, personality, gender, or the designer of the message. In this chapter, we evaluate the effects of the motivational messages over the course of three months in-the-wild. We compare the results of two conditions, where we either send a daily motivational message tailored to the stage of the participant as per recommendation of the TTM or we send a daily random motivational message. We measure whether there are differences in how motivating participants perceive the messages to be, whether there are changes in their physical activity levels, or whether their are changes in their confidence about doing physical activity.

7.1 Introduction

In previous chapters (Chapter 4, 5 and 6), we elicited, categorized and evaluated peer- and expert-designed motivational messages. We found that, through the use of rigorous coding, the messages we elicited could be categorized into behavior change strategies. Moreover, through self-report survey evaluations, we found that based on the individual differences of stage, personality and gender, different people perceived different strategies as motivating, which is a good basis for tailoring. However, we do not yet know if these messages and individual differences will also result in actual changes, for example changes in behavior or in self-efficacy. Therefore, in this chapter, we present our long-term study[1,2] where we study the effects of a daily motivational message tailored to the recommendations of the TTM for the stages.

[1]We thank Bryan Oostra for his efforts in designing and programming the Android application used in the long-term study.

[2]We thank Yselle van Praet for her efforts in setting up and carrying out post-experiment interviews.

We make the following contributions: contrary to what we expected, we show that 1) people in the random condition engaged in more physical activity over a duration of three months than people in the tailored condition, as shown by sensor data (but not by self-reported data); 2) people in the random condition at times reported higher self-efficacy and decisional balance over a duration of three months than people in the tailored condition; 3) people in the random condition rated the messages they received *more* motivating over the duration of three months than people in the tailored condition.

The chapter is outlined as follows: first we discuss the hypotheses (section 7.2), the method, (section 7.3) data analysis (section 7.4), and the results of the analyses for the longitudinal study (section 7.5). We end the chapter with a discussion (section 7.6) and a conclusion (section 7.7).

7.2 Hypotheses

We started this dissertation with the goal to motivate people to change or maintain their behavior, in this case physical activity behavior. That is what our main overarching research question for this dissertation (see section 1.2) focuses on.

How can people be motivated to inherently change their physical activity behavior using technology?

We discussed that designing and developing effective motivational technology to motivate people to change or maintain their behavior is a challenge. To address this challenge, we decided to base our motivational strategies on existing behavior change theory and to tailor these strategies to characteristics of the user to increase the effectiveness of the strategy. In previous chapters we presented our approach to designing theory-based motivational strategies by operationalizing and coding the processes of change from the TTM as text messages. Furthermore, we evaluated whether there are individual differences in how people evaluate these messages that can be used to tailor to, like their stage of change, personality, gender, or the designer of the message. In this chapter, we evaluate the effects of the motivational messages over the course of three months in-the-wild, which can in turn answer the main research question.

We investigate the effects of receiving a daily motivational message, chosen either according to the recommendations of the TTM for the stages or at random. We evaluate the effects of receiving these messages on physical activity, self-efficacy, decisional balance, perceived motivationalness of the messages and other measures. We expect that messages chosen according to the TTM will have more effect than when the messages are randomly chosen. We hypothesize that:

H1: People who receive messages tailored to their stage of change over three months time perceive the messages as more motivating (H1a), have more confidence in their ability to engage in physical activity (H1b), and engage in more physical activity than people receiving random messages (H1c).

Although the impact of an intervention in the form of a daily motivational message, either random or tailored, might seem minor at first, we expect that, over the course of three months, an appropriate message for someone's stage of change will gradually change their perspective on their physical activity behavior, and in turn, this will show in their confidence in their ability to engage in physical activity, their perception of the messages, and their actual physical activity levels.

7.3 In-the-wild experiment

The study is set up with a mixed methods sequential explanatory design with an emphasis on the quantitative data collection [Creswell, 2009]. This means that we first collect quantitative data, as a longitudinal in-the-wild randomized experiment and then (sequentially), we collect qualitative data to better explain and understand the quantitative findings. For the quantitative longitudinal data we have repeated observations of sensor data and self-reported data. Participants are randomly assigned to an experimental condition (either the 'treatment' or the 'control'), and the experiment takes place during the everyday life of participants. For the qualitative part, we have open-ended questions in the final questionnaire and post-experiment interviews with a selection of the participants. In contrast to our previous studies, this means that we have less control over the study, but more ecological validity and insight into the real effect and causality of the intervention. The study was set up through an Android application. The application was designed to be available to everybody, and was therefore published in the Android Play Store.

Although many HCI researchers aim to promote long-term behavior change through their technology, evaluations confirming long-term behavior change are rarely carried out [Klasnja et al., 2011]. The main focus is on developing the technology. In health sciences, however, randomized controlled trials (RCTs) are the gold standard in evaluating the effects of treatments or interventions.

Even though the minimum time for observing long term behavior change is considered six months [Prochaska and Velicer, 1997], and it is not uncommon to see randomized trials that last two years or more, the minimum time for a randomized trial is usually considered to be three months [King et al., 1998; Heath et al., 2012]. Moreover, there is evidence to suggest that, at least for physical activity interventions, three months is long enough to detect true change [Marcus et al., 1998a,b; Geller et al., 2012]. Even more so, in a recent meta analysis it is suggested that "... an optimal PA behavior change would be observable during the first 3 months of intervention, whereas there would be no additional effect of pursuing intervention" [Bernard et al., 2017, p. 232], suggesting that extending the intervention for more than three months might not be beneficial. Therefore, running our experiment over the course of three months, should be enough to observe the effects of the intervention.

7.3.1 Participants

The invitation of participation started the 21th of April 2017 and continued until the 13th of June 2017. The respondents had the option to participate anonymously (without filling in a name or email address). In total we had 127 respondents. After removing duplicates and faulty installations, we were left with 118 respondents, of which 62 were female and 56 were male, with 58 (24 male) in the random condition and 60 (32 male) in the tailored condition. Of these 118 respondents, 106 lived in the Netherlands at the time. The age varied between 19 and 63 years with an average of 27 (SD = 8.01).

7.3.2 Experimental conditions

Participants were randomly distributed in one of two conditions: the random condition, or the tailored condition. For the random condition, participants received a random message from one of the ten possible strategy categories with the weights for choosing the categories equal. For tailored condition, participants received a random message from one of the three to five categories suitable for the stage they are in according to the TTM. Depending on the stage that the participant self-reported to be in, the possible categories would be:

Stage Precontemplation (1): CR, DR, ER, SOL
Stage Contemplation (2): CR, DR, ER, SOL, SR
Stage Preparation (3): SR, SEL, HR, CC
Stage Action (4): SEL, HR, CC, RM, SC
Stage Maintenance (5) CC, RM, SC

However, due to programming problems, the app did not always work as planned. An error in the code caused some (12.6%) of the messages sent in the tailored condition to be completely random (not based on stage). This happened when the app did not have a list of stage-appropriate messages to send, causing fail-safe code to send a random message instead. This means that the tailored condition contains some noise of random messages and was therefore slightly less tailored than intended (as can be seen from Table E.1 in the Appendix). Moreover, due to an error in choosing which message to send in the random condition, the randomization was not on category level, but on message level, resulting in smaller chances of sending messages from categories with fewer messages (DR, SOL, SC).

7.3.3 Daily messages

Only peer-designed messages were used for this experiment. We chose to use peer-designed messages because in the previous chapter we showed that, overall, peer-designed messages were rated as more motivating and because we could select messages from a larger dataset (2886 peer-designed vs 377 expert-designed messages). For both conditions a dataset of 267 peer-designed messages divided over ten categories was used: Consciousness Raising (30 possible messages), Dramatic Relief (26 messages), Environmental Reevaluation (30), Social Liberation (10), Self-reevaluation (30), Self-liberation (30), Helping Relationships (30), Counterconditioning (30), Reinforcement Management (30), Stimulus Control (21).

7.3.4 Design of the app

The application was designed through an iterative user-centered design approach to be as clean and non-intrusive as possible. An initial iteration started as a project for students in February 2016, first testing the sending of messages through lo-fi prototypes, then developing an actual application, and finishing with a small user study testing tailored versus random messages over the span of three days.

The lessons and recommendations from this project were taken as a starting point for developing the current application, starting November 2016. Notable recommended improvements were: Google Fit integration to measure real physical activity, robust data handling, and data storage in the cloud. Beta testing with versions of this application started January 2017. After several iterations, the application was judged robust in April 2017, and the first participants were recruited 21th of April 2017.

The daily motivational messages were sent as notifications (see Figure 7.1), but did not require immediate attention and stayed in the notification area until resolved. The notification could be resolved easily through the notification area itself, but also in the app. The app itself did not contain any features except the history of the rated messages.

7.3.5 Task

The participants in either condition were asked to rate one motivational message each day for three months. The message was delivered as a notification in the application. To avoid repetition, we tracked the messages sent, so they were rotated out of the possible selection. Moreover, they were asked to fill in a short survey once every two weeks, totaling six surveys, and one final survey.

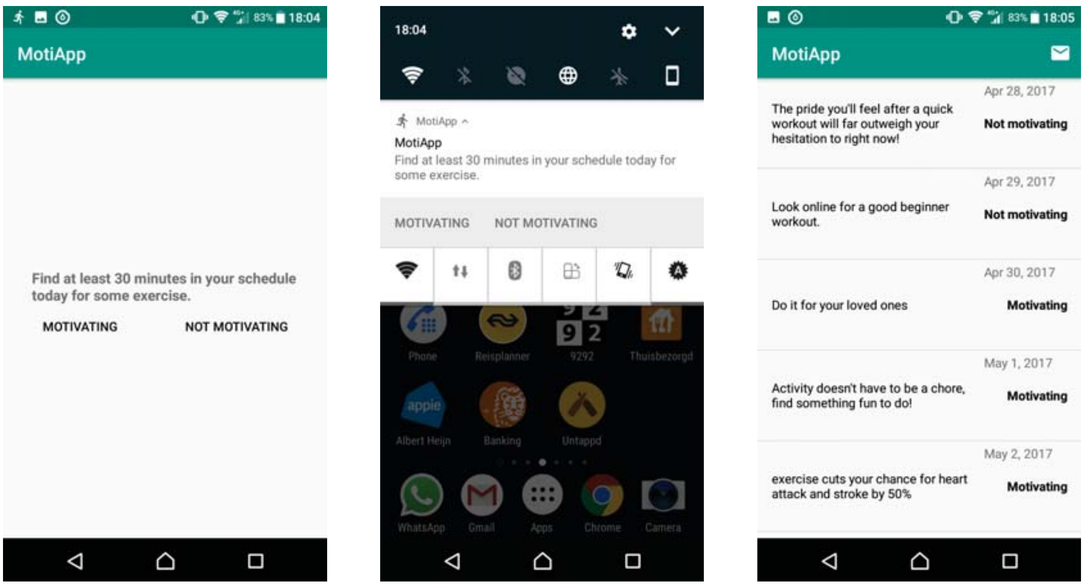

Figure 7.1: Three screenshots of the app. On the left, a motivational message in the app. In the middle, a motivational messages in the notification area. On the right, the history of rated messages.

7.3.6 Measures

After installing the app, we asked the participants their name (optional), email (optional), age, gender, living location, understanding of the English language and how they came across the application. We also measured their personality, through a 10-item version of the 50-item IPIP representation of the revised version of Costa and McCrae's [Costa and McCrae, 1992] NEO Personality Inventory[3], their current stage of change[4] [Norman et al., 1998], and their exercise activities, through the Godin Leisure-Time Exercise Questionnaire [Godin and Shephard, 1997].

During the experiment we asked participants in both conditions to rate their daily motivational message with either 'Motivating' or 'Not motivating' (used for H1a), and, through a biweekly survey, we asked them to rate their self-efficacy[4] [Benisovich et al., 1998] (used for H1b), decisional balance[4] [Nigg et al., 1998] for exercise (also used for H1b, decisional balance is split into the positive and negative aspects of exercise, called the pro's and con's of exercising), their current stage of change[4] [Norman et al., 1998], and their exercise activities [Godin and Shephard, 1997] (used for H1c). Moreover, we measured their actual physical activity level for running, walking, and biking through Google Fit if participants also installed that app (also used for H1c).

After the experiment, we asked participants questions concerning their perception of the motivational messages they received in terms of the interruptiveness, attentiveness to, likeability of, motivationalness, general relevance and personal relevance. We also asked them, in freeform, to name positive and negative experiences with the app, if possible. Moreover, through an adapted version of the Technology Acceptance Model [Davis, 1989], we measured the perceived usefulness of, perceived ease of use of, attitude towards use of, and intention to use the application. The data from these questions has not been used in this dissertation.

7.3.7 Procedure

Participants were recruited by convenience sampling based on the social network of the primary researcher and subsequent snowball sampling. After installing the app, but before starting the study, the participants were informed of the goal of the study (testing how motivating the text messages are), the run time of the study (3 months), the requirements of the study (rate a daily text message, fill in biweekly surveys and an end survey), and they were presented with a consent form. If participants consented, they proceeded to the initial questionnaire (see the previous section 7.3.6). After filling out this questionnaire, they landed on the main page of the app where during the study the history of rated messages was displayed. As explained on the landing page, however, during the first week no messages were sent, this started from the second week. After the second week, the first biweekly questionnaire was sent, repeating every two weeks. After twelve weeks, an end questionnaire was presented through a notification. After filling out this end questionnaire, participants were debriefed and informed that they could deinstall the application, or keep using it if they preferred to, with only the daily motivational messages function.

[3] `http://ipip.ori.org/`
[4] `http://web.uri.edu/cprc/measures/`

7.3.8 Semi-structured interviews

A selection of non-anonymous participants was invited to participate in an interview, either after they finished or quit the experiment. In total, 30 participants were interviewed in-depth about their experiences with the app. Sixteen of these from the tailored condition (eight who finished, eight who did not finish), fourteen of these from the random condition (seven who finished, seven who did not finish). The semi-structured interview was designed to built on some of the insights from the quantitative data of the participant, to gain additional insight into the results found, and into the use or lack of use of the app. In this chapter, we will use some of these insights that relate to the different conditions that we used, exemplified through quotes, to try to explain or substantiate some of the results.

7.4 Data analysis

The data (e.g., ratings of messages, questionnaires, Google Fit data) from the application was stored in an online database[5]. The database was split in different sub-databases, a database containing activities registered through the Google Fit app, a database containing messages rated by the participants, a database containing the responses to the final surveys, a database containing the responses to the biweekly surveys, a database containing registration of Google Fit installations, a database containing the registration of Google Fit goals and a database containing details of the participants from the initial installation and questionnaire. Each of these databases contained data with a participant ID. Through the use of R [Team, 2016] we transformed these databases to combined datasets.

In total, 118 participants installed the app. Their *self-assessed* stages of change: 55 participants rated themselves to be in the Maintenance stage (M), 21 in the Preparation stage (P), 19 in the Contemplation stage (C), 16 in the Action stage (A), and 7 participants rated themselves in the Precontemplation stage (PC).

Many people did not rate or receive every messages or fill in every survey. Moreover, a notably smaller group used the app over a longer term. Although we are interested in the general use of the app, including the attrition, some of the questions we are interested in are better answered with the data of people who used the app over a longer period of time. To this end, we created a smaller dataset with data of people arguably using the app over a longer time. As a selection criterion, we used the information of whether the participant filled in the end survey or not. A total of 47 participants did. These 47 participants also include a few participants (about 5) who rated very few messages (and biweekly surveys). However, we cannot know for sure if having rated very few messages (and biweekly surveys) also implies having seen very few messages, and therefore having little exposure to the experimental manipulation (the messages), or if it just implies that those participants did not enjoy rating a message daily. Therefore we considered it best to include these participants in the 'using the app over a longer period of time' group. On the other hand, by selecting on participants who filled out the end survey, we excluded some participants

[5] restdb.io

(about 7) who rated quite some messages, but not the end survey (yet), and did fill in a notable number of biweekly surveys. Arguably, they have had good exposure to the experimental manipulation, could have participated until the very end and their data on rated messages and biweekly surveys is valuable for the group of longer-time app users. However, we cannot know *with certainty* that they participated until the end (and just missed or did not come around to the end survey), therefore we considered it best to exclude these participants from the 'longer period of time' group.

Looking only at the participants who finished the end survey gives us a subset of 47 participants (24 male, 23 female), with 23 participants (9 male, 14 female) in the random condition, and 24 (15 male, 9 female) in the tailored condition. Their *self-assessed* stages of change: 17 participants rated themselves to be in the Maintenance stage (M), 12 in the Contemplation stage (C), 9 in the Preparation stage (P), 6 in the Action stage (A), and 3 participants rated themselves in the Precontemplation stage (PC). 42 participants were living in the Netherlands at the time. To paint a better picture of the results of our study, we will report on statistics of the dataset that best answers the hypotheses.

A Chi-square test was carried out in SPSS to see if the condition (i.e., tailored or random) had an effect on how many messages were rated either 'Motivating' or 'Not motivating', for both the data of all participants, and the smaller group of longer-term participants. To see if the self-reported activity changed over time due to the condition, we performed a linear-mixed model in R. This approach was chosen over a repeated-measure ANOVA due to the notable amount of missing data in the biweekly questionnaires and the linear-mixed model being better suited for this missing data. Moreover, four ANOVA's were carried out in SPSS to see if the condition had an effect on how many activities (activities automatically registered by Google Fit are running, walking and biking) or how many minutes of activities were performed by the participants, for both the data of all participants, and the smaller group of longer-term participants. To see if the self-reported decisional balance and self-efficacy changed over time due to the condition, we performed a linear-mixed models in R. Again, this approach was chosen over a repeated-measure ANOVA due to the notable amount of missing data in the biweekly questionnaires and the linear-mixed model being better suited for this missing data.

7.5 Results

We present the results of the longitudinal in-the-wild study by looking at the data of *all* participants, but also by focusing on the data of longer-term users as explained in the previous section. To investigate the influence of the tailored daily messages compared to random daily messages, we looked at: how motivating the messages were perceived, biweekly reported self-efficacy, decisional balance (split into the pro's and con's of engaging in physical activity), and physical activity, and recorded sensor data for physical activity.

7.5.1 Rated messages (H1a)

We expected that people self-reporting in the tailored condition, would rate more messages motivating than in the random condition (H1a), however this hypothesis is not supported by the data and analysis. In fact, for the data of the longer-term users we found an opposite significant effect. From Table 7.1 it can be seen that the distribution of the motivating and not motivating-rated messages over the conditions for all the data does not resemble our expectation: there does not seem to be a difference in how many messages are rated motivating between the tailored and random condition. The results show that there is no significant association between the condition and the rating of the messages ($\chi^2(1) = 1.335, p < .248$). From Table 7.1 it can also be seen that the distribution of the motivating and not motivating-rated messages over the conditions for the data of the longer-term users also does not meet our expectation: there does seem to be a difference in how many messages are rated motivating between the tailored (49.7%) and random condition (55.2%), but this seems to be the opposite direction of our hypothesis (H1e). The results show that there is a significant association between the condition and the rating of the messages ($\chi^2(1) = 11.580, p = .001$).

Dataset	Messages (avg.)	Motivating (exp.)	Not motivating (exp.)
all users (N = 118)	5483 (46.47)	2743	2740
Tailored (N = 60)	2969 (49.48)	1464 (1485.3)	1505 (1483.7)
Random (N = 58)	2514 (43.34)	1279 (1257.7)	1235 (1256.3)
Longer-term (N = 47)	3724 (82.93)	1776	1727
Tailored (N = 24)	1961 (81.70)	974 (1025.8)	987 (935.2)
Random (N = 23)	1763 (76.65)	974 (922.2)	789 (840.8)

Table 7.1: The number of messages rated in total, per condition. The number of messages divided by rating of Motivating or Not Motivating and the expected number from Chi-square test, per condition.

7.5.2 Self-reported self-efficacy and decisional balance (H1b)

We expected that people in the tailored condition would be more confident in their ability to do physical activity, which we operationalized by measuring both (H1b) self-reported self-efficacy and decisional balance (split into the pro's and con's of exercising), however this hypothesis is not supported by our data and analysis. In fact, for the pro's of decisional balance and for self-efficacy, the data shows a trend where the random condition is scoring significantly higher in the later weeks in respect to the tailored condition.

For the self-reported decisional balance and self-efficacy data (Figures 7.2, 7.3, and 7.4), we ran three linear mixed-effects model analyses in R similar to the previous self-reported physical activity analysis. We only ran these analyses on the data of the longer-term participants. Our models included the reported score as the outcome (either decisional balance pro's, con's or self-efficacy), and as fixed effects: the week of the reported score (the over time data), the condition of the participants (tailored or random), and the interaction between the week of the reported score and the condition of the participants. The participants were included as random effect.

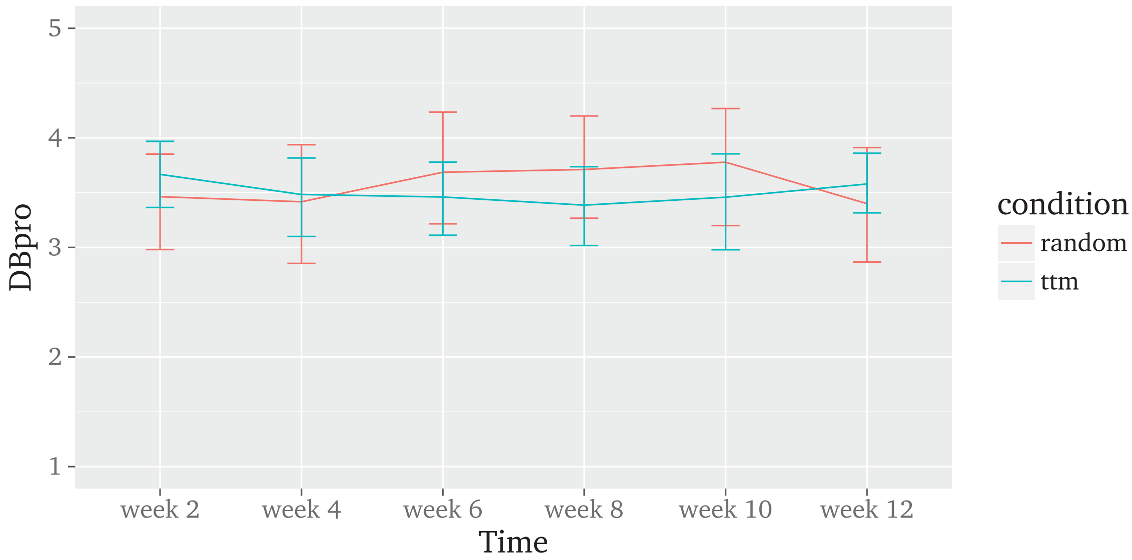

Figure 7.2: The mean self-reported decisional balance pro's scores and mean standard error bars by week divided between the random and tailored condition.

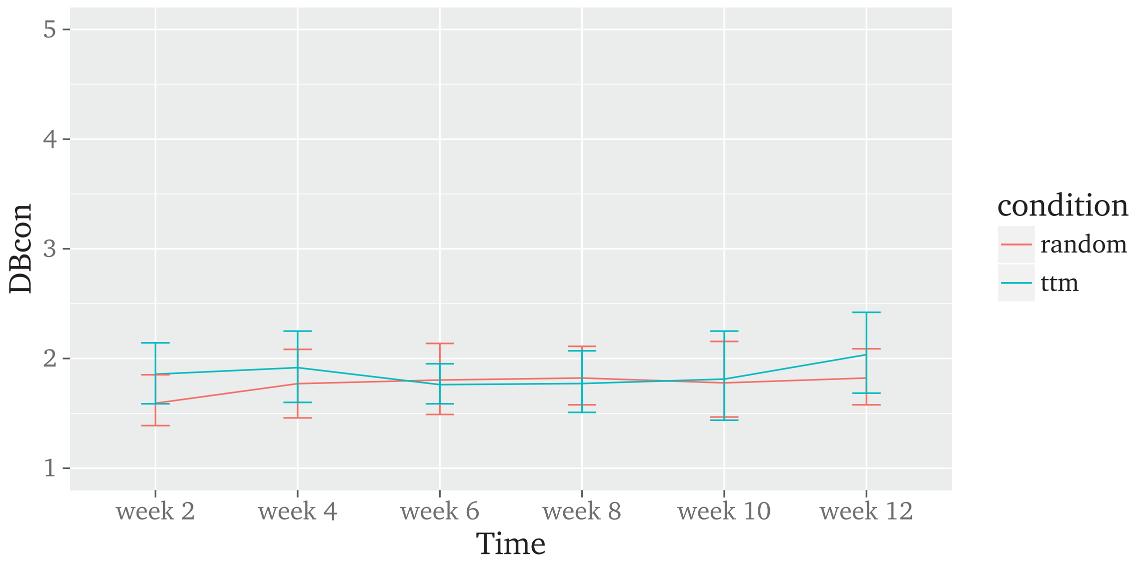

Figure 7.3: The mean self-reported decisional balance con's scores and mean standard error bars by week divided between the random and tailored condition.

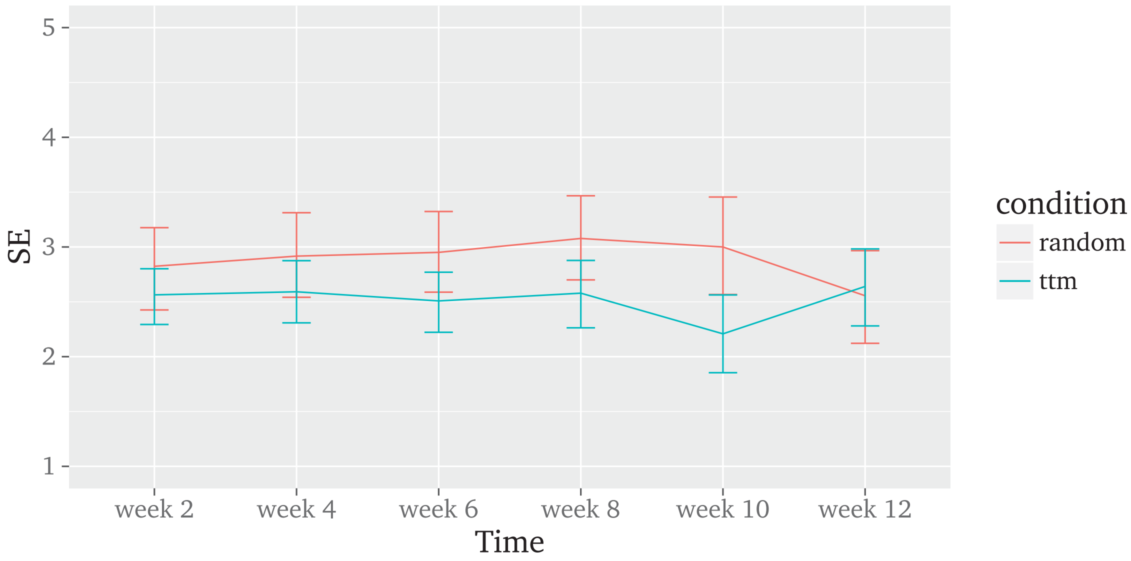

Figure 7.4: The mean self-reported self-efficacy scores and mean standard error bars by week divided between the random and tailored condition.

We report on the summaries of the models (Table 7.2, 7.3, 7.4, 7.5, 7.6, and 7.7), with the week 2's (two weeks after installation of the app) self-reported decisional balance pro's, con's or self-efficacy as the reference level.

Looking at the effects (Tables 7.3, 7.5, and 7.7), we found that neither the condition (random or tailored) nor the week of the self-reported scores of decisional balance pro's, con's or self-efficacy makes a significant difference in reference to the self-reported scores at the reference level, except for the interaction between the decisional balance pro's and the condition, where a negative estimate difference is significant for week 8 and week 10, and the interaction between self-efficacy and the condition, where a negative estimate difference is significant for week 10. This means that the random condition is scoring significantly higher in those weeks in respect to the tailored condition (seen for decisional balance pro's in Figure 7.2 by the random condition scoring higher than the tailored condition, and seen for self-efficacy in Figure 7.4 by the dip in the score for the tailored condition).

Random effects	variance	SD
UserID (Intercept)	0.5076	0.7125
Residual	0.2729	0.5224

Table 7.2: The variance and standard deviation of the random effect of the decisional balance pro's model: respondents. Number of observations: 212, number of respondents: 45.

Fixed effects:	Estimate	Std. Error	df	t value	Pr(> $\|t\|$)
(Intercept)	3.49731	0.20016	88.93000	17.472	<0.0001 ***
DBpro.week4	-0.07523	0.18480	165.16000	-0.407	0.6845
DBpro.week6	0.15558	0.18402	166.59000	0.845	0.3991
DBpro.week8	0.19237	0.18813	164.64000	1.023	0.3080
DBpro.week10	0.27991	0.18813	164.64000	1.488	0.1387
DBpro.week12	-0.01312	0.18957	165.64000	-0.069	0.9449
conditiontailored	0.16052	0.27336	88.30000	0.587	0.5586
DBpro.week4:tailored	-0.10482	0.24935	163.83000	-0.420	0.6747
DBpro.week6:tailored	-0.28808	0.24632	164.25000	-1.170	0.2439
DBpro.week8:tailored	-0.54706	0.25198	162.89000	-2.171	0.0314 *
DBpro.week10:tailored	-0.62283	0.25753	162.71000	-2.418	0.0167 *
DBpro.week12:tailored	-0.07426	0.25363	163.64000	-0.293	0.7700

Table 7.3: The estimates and standard error of the fixed effects and interactions effects of the model: self-reported decisional balance pro's score per week, condition, and self-reported decisional balance pro's score per week and condition. Significant effects reported for '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1.

Random effects	variance	SD
UserID (Intercept)	0.3575	0.5979
Residual	0.1012	0.3181

Table 7.4: The variance and standard deviation of the random effect of the decisional balance con's model: respondents. Number of observations: 212, number of respondents: 45.

Fixed effects:	Estimate	Std. Error	df	t value	Pr(> $\|t\|$)
(Intercept)	1.668009	0.151484	65.920000	11.011	$<$0.0001 ***
DBcon.week4	0.095302	0.113095	160.950000	0.843	0.401
DBcon.week6	0.059158	0.112745	161.920000	0.525	0.601
DBcon.week8	0.156465	0.115079	160.540000	1.360	0.176
DBcon.week10	0.153480	0.115079	160.540000	1.334	0.184
DBcon.week12	0.198546	0.116056	161.220000	1.711	0.089 .
conditiontailored	0.219334	0.207055	65.520000	1.059	0.293
DBcon.week4:tailored	0.002467	0.152417	159.850000	0.016	0.987
DBcon.week6:tailored	-0.103483	0.150620	160.160000	-0.687	0.493
DBcon.week8:tailored	-0.249822	0.153913	159.210000	-1.623	0.107
DBcon.week10:tailored	-0.258109	0.157277	159.080000	-1.641	0.103
DBcon.week12:tailored	-0.172004	0.155009	159.710000	-1.110	0.269

Table 7.5: The estimates and standard error of the fixed effects and interactions effects of the model: self-reported decisional balance con's score per week, condition, and self-reported decisional balance con's score per week and condition. Significant effects reported for '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1.

Random effects	variance	SD
UserID (Intercept)	0.4305	0.6562
Residual	0.1695	0.4117

Table 7.6: The variance and standard deviation of the random effect of the self-efficacy model: respondents. Number of observations: 212, number of respondents: 45.

Fixed effects:	Estimate	Std. Error	df	t value	Pr(> $\|t\|$)
(Intercept)	2.72662	0.17435	74.86000	15.639	<0.001 ***
SE.week4	0.05915	0.14601	162.14000	0.405	0.6859
SE.week6	0.11783	0.14548	163.36000	0.810	0.4191
SE.week8	0.14990	0.14860	161.65000	1.009	0.3146
SE.week10	0.13768	0.14860	161.65000	0.927	0.3555
SE.week12	-0.15907	0.14980	162.51000	-1.062	0.2899
conditiontailored	-0.14128	0.23821	74.35000	-0.593	0.5549
SE.week4:tailored	-0.01614	0.19689	160.87000	-0.082	0.9348
SE.week6:tailored	-0.11505	0.19453	161.24000	-0.591	0.5551
SE.week8:tailored	-0.18582	0.19889	160.06000	-0.934	0.3516
SE.week10:tailored	-0.50337	0.20325	159.90000	-2.477	0.0143 *
SE.week12:tailored	0.13129	0.20025	160.69000	0.656	0.5130

Table 7.7: The estimates and standard error of the fixed effects and interactions effects of the model: self-reported self-efficacy score per week, condition, and self-reported self-efficacy score per week and condition. Significant effects reported for '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1.

7.5.3 Self-reported and recorded physical activity (H1c)

We expected that people who receive messages tailored to their stage of change over three months time will do more physical activity than people receiving random messages (H1c). We operationalized this by measuring self-reported physical activity and recorded physical activity. For self-reported physical activity, this hypothesis is not supported by our data and analysis. For this self-reported activity data (see Figure 7.5), to test whether there were differences in how much activity people reported over time (by way of the biweekly surveys) between the conditions, we ran a linear mixed-effects model analysis in R [Team, 2016] with the lme4 package [Bates et al., 2015] and to output significant differences, the lmerTest package [Kuznetsova et al., 2016]. We only ran this analysis on the data of the longer-term participants. To accommodate for hierarchical completeness, our model included the reported activity score as the outcome, and as fixed effects: the week of the reported score (the over time data), the condition of the participants (tailored or random), and the interaction between the week of the reported score and the condition of the participants. The participants were included as random effect.

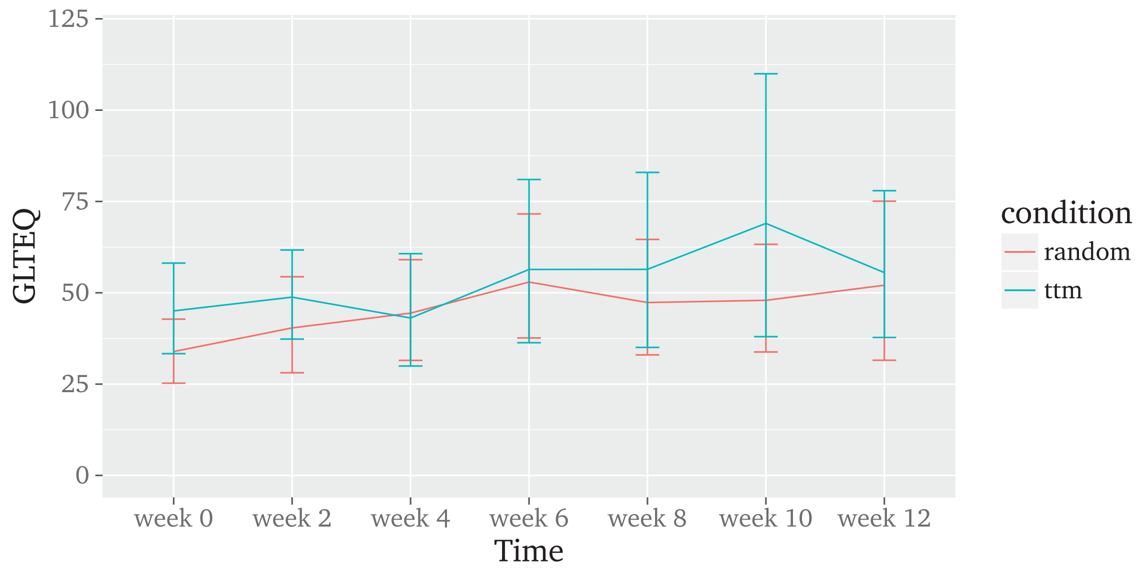

Figure 7.5: The mean self-reported activity scores and mean standard error bars by week divided between the random and tailored condition.

We report on a summary of the model (Table 7.8 and 7.9), with the week 0 (at the app installation) self-reported activity as the reference level (benchmark to which to compare the scores of the other categories relatively). The intercept score is the score on the reference level while all other factors are zero, meaning the week of the self-reported activity and the condition (random).

Looking at the effects (Table 7.9), we found that the condition (random or tailored) does not make a significant difference in reference to the self-reported activity score (intercept reference level) for week 0, but that the week of the self-reported activity does, indicating that the physical activity performed increases over the weeks. However, this is not due to the interaction between the condition and the week of the self-reported activity, because the interaction does not make a significant difference in reference to the self-reported activity score (intercept reference level) for week 0.

Random effects	variance	SD
UserID (Intercept)	1196.7	34.59
Residual	399.2	19.98

Table 7.8: The variance and standard deviation of the random effect of the GLTEQ model: respondents. Number of observations: 259, number of respondents: 47.

Fixed effects:	Estimate	Std. Error	df	t value	Pr(> $\|t\|$)
(Intercept)	33.913	8.330	70.260	4.071	0.000121 ***
GLTEQ.week2	6.637	6.450	205.290	1.029	0.304743
GLTEQ.week4	10.531	6.731	206.030	1.564	0.119245
GLTEQ.week6	17.500	6.590	205.890	2.656	0.008536 **
GLTEQ.week8	14.727	6.884	206.120	2.139	0.033576 *
GLTEQ.week10	12.499	6.884	206.120	1.816	0.070853 .
GLTEQ.week12	18.369	6.867	205.900	2.675	0.008071 **
conditiontailored	11.129	11.657	70.260	0.955	0.343016
GLTEQ.week2:tailored	-7.030	8.823	203.770	-0.797	0.426528
GLTEQ.week4:tailored	-8.110	9.089	204.220	-0.892	0.373276
GLTEQ.week6:tailored	-8.608	8.923	204.090	-0.965	0.335809
GLTEQ.week8:tailored	-7.952	9.275	204.400	-0.857	0.392251
GLTEQ.week10:tailored	4.332	9.522	204.440	0.455	0.649648
GLTEQ.week12:tailored	-12.682	9.257	204.210	-1.370	0.172195

Table 7.9: The estimates and standard error of the fixed effects and interactions effects of the model: self-reported activity score week, condition, and self-reported activity score week and condition. Significant effects reported for '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1.

For the recorded running, walking, and biking Google Fit data, this hypothesis (H1c) is also not supported by our data and analysis. We carried out separate univariate analyses of variance (ANOVA) with the condition as predictor variable and the number of activities and duration as separate outcome variables to see if the number of activities and average duration differed between the conditions, for both the data of all participants, and the smaller group of longer-term participants. From the descriptive statistics in Tables 7.10 and 7.11 it can be seen that average number of activities and the average duration of the activities for both conditions and across both datasets is very similar. The results show that neither of these relations is significantly different between conditions for the data of all participants (activities: $F(1,53) = 0.848; p = .361$, and duration: $F(1,4594) = 1.385; p = .239$), nor for the data of the longer-term participants when it comes to activities (activities: $F(1,25) = 1.238; p = .276$). However, there is a trend for the duration of activities being longer for the longer-term participants in the random condition than in the tailored condition (duration: $F(1,3108) = 3.838; p = .050$), which is opposite to the hypothesis.

Dataset	Act. total	avg. activities (SD)	CI
all users (N = 55/118)	4596	83.56 (56.23)	68.36 – 98.76
Tailored (N = 28/60)	2532	90.43 (55.85)	68.77 – 112.09
Random (N = 27/58)	2064	76.44 (55.78)	53.98 – 98.91
Longer-term (N = 27/47)	3110	119.79 (43.58)	95.05 – 135.32
Tailored (N = 14/24)	1759	125.64 (39.98)	102.56 – 148.73
Random (N = 13/23)	1351	103.92 (60.16)	67.57 – 140.28

Table 7.10: Number of activities total, per condition. Average number of activities and the standard deviation, and confidence interval per condition.

Dataset	Act. total	avg. duration (SD)	CI
all users (N = 55/118)	4596	46.83 (45.01)	45.52 – 48.13
Tailored (N = 28/60)	2532	47.53 (47.81)	45.67 – 49.39
Random (N = 27/58)	2064	45.96 (41.31)	44.18 – 47.74
Longer-term (N = 27/47)	3110	47.83 (46.38)	48.02 – 51.48
Tailored (N = 14/24)	1759	48.23 (51.20)	45.84 – 50.63
Random (N = 13/23)	1351	51.72 (46.62)	49.24 – 54.21

Table 7.11: Number of activities total, per condition. Average duration of activities and the standard deviation, and confidence interval per condition.

7.6 Discussion

In this section, we first consider the technical problems that occurred during our experiment, because these are cause for careful interpretation of the results. Then, we proceed to discuss the results of the analyses and the limitations of the experiment design. Three technical problems with the application presented themselves during the experiment and the analysis of the data. Nevertheless, we argue that the long-term in-the-wild experiment provides a contribution in the form of i) the results of the evaluation of behavior change strategies as text messages long-term and in-the-wild, and ii) the lessons learned designing and using technology in-the-wild over a longer period of time.

Concerning our technical problems, from the data we can see that some participants have received and rated more than one messages a day. This is probably due to reboots of the phone, causing resending of messages. Moreover, due to an error in choosing which message to send in the random condition, the randomization was not on category level, but on message level, resulting in smaller chances of sending messages from categories with fewer messages (DR, SOL, SC). And, as mentioned earlier, the app did not always work correctly and this resulted in 12.6% of the messages in the tailored condition actually not belonging to one of the tailored processes fitting the participant's stage of change. This was due to an misrepresentation of the weights to pick messages from categories. These weights should have been equal across categories, but they were not. Due to errors in the weight calculation of the algorithm, most of the time (90%) when the first category should have been picked of the possible categories list for a participant (e.g., if possible categories for a participant in Precontemplation were: CR, DR, ER and SOL, then CR should be picked), the weight calculation ended up on a negative number and as a fail-safe picked a random message. In short, it meant that messages from one of the categories for a participant in the tailored condition would be shown a lot less, and instead would show a random message from any of the categories (as can be seen in the overview Table E.1 in the Appendix). All these problems interfere in a notable way with the experimental condition and are therefore cause for (very) careful interpretation of the results.

7.6.1 The effects of receiving tailored or random messages (H1)

With the experiment we set out to investigate how communicating crowd-designed motivational messages to people influences their physical activity behavior (to answer this dissertation's main research question), and whether sending daily motivational messages from a process fitting a participant's stage of change in line with the TTM could motivate that person to change their behavior (for example a participant in stage Precontemplation would receive a daily message from either the process of CR, DR, ER, or SOL). To compare, other participants received messages *not* according to this model, but totally at random (but still messages from one of the ten possible processes). We expected that sending motivational messages from the processes that fit a participant's stage of change, would be significantly more motivating than sending random motivational messages. Over the course of three months (i.e., the experiment duration), we expected that people that receive messages tailored to their stage of change over three months time perceive the messages as more motivating (H1a), have more confidence in their ability to do physical activity (H1b), and do more physical activity than people receiving random messages (H1c).

Although we could test these hypotheses with the data of every participant (however little), we felt that some of these hypotheses are better tested with only or also the data of people who used the app over a longer period of time. We selected these people based on whether the participant filled in the end survey or not.

Looking at how the daily motivational messages were rated, we expected that participants in the tailored condition, would rate significantly more messages as motivational (H1a), because they got messages from processes fitting their stage of change. We found, however, that there were no significant differences for the data of all participants, and a contradicting significant difference for the data of the long-term users.

Concerning the constructs of the Transtheoretical Model that are indicators for behavior change, self-efficacy and decisional balance, we expected that people in the tailored condition would have more confidence in their ability to do physical activity, and therefore score their self-reported self-efficacy higher, and would score their self-reported decisional balance pro's higher and their self-reported decisional balance con's lower (H1b). In this case, we only looked at the data from long-term users since we are interested in self-efficacy and decisional balance over time. We found that for the interaction between the decisional balance pro's and the condition in week 8 and week 10, the scores in the random condition where significantly higher. And for the interaction between self-efficacy and the condition in week 10, the scores in the random condition where significantly higher.

For people's physical activity level, we expected that participants in the tailored condition would perform more physical activity, measured from both self-report and sensor data (H1c). Again, we only looked at the data from long-term users since we are interested in physical activity over time. For self-reported physical activity, no significant differences for the condition or the interaction of the condition with time were found. However, week 6, 8, 10 and 12 did score significantly higher in reference to the self-reported activity score (intercept reference level) for week 0, suggesting that overall *for both conditions* the self-reported physical activities did go up compared to when people installed the application (baseline score). For the sensor data, we found that there is a trend ($p = .050$) for the average duration of activities overall, being longer in the random condition than in the tailored condition, which is opposite to the hypothesis.

Overall, we cannot accept any of the hypotheses. In fact, we found a number of significant differences in the opposite direction of the hypotheses. In the remainder of this discussion section, we proffer some interpretations for these results and we discuss the limitations to the work, which are cause for careful interpretation of the results.

First of all, the technical problems complicate the interpretations of these results. The technical problems we suffered could explain the absence of significant results. However, the malfunctioning of the app does not offer insight into why the random messages seemed to have more effect, because in a sense it made the tailored condition 'more random', and in that sense, closer to the random condition, which would normally mean that you would see less (or even no) effects.

The hypotheses were built on several assumptions, from our previous work, but also from theoretical foundation. For example, we assumed that the Transtheoretical Model works for physical activity behavior change. We assumed that the processes that underlie behavior change could be captured through our codebook describing the processes of change. We assumed our coding and subsequent selection of coded messages worked and resulted in messages that were fair representations of the processes of change. We assumed that people reading these fair representations of messages fitting their stage would be motivated more by them than people reading these fair representations randomly selected. We assumed that people honestly fill in their stage of change. We assumed that people would honestly self-report their activity, self-efficacy and decisional balance. And we assumed three months of daily motivational messages would be enough to see and measure behavior change, or at least its precursor in the form of higher self-efficacy and decisional balance. These are a lot of assumptions to build on and any of these assumptions being faulty could lead to the absence of results. However, we have tried to test these assumptions where possible, and we argue that these assumptions are reasonable and therefore accepting them is reasonable. Moreover, any of these assumptions being faulty would not necessarily lead to finding the results in the opposite direction of our hypotheses, as we did.

One possible explanation for finding results opposite to our expectations could be that, somehow, the messages in tailored condition were felt to be too repetitive. From the post experiment interviews, we found some support for this explanation. This is illustrated by for example: “I did not notice the content of the messages change over time. I had the feeling that there were duplicate messages, and then I thought ‘did I not already see this before in a different phrasing’” (R7 - tailored condition), and “I did get a lot of the same sort of messages” (R27 - tailored condition). This repetitive feeling could be explained by the lack of depth in the messages. Although we engaged in rigorous coding and had good reliability for the messages representing the processes, motivational text messages in this format cannot always capture the whole range and depth of a process of change. This would lead to mismatched expectations about the impact of the text messages and would undermine the stage relation of the ‘full’ process that we do not test through the use of these messages that could be, in sense, a poor substitute. It could be then, that a broader selection of motivational messages with different themes (as weak representations of the processes) were more motivating than the repeated selection of motivational messages from a select few categories (as weak representations of the processes).

Another possible explanation for this repetitive feeling could be tied to the construct of the stages of change. Although the stages provide huge heuristic value in the form of conceptualization and measurement of behavior change to be used in something as practical as an app, the ostensibly arbitrary definition of the stages has been a point of critique (e.g. [Littell and Girvin, 2002]). An interpretation would that the stages are not as important as expected and that the processes could be used more stage-independently. This would align to the results discussed in Chapter 5, where we concluded that the way people rated the messages for the processes in the evaluation survey did not match the expectation based on what processes should be most relevant for the stage they are in. It could be then, that the repeated selection of motivational messages from a select few categories (even though they are considered more stage-relevant) was not so motivational because of the limited themes represented, and that in the random condition, the broad selection of motivational messages with different themes was more motivating because there was higher chance that one of these themes (categories) would be salient, or at least less repetitive, for the participant.

The last possibility is that the results are just due to random variation. Most of the results are based on a relatively small sample size, the effects found are small, and the standard errors overlap. It is reasonable to consider that the results are due to random variation and no conclusive evidence for any type of tailoring has been found.

Using theory in practice is not always easy and effective. Although there is a general consensus on the value of most behavior change theories and also the Transtheoretical Model, there is also still plenty of room to increase the effectiveness and salience of such theories by identifying more determinants (e.g., personality) for specific situations (e.g., the exercise domain) and by revealing new dependencies between them.

7.6.2 Limitations of the current work

Next to the technical problems, the study also suffered from some other limitations. Because the study was conducted in-the-wild, we had less control over the study. This resulted in, amongst other things, missing data in rating the daily motivational messages, the biweekly surveys and the end survey. This was augmented by the three months duration of the experiment, which is quite a long time for participants to consistently fill in requested data. On the other hand, three months can be considered a short time to expect real behavior change. Also, the runtime of the experiment (people that participated could start from end of April to end of June), was partly during the summer holidays, which could, for example, interfere with people's normal (physical activity) habits and day-to-day phone use. Another limitation could be the attrition rate of the participants. From the 118 starting participants, 47 filled in the final survey, which means 60% of the participants did not fill in the final survey. Although this might seem like a low number, this is similar to other research, considering the length of our study. For example, Kaptein et al. [2012] performed a two-week study where they sent people messages to reduce snacking behavior, where they report that more than half did not fill out their final diary. In terms of data analysis, a limitation is the large number (17 out of 47, or 55 out of 118) of participants in the maintenance stage, which are (arguably) less informative because they will not change much because they already exhibit the preferred behavior.

7.7 Conclusion

In this chapter, we described the in-the-wild experiment lasting three months, where we sent participants motivational messages fitting the processes of change from the Transtheoretical Model, either at random or tailored to the stage they where in. We compared the results of these two conditions to see whether people that receive messages tailored to their stage of change over three months time perceive the messages as more motivating (H1a), have more confidence in their ability to do physical activity, measured through self-reported self-efficacy and decisional balance (H1b), and do more physical activity than people receiving random messages, measured through self-reported physical activity and recorded physical activity (H1c). The findings show, however, that none of these hypotheses can be supported. In both conditions the self-reported physical activity scores went up significantly over the weeks of the experiment, which could suggest that motivational text messages in general, not specifically tailored messages, convinced people to do more physical activity. In fact, it seemed that participants in the random condition rated more messages as motivating, had a longer duration of activities measured by Google Fit, self-reported higher scores on decisional balance pro's, and self-reported higher scores on self-efficacy. These findings seem to point to an effect opposite of what we expected: random messages motivate people more to do physical activity than messages tailored to their stage in accordance to the processes of the TTM. An interpretation of these findings, similar to our interpretation of the results in the previous Chapter 5, might be that tailoring to stage is not as important as expected. That the broad selection of motivational messages from different categories is more motivating than the repeated selection of motivational messages from a select few categories (even though they are considered more stage-relevant). In other words, it is better to try to motivate people with a wide variety of messages with different themes, so that there is higher chance that something will be salient for the person, in contrast to repeating a few themes over and over. This could also align to our conclusion in Chapter 5 where we concluded that the way people rated the processes did not match their stage expectation, and that in turn personality and gender had significant relations with the process-categories. On the other hand, due to small sample size, small effects, and overlapping standard errors, it is also reasonable to consider that the results are due to random variation and no conclusive evidence for any type of tailoring has been found. Unfortunately we were unable to tailor also to personality and gender in this experiment, but this should definitely be considered for future work. Although in a different way than we imagined, our work shows that the content of the motivational text messages matters and can easily have a positive, or negative, impact.

8 | Discussion

The previous chapters presented the relevant background, research, and results to answer the research questions for this dissertation. This chapter discusses the implications of these results, to what extent the research questions can be answered by these results, but also what the limitations are of the research.

This dissertation aimed to answer the main overarching research question of **how people can be motivated to inherently change their physical activity behavior using technology?** The first step in answering this question, is to figure out a way to motivate people to change their behavior. Literature discussed in Chapters 2 and 3 showed that theory-based behavior change strategies are effective in motivating people to change their physical activity behavior using technology. But how do you shape these theory-based strategies for practical use? The design of such strategies is not trivial, nor is there any content for such theory-based strategies readily available. So the first research question this dissertation aimed to answer is *how can theory-based strategies be translated into a real-world technology-based intervention?* Getting people to change their behavior can be challenging, because behavior change is an intricate process, individual differences make it difficult to generalize this process, and people do not all have the same motivations. This dissertation set out to answer *how does tailoring the intervention to individual differences influence people's motivation for physical activity?* To increase the impact of the theory-based strategies, it is important to tailor these theory-based strategies to individual differences. Next to tailoring the strategies, another factor that influences the perception of the strategy is who designed the strategy. Therefore, the third research question was *to what extent does the expertise of the designer of the intervention's motivational content influence how motivating the intervention is perceived?* Using the insights gathered from answering these questions, this dissertation aimed to answer the main overarching research question. But first, the next section will reiterate the main findings of this dissertation.

8.1 Findings

The research described throughout the dissertation has several findings worth highlighting. The findings have implications for anyone developing motivational technology, as well as for HCI in general. In particular, the findings have implications for theoretical behavior change, for designers of behavior change strategies, and for answering the research questions.

Firstly, our approach to operationalizing the processes of change (which are the behavior change strategies, and also the coding categories) as text-based motivational messages through the use of scenarios, crowdsourcing, and deductive coding offered valuable and interesting results. The scenarios proved to be realistic but general enough to provide a clear description of the stage of change of a fictitious persona, while not triggering participants to design motivational messages containing too much details laid out in the description. As shown in Chapter 4, the distribution of the messages over the coding categories was different for each of the stages of change described in the scenarios, and more or less in line with what we would expect (more Experiential messages for early stages, more Behavioral messages for later stages) for the scenarios based on the relation between the stages and processes of change as described in the Transtheoretical Model. In the evaluation after, the selection of coded messages to represent the process categories showed very good reliability. This suggests that the five messages we selected for each category were a good fit for the process of change they represented. Moreover, when we repeated the coding and evaluation with experts, we found similar results. Overall, these findings suggest that the approach used to elicitate text-based motivational messages through scenarios (crowdsourced or otherwise), deductively code the messages, and evaluate the messages after, has proven feasible, successful, and hopefully, replicable.

Secondly, through evaluation surveys we found that indeed there are individual differences in how motivating processes of change are perceived. Although the results were not the same for the processes of change measured through a questionnaire and the processes of change represented by motivational messages, in both instances individual differences were found. For the processes of change measured through a questionnaire we found that personality traits (E and N) relate to the stages of change and personality traits and the stages of change relate to preferences for certain processes of change (Chapter 4). This supports the general understanding of the processes of change, and is also useful for other researchers using and measuring the processes of change in a similar fashion. For the processes of change represented by our motivational messages, we found significant relations between process-categories (CR, ER, SOL, SR, SEL, HR, CC, and SC) and personality traits (Openness, Extraversion, and Agreeableness), and between process-categories (CR, DR, ER, and SR) and gender. Moreover, we found that the way people rate the messages (processes) on how motivating they are does not always match the expectation of what processes should be most relevant for the stage of change people are in (Chapter 5). This is helpful for others also operationalizing the processes of change, in that the stages may not play the role they expected, but that impact could be improved by tailoring to personality and gender. Also, this is insightful for others operationalizing behavior change strategies in general, in that the operationalized version of the strategies may not evoke the same results as what is theoretically expected.

Thirdly, we find that the designer of the motivational message also plays a role in how the message is perceived. Although this might seem self-evident at first, because it seems likely that contrasting groups such as peers and experts will come up with different messages, in this case it is not so self-evident. This is because the evaluation of designer was carried out by comparing peer-designed messages representing a process of change with expert-designed messages representing the same process, with the messages selected to represent the complexity of the processes equally well. We found that expert-designed messages were perceived more motivating by people who are yet unwilling to change their exercise behavior, and peer-designed messages were perceived more motivating by people who are either planning on changing, are already taking steps to change, or are maintaining their exercise behavior. To get more insight into why we found these differences, we examined and compared the messages on a linguistic level. To see, for example, whether the differences could be explained by experts designing longer messages, or using more complicated words. We found, however, that there was little difference between the two message sets, except for expert-designed messages scoring higher in authority (Chapter 6). This finding is relevant for anyone designing motivational messages and operationalizing behavior change strategies. The designer of the message or strategy is not always considered thoroughly, but the finding shows that the designer can have an impact, specifically for people with differences in their willingness to change their behavior, even when these messages or strategies have been coded and selected to be of a similar strategy.

Fourthly, in-the-wild motivational text messages could work. We expected to see that sending tailored messages would result in more physical activity or higher scores on precursors of increasing physical activity (self-efficacy and decisional balance) compared to sending random messages. We found that in both conditions the self-reported physical activity scores went up significantly over the weeks of the experiment, suggesting that both or neither tailored and random motivational text messages work, or that the increase in physical activity is an artifact of the experiment (e.g., just installing the application, getting invited to the experiment, having the feeling of being observed, or thinking it was expected could have triggered people to engage in more physical activity). However, these types of artifacts of the experiment usually wear off over time, while the results show physical activity increasing over the three month period. This could indicate that the effect is not an artifact of the experiment. We found that random messages were perceived as more motivating, and resulted in longer activities (measured through Google Fit) and higher scores on self-efficacy and decisional balance (Chapter 7). It is important to note that the technology used in this experiment suffered from some technical problems. However, these technical problems do not explain why we found significant results in the opposite direction of our hypothesis. What could explain these results is the fact that some people, through post experiment interviews, noted that the messages in the tailored condition were too repetitive. Another interpretation could be that the stages are not as important as expected (e.g., [Littell and Girvin, 2002]) and that the processes could be more motivating if used more stage-independently. However, since the results are based on a relatively small sample size, small effects are found, and the standard errors overlap, another important interpretation to consider is that these results are due to random variation. Although these results are unexpected, it could be argued that they still show the underlying mechanism that was assumed: the content of the motivational messages matters to motivate people to change their physical activity behavior.

8.2 Findings in light of the research questions

With the main findings of this dissertation discussed, we can look at the research questions to see if we can answer them.

To start with the first research question first — *How can theory-based strategies be translated into a real-world technology-based intervention?* — this question can be answered with our first finding. Overall, it seems that scenario-based elicitation tasks can be used in crowdsourcing to design theoretically-grounded behavior change strategies. This question was born out of necessity. At the time the research started, no established methodology was in place to translate theory-based strategies to real-world technology-based interventions. Related work did sometimes use translated theory-based strategies, but it was not common to describe the approach that was followed to get to these translations, and therefore replicating these approaches was infeasible. This dissertation describes research with an innovative approach to translating theory-based strategies, where online crowdsourcing, scenarios, coding and codebooks are combined and subsequently documented for future use. The answer to the first research question is that a feasible and efficient way to translate theory-based strategies to real-world technology-based interventions is to follow the approach we used of elicitating text-based motivational messages through scenarios (crowdsourced or otherwise) and deductively coding the messages.

The second research question — *How does tailoring the intervention to individual differences influence people's motivation for physical activity?* — can be answered with our second and fourth finding. It seems that people evaluate the intervention or strategy differently based on either stage, personality, or gender. Of course, many other factors could be considered when tailoring to people, and many other factors, for example the context, the timing, or the location of a person can influence the impact of an intervention. However, based on the findings of this dissertation the answer to the research question is that tailoring of theory-based interventions to individual differences, like gender, personality, or stage of change can indeed influence how the intervention is evaluated, and therefore the impact of the intervention.

The third research question — *To what extent does the expertise of the designer of the intervention's motivational content influence how motivating the intervention is perceived?* — can be answered with our third finding. In most literature, the designer of a strategy or message is not considered at great length. Having peers design these strategies or messages has been a development in recent literature, where before having experts design strategies has been the norm. In this dissertation both groups have been asked to design messages using the same approach. The interventions designed by the crowd did not differ from the expert-designed interventions in terms of linguistic differences. However, there were differences in how the interventions were evaluated based on the stages of the participant. Overall, it seemed that expert-designed messages were perceived more motivating by people who are yet unwilling to change their exercise behavior, and peer-designed messages were perceived more motivating by people who are either planning on changing, are already taking steps to change, or are maintaining their exercise behavior. Overall, based on the findings of this dissertation the answer to the research question is that the designer of theory-based interventions does indeed influence how the intervention is evaluated, and therefore the impact of the intervention.

The main overarching research question — *How can people be motivated to inherently change their physical activity behavior using technology?* — can be answered by our fourth finding, although all findings contribute to this answer. Overall, it seems that the effect of sending these crowd-designed interventions of theory-based strategies is there, but not as expected. We expected that tailoring the strategies to the stages based on the Transtheoretical Model would result in more physical activity or higher scores on precursors of increasing physical activity (self-efficacy and decisional balance) compared to sending the strategies randomly. It seemed, however, that sending the strategies randomly resulted in more physical activity and higher self-reported self-efficacy and decisional balance. For the specific study, this was an unexpected result, the relation between manipulation (sending tailored or random messages) and outcome (e.g., self-efficacy, physical activity) did not turn out as expected. From the perspective of the broader research question, this unexpected result can still answer the question. Overall, based on the findings the answer to the research question is that sending theory-based strategies in the form of motivational messages seems to influence people in physical activity behavior (just not in the way that was expected). However, due to the limitations of the study, the results should not be interpreted as conclusive evidence.

8.3 Limitations

Limitations are a part of running studies. Throughout the studies discussed in this dissertation we ran in to several limitations. The chapters separately address the limitations of the studies. However, some of the noteworthy, key limitations are reiterated here.

Firstly, there are **limitations to the TTM**. The model has been used frequently, but has also received a share of criticism (e.g., [Joseph et al., 1999; Littell and Girvin, 2002; Bridle et al., 2005]). Most criticism revolves around the construct of the stages of change. Although the stages provide huge heuristic value in the form of conceptualization and measurement of behavior change to be used in something as practical as an mobile phone application, the ostensibly arbitrary definition of the stages has been a point of discussion (for example see this excellent critique [Adams and White, 2004] and subsequent excellent commentary [Brug et al., 2005]). The stages of change are arguably a seemingly artificial way to put people into a certain stage of their complex health behavior, such as physical activity behavior change, and the staging algorithms to do this lack validity. These concerns should be taken into account when interpreting the TTM results, specifically those results involving the stages of change. However, conceptual models like this are never completely accurate. If the model would be completely accurate, it would represent reality. It might be more reasonable to consider whether there is value to the model and improve upon it, then to argue that it is not completely accurate and dismiss it altogether.

Secondly, there are **limitations to motivational messages as representation of theory**. To use text messages as representations of the processes of change assumes that these, on average ten to eleven word text messages can capture the subtleties of a behavior change strategy that in original versions would, for example, be conveyed through personal feedback, education, confrontation, interpretation, bibliotherapy, or media campaigns. Moreover, using text messages as representations of the processes of change assumes that the processes could be captured through the use of coding and a extensive codebook.

Thirdly, there are **limitations to adopting (online) surveys**. As mentioned in several chapters, we gathered the respondents through Amazon Mechanical Turk, which could mean a misrepresentation of the general adult population of a Western society, although studies have reported that AMT generally has quite a good representation for online survey standards [Mason and Suri, 2012]. However, we are aware that there can be limitations to how representative the sample is, as not all types of people would be equally likely to register for AMT, nor participate in these surveys. The surveys gathered data at one specific point in time (cross-sectional data), which only provides evidence for correlation, not causation. For example, in Chapter 4 we argued that people with certain personalities might not progress well through the stages of change, because personality scores were different between the stages. However, the results only provide evidence for correlation, therefore it could also be that people's personalities change when they progress through the stages, which would also be a valid explanation as to why there are personality score differences between stages. Moreover, the use of surveys results in limitations to our understanding of the effectiveness of a behavior change intervention as a metric to judge its success. A large part of this dissertation has focused on the effectiveness of an intervention by asking participants in a survey to rate how motivating they perceive the intervention to be. However, there is no guarantee that how motivating participants perceive the intervention to be leads to actual behavior change when used in real life. We aimed to test this in our in-the-wild study, but the results proved mostly inconclusive. Related to this, surveys involve self-assessment from participants, however people are not necessarily able or willing to truthfully self-assess certain questions, such as which stage of change they are in or whether a certain text message is really motivating or not.

Fourthly there are **limitations to the participant samples**. The online surveys are all set up through SurveyMonkey with participants gathered through Amazon Mechanical Turk. To ensure the quality of English used in the text messages designed by the participants, we restricted the participant pool to people located in the United States. Subsequent evaluation studies were also set up through SurveyMonkey with participants gathered through Amazon Mechanical Turk. Also, all participants recruited through Amazon Mechanical Turk received monetary compensation (at a rate of $6 an hour). However, the survey to elicit motivational text messages from experts was distributed in the Netherlands and Germany because this allowed us to approach the experts in person. The experts also did not receive monetary compensation. Participants for the final in-the-wild study were recruited through convenience sampling based on the social network of the primary researcher and subsequent snowball sampling. This meant that the vast majority of participants were Dutch. Moreover, the sample of this study was relatively small (118 participants started, 47 participants finished). Also, participants did not receive monetary compensation.

Fifthly, there are **limitations to the in-the-wild study**. The study was set up to send one daily motivational message either according to a tailored plan or at random. However, due to technical problems, some participants received and rated more than one message a day, and, some participants in the tailored condition received and rated messages (12.6%) actually not belonging to their tailored plan. Also, we assumed that three months of daily motivational messages would be enough to see and measure behavior change, or at least its precursor in the form of higher self-efficacy and decisional balance, however it has been argued that seeing changes in behavior could take a minimum of six months up to two years. Moreover, the data gathered in this study includes a fair share of missing data, due to either technical problems or participants not using the app daily.

It is important to consider that the discussed limitations are possible limitations, not necessarily probable limitations. It is valuable to acknowledgement and consider the possible limitations to the work to improve future work and interpret the results. Overall, we argue that the results of this dissertation are fairly generalizable, for example personality, gender, stage of change or designer of the message will probably also influence the perceived motivationalness of strategies in different contexts, however this might not be for exactly the same strategies or for exactly the same personality traits, gender, stage or message designer. Moreover, the approach to operationalize the processes of change as text-based motivational messages through the use of scenarios, crowdsourcing, and deductive coding is likely to be valuable in different contexts, although for optimal effect, adjustments to the scenarios, surveys or codebook might be needed. Lastly, the in-the-wild study proved that sending theory-based strategies in the form of motivational messages does influence people in physical activity behavior, however more replication in identical and different contexts is needed to really find out the underlying mechanisms.

9 | Conclusion

The previous chapter discussed to what extent the research questions could be answered. This chapter summarizes the work done, highlights the contributions, and outlines future work based on the direction of this dissertation.

This dissertation presented research in which we investigated, designed, and evaluated behavior change strategies that can be used to motivate people to change their physical activity behavior. The research showed how theory-based strategies can be translated to real-world technology-based interventions, to what extent tailoring of these theory-based interventions to individual differences can influence the impact of the intervention, and to what extent theory-based interventions can effectively empower people to change their physical activity behavior using technology. In this chapter, the contributions of the work are summarized and final remarks are made, and the chapter ends with a perspective on future work. The work presented in this dissertation contributed to the challenge of designing and developing effective motivational technology based on behavior change theory in several ways.

Firstly, this work had a methodological contribution in terms of exploring an approach to operationalize the strategies suggested in behavior change theories and models to be used in practice. The work showed that these strategies can be operationalized by elicitating text-based motivational messages through scenarios (crowdsourced or otherwise) and deductively coding the messages. Although using elicitation and coding is not new, we do not know of other research applying these methods to the operationalization of behavior change strategies.

Secondly, this work had an empirical contribution to the challenge of designing and developing effective motivational technology by exploring characteristics of the user to which the interventions or strategies could be tailored. Although the results apply to the strategies that were tested specifically, the implications of our results (that impact of interventions could be improved by tailoring to other characteristics) can be more generalizable in the sense that other behavior change strategies could also benefit from exploring other characteristics that could impact the effectiveness above and beyond what is formalized in the specific theories or models. In this sense, the empirical contribution could also be viewed as a (minor) theoretical contribution to the behavior change theory literature at large and the Transtheoretical Model literature in specific.

Thirdly, this work had another empirical contribution to the challenge of designing and developing effective motivational technology by showing that the designer of the operationalized behavior change strategy also matters. The work showed that messages fitting the same behavior change strategy were perceived differently (motivationally) based on the designer of the message (expert or peer) and based on the stage of behavior change people were in, with expert-designed messages being preferred early in behavior change, while peer-designed messages being preferred later in behavior change. In a similar manner as the second contribution, this empirical contribution could also be viewed as a minor theoretical contribution.

Fourthly, the work contributed empirically to the challenge of designing and developing effective motivational technology by showing that the content of the motivational messages matters to motivate people to change their physical activity behavior in the sense that people were more motivated by random messages from ten behavior change strategies than tailored messages (based on their stage of change) from three to five behavior change strategies. In a similar manner as the second and third contribution, this empirical contribution could also be viewed as a minor theoretical contribution. Specifically, this could be a minor theoretical contribution to the literature that disputes the use and strict progression of the stages of change as is, and would opt for a looser definition of the stages, which would also involve a looser use of the processes fitting the stages.

Fifthly, the work also has a contribution in the form of datasets, as the fifty expert and peer-designed messages used to operationalize the ten processes of change behavior change strategies are available in this dissertation (see Table B.1 and C.1). Moreover, the original datasets of motivational messages (audio recording of peer-designed messages and text version of peer and expert-designed messages) are available for researchers doing similar work upon request from the author of this dissertation.

Thus far, the most important contributions of the research described in this dissertation have been highlighted. We think that the findings also contribute to the wider field. Therefore, in this section we reevaluate the contributions and the dissertation in light of some of the most important challenges discussed by current or emerging leaders of the field in The 2016 Theme Issue of the American Journal of Preventive Medicine: Digital Health: Leveraging New Technologies to Develop, Deploy, and Evaluate Behavior Change Interventions [Yardley et al., 2016a].

The Theme Issue discusses the future of the field (of "Digital Health Behavior Change Interventions") extensively through five papers [Patrick et al., 2016; Hekler et al., 2016; Yardley et al., 2016b; Murray et al., 2016; McNamee et al., 2016] and a commentary [Kelly, 2016]. As is mentioned in the papers, the next steps or remaining challenges are many. A selection of these steps and challenges are highlighted here because they are valuable and relate to approach and contributions of this work and show the relevance and timeliness of the research questions and the findings.

For example, in one of the Theme Issue papers, Hekler et al. [2016] argue that "It is essential for advancing behavioral science not only to focus on building computational models but also on the development of these models and behavioral theories in a collective mindset where each group of scientists are clearly specifying when a theory/model will and will not be useful and, by extension, the interventions that are created based on the theories." From the translation of theory-based strategies to practical interventions, to finding out to what extent communicating crowd-designed motivational messages to people influences their physical activity behavior, this dissertation has focused on getting and describing specific results, and describing the specific circumstances leading to these results. This is exemplified in the approach taken to translating theory-based strategies, in the evaluation of comparing different designers of strategies, and in the evaluation of the messages representing the strategies in a three-month-lasting field study.

Yardley et al. [2016b] discuss future, more elaborate work in engagement and tailoring, but note that "conventional tailoring of content to match an individual's demographic characteristics, may still be required to ensure that users are not presented with material they find so alienating or demotivating that they abruptly cease using the intervention." The need for "conventional tailoring", as it is called in this paper, is one of the main arguments of this dissertation, in particular tailoring to demographic characteristics and individual differences like personality, gender and stage of change. As is exemplified by the results described in Chapter 5, where some of the strategies rated on a scale from 1 ("Very demotivating") to 5 ("Very motivating") with 3 as neutral scored lower than a 3, meaning that the strategies were considered, in certain circumstances, demotivating. To quote the discussion section of that chapter: "In that sense, the measurement of how motivating these messages are perceived is not only input for what strategies to definitely use, but also for what strategies to avoid for which users to increase the likelihood of this system being used over longer periods of time." In the rated evaluations of our strategies in this dissertation, the chosen scale was intentionally from demotivating to motivating, so that we could account for strategies that some people could find alienating or demotivating, and through conventional tailoring we could avoid using these strategies on those people.

Murray et al. [2016] mention a guideline that stipulates that there should be "reasonable confidence that the intervention plus delivery package can be implemented with high fidelity". Again, this guideline resonates with the overall setup of the research described in this dissertation. The approach to realizing the design of the theory-based strategies as real-world technology-based interventions has been to stay as true as possible to the original theoretical strategy. The definitions of the stages of change were taken to shape the content of the scenarios that were used to trigger people to design motivational messages. The definitions of the processes of change were used to develop of codebook representing these processes of change as closely as possible while considering the text message form. The evaluations of selections of the text messages were, in part, to confirm the internal consistency of the messages selected for each process-category. So all in all, many of the steps taken in the research described in this dissertation have been to ensure with reasonable confidence that the interventions plus the delivery package were implemented with high fidelity.

Overall, discussing the approach and contributions of this dissertation in light of an extensive Theme Issue discussing the future of Digital Health Behavior Change Interventions shows that the approach and contributions of this dissertation are relevant and timely.

Considering the current state of technology, research done in the field, and our contributions, we can also reflect on some of the next steps or leaps that could be taken to advance the field. Technology is ever advancing, and implementing and using more of these advances will open up more and more possibilities in sensing and monitoring people in daily life. The ubiquity and pervasiveness of more technology with sensors can result in more data, more accurate data, and real-time data. These types of data can be used in a myriad of ways, for example, real-time sensor data can be used for real-time interventions, and more and bigger data can be used for more accurate machine learning algorithms. However, these are not next steps that would logically follow from this dissertation, since this dissertation did not focus on (the development of) technological advances.

The advancement of technology could provide next steps that would follow from this dissertation in a different way, for example: faster and better testing of motivational strategies and theory development. This could happen in the 'classical' sense, similar to the approach used in this dissertation, where motivational strategies are designed and evaluated based on theories and models and subsequently implemented and evaluated in-the-wild. Combining this with a robust app or any other robust technology that can deliver motivational strategies (in text form or other) would provide a platform to easily test and evaluate strategies through the convenience of a ubiquitously available app in any of the app stores. This would in turn hopefully result in more precise, quantitative, and testable theories, models or strategies. Considering this in light of our own approach, new strategies can be easily crowdsourced, evaluated, added to the app and tested. In this way, future research could incorporate larger studies with control groups, which are usually not feasible for HCI research [Klasnja et al., 2011], but which are the gold standard in health sciences for the evaluation of effects of interventions.

Another option already discussed in Chapter 5, is that this type of research of investigating all the different factors that might relate to certain strategies could be done by an adaptive system [Kaptein et al., 2012]. An adaptive system would not need any prior input on how the factors relate, and could find the most optimal model for users by testing all the strategies. In Chapter 5 we argued that: "for the purpose of our system, we feel that this does not fit our research approach in two ways. First, the use of an adaptive system with a wide variety of possible strategies and no prior input, has a high risk of being abandoned by users, because on average, the users will be exposed to a number of strategies that might not be motivating or even demotivating them (one of our results of the evaluation), resulting in a higher chance of abandoning the technology. Second, starting from an adaptive system with no prior input might optimize the model in an more effective way compared to a predetermined static model, but it will not help explaining or interpreting why the model works. For example, we could find that a certain group of messages is highly effective for some user features, but if these messages do not explicitly group on a certain theme or underlying construct it will be hard to interpret and replicate the results." After our in-the-wild evaluation, however, it seems that our first argument is not valid in our particular case. The wide variety of possible strategies that we used in our random condition were received and perceived more positively than our tailored condition. In light of that, an (at least partially) adaptive system would be a logical next step for our system, and possibly also for similar systems in general. Moreover, recent advancements in open source adaptive systems [Kaptein and Kruijswijk, 2016; Kruijswijk et al., 2016] make it easier for anyone to use an adaptive system with a multitude of strategies to explore which strategies could work. Nevertheless, we think that it would still be valuable to base the initial system on an already tested model, to help interpret and explain when the system works or does not work.

Another venue to explore would be to take a step back in how we currently design motivational messages. Analyzing text is nothing new, and programs (e.g., LIWC) exist to do this for you, as shown in Chapter 6. However, in the context of behavior change interventions it is not very common to do linguistic research on the motivational messages used as intervention. In that sense, another step forward might be to take a step back again and really build up the messages we send as interventions and curate the content.

This dissertation presented research that investigated, designed, and evaluated behavior change strategies that can be used to motivate people to change their physical activity behavior. Although the research contributed new knowledge to the challenge of 'how can people be motivated to change or maintain their physical activity behavior', motivating people to change their behavior has still proven challenging. As stated in the introduction: "changing behavior is an intricate and difficult process, and individual differences make it difficult to generalize this process. People do not all have the same motivations, which makes it difficult to successfully motivate all types of different kind of people to do more regular physical activity." So what work needs to be done to get closer to really motivate people to change their, in this case physical activity, behavior?

New developments in technology will offer faster and better algorithms, more and better data, and bigger and better computational models. However, these developments do not necessarily make behavior change easier. To advance the field of digital health behavior change interventions, we believe a few things are essential: to keep using theories and models as the basis for intervention strategies and to keep updating theories and models based on new insights, to specify when and why certain strategies will work, to develop interventions with high fidelity, and to document the development of interventions so that they can be reproduced with high fidelity. In this way, we do not just get more data on behavior change, we can help make behavior change easier.

A | Codebook used for peer- and expert-designed motivational messages

Process	Consciousness Raising/Increasing Knowledge — Code: CR
Brief definition	Increased awareness of causes, consequences and cures for not being physical active.
Full definition	CR is a process that involves increased awareness of causes, consequences and cures for not being physical active. The intention is increasing the knowledge of unaware individuals with objective information.
Practical def.	Encourage subject to read and think about physical activity (cognitive process)
Perspectives	To start/trigger or advance the process: messages that give (objective) information about the benefits or disadvantages of not exercising.
Inclusion	Arguments with information (mostly) facts about benefits or disadvantages for health; (objective) confrontations with diseases; prevention for diseases
Exclusion	Subjective arguments why people should exercising; benefits for appearance; a proposal
Example incl.	“It can prevent all types of diseases like Diabetes and cancer” “Exercise can help you live longer” “Exercise will help you to be healthy and fitt.”
Example excl.	“You worked hard for everything, why not also for your health?” - This would fit better with SR “You will look much better; You will feel better when you exercise.” - This would fit better with SR

Process	Dramatic Relief/Emotional Arousal/Being aware of risks — Code: DR
Brief definition	Produces increased emotional experiences for not being physical active.
Full definition	DR is a process that gives individuals an increasing in emotional experiences (like feeling fear, anxiety, worries) when they are not physical active.
Practical def.	Dramatically making the subject aware that being inactive is very unhealthy (cognitive process)
Perspectives	To start/trigger or advance the process: messages that ‘play’ on the dangers of not exercising (something shocking) or the subjective benefits of exercising (something desirable). To distinguish: When an DR messages also contains information, the difference between DR and CR is in the more subjective/dramatic/exaggerating feeling (good or bad) of the messages.
Inclusion	Warnings that elicits an emotional response, like making individuals scared about the risks with eventually an action that can be taken.
Exclusion	Objective arguments of the benefits of exercise
Example incl.	“You get cancer and die when you not exercise”
Example excl.	“It prevents diseases.” - This would fit better with CR

Process	Environmental Reevaluation/Caring about consequences to others — Code: ER
Brief definition	Consideration of how the individual affects others in his social environment
Full definition	ER is a process to increase the individual's consideration of how his physical inactivity or activity affects his social environment, like family and friends.
Practical def.	Encourage the subject to recognize how his inactivity affects his family, friends, and coworkers (cognitive process)
Perspectives	To start/trigger or advance the process: messages that suggest or remind you not exercising might have a negative effect on family, friends or environment; or that trying exercise would have an positive effect on family, friends or environment. To distinguish: Messages in this category can usually also fit in other categories like SR, CR or DR or even SEL, but fit into ER because the main topic is family, friends or social environment.
Inclusion	Reminders/hints for recognizing what influence his (good or bad) behavior has on his social environment; Raises awareness about his function as positive or negative role model; Implies to notice the effect the individual has on others.
Exclusion	Packaged compliments
Example incl.	"Just think about how your wife will notice" "Let's get you in shape so you can play with your kids!"
Example excl.	-

Process	Social Liberation/Increasing healthy opportunities — Code: SOL
Brief definition	Awareness of alternatives or social opportunities when being physical active.
Full definition	SOL is a process to increase the awareness of the individual's to see social opportunities or alternatives if being physical active in the society.
Practical def.	Help the subject to increase her awareness of opportunities to be physically active (cognitive process)
Perspectives	To start/trigger or advance the process: messages that suggest or remind you not exercising might not fit anymore into today's society; or that if you want to try exercising, the society around you has plenty of opportunities to do so.
Inclusion	Options of social opportunities the society offers for physical activity; Implies to notice possibilities in the society.
Exclusion	What they should do to be physically active
Example incl.	"If you need help: there are programs online, local gyms"
Example excl.	-

Process	Self-Reevaluation/Comprehending benefits — Code: SR
Brief definition	Cognitive and emotional reappraisal about self-image
Full definition	SR is a process that contains an emotional and cognitive (re)evaluation of the individual's self-image about being or not being physical active.
Practical def.	Help the subject to understand the personal benefits of being physically active (mostly cognitive process and a bit behavioral process)
Perspectives	To start/trigger the process: messages that suggest not exercising might conflict with personal values or that trying exercise would match very well with personal values. To advance the process: messages that tell you how not exercising conflicts with personal values or going to exercise will match very well with personal values (persuasive argument).
Inclusion	Triggers people to think about his or her self-image; Imagery, how would it be if? ; Proposals: if you exercise, you will feel great; Reflections; persuade arguments that helps to make a decision. Benefits for appearance. Contains mostly thing like "you will/would [look great] if", "think about".
Exclusion	Only suggestions or imperatives without arguments why; or if the messages suggests the individual has already started. This would fit better with SEL
Example incl.	"Think of how great you will feel when you get started", "You should up and do something, you will feel great", "You will look much better!"
Example excl.	"You should go on a walk." - This would fit better with SC "Each day you do this, it will get easier." - This would fit better with SEL (assuming somebody already started here), as reinforcing the goal and commitment.

Process	Self-Liberation/Committing yourself — Code: SEL
Brief definition	The choice and commitment to change in being physical active and also believe in it.
Full definition	SEL is a process that helps individuals in their choice and commitment to be physical active and their believe to stay active.
Practical def.	Encourage the subject to make (and keep) promises, plans, and commitments to yourself to be active on short term (mostly behavioral process and a bit cognitive process)
Perspectives	To start/trigger the process: messages that instigate you to make a commitment (to someone/yourself) or set a goal for exercise or messages that drive you to accept responsibility for starting to exercise. To advance the process: messages that help and remind you of your commitment (to someone/yourself) or your goal for exercise. To distinguish: Perspective is usually on a short-term goal
Inclusion	Start exercise: keywords like that say "goal" or "commitment" Keep exercising: "don't give up", "you can do it", "stay motivated" etc. or paraphrases of this.
Exclusion	Sentences where the "don't give up" and others are just 'added' and not the main topic
Example incl.	"Don't give up, you can do it!"; "Keep it up!"; "Keep going"
Example excl.	"Don't give up, you're dreams about to come true" - would fit better in RM because of the (future) reward being the most important part of the message. "Get up and do it" - it contains no commitment or promise, so it fits better in SC.

Process	Helping Relationships/Enlisting social support — Code: HR
Brief definition	Support from others for being physical active
Full definition	HR is a process that involves helping the individual during attempts to be or to stay physical active by giving support.
Practical def.	Encourage the subject to find a family member, friend, or coworker who is willing and able to provide support for being active or be the support (behavioral process and cognitive process)
Perspectives	Direct messages: messages that directly function as a “friend”, by being personal and caring Indirect messages: messages that suggest you to engage in a ‘helping relationship’, by seeking support of friends, or suggest to do something with friends. To distinguish: Messages in this category can usually also fit in all other categories, but fit into HR because the messages are more personal, very often with “I” added in the sentence.
Inclusion	A personal tone; Feeling understood: Seems like as a friends say it; Mostly contains the word “I”, like: I know it’s hard. A compliment from “I” perspective
Exclusion	-
Example incl.	“I’m so proud of you!” , “I know it’s hard, but I think you’re doing really great!”, “Play some sports with friends”
Example excl.	“you’re doing great” - This would fit better in RM because it lacks a personal touch (e.g. “I think you’re doing great”).

Process	Counter Conditioning/Substituting alternatives — Code: CC
Brief definition	Learn to substitute alternative healthy behavior for problem behavior
Full definition	CC is a process that helps individuals to learn and seek to ways of being physical active when encountering barriers.
Practical def.	Encourage the subject to participate in physical activity when she is tired, stressed, or unlikely to want to be physically active (behavioral process and cognitive process)
Perspectives	Direct messages: messages that *suggest* that you could replace a bad behavior with a good behavior. Or messages that suggest you could do the good behavior (to replace the bad behavior can be implicit). Indirect messages: messages that *suggest* ways to overcome or avoid certain barriers (like time, repetitiveness, muscle ache) that prevent you from doing exercise. To distinguish: The suggestions can be on learning behavior (behavior change) on the long-term, so this can be cognitive. Or behavioral, when suggesting good behaviors.
Inclusion	Advise for being physical active when encountering barriers; keywords can be “time” and “don’t feel like X, do Y” perspective is usually on long-term
Exclusion	Imperatives
Example incl.	“Have fun with it, don’t make exercising a boring chore” “How about going to swim? That sounds fun!”
Example excl.	“Park the car further away from the building and walk to it” - because this is more a behavior reminder, it fits better in SC.

Process	Reinforcement Management/Rewarding yourself — Code: RM
Brief definition	Consequences for doing healthy behavior
Full definition	RM is a process that provides consequences for doing physical activities.
Practical def.	Encourage the subject to praise himself and reward himself for being physically active or be the praise and reward (behavioral process)
Perspectives	Direct messages: messages that directly function as a reward (like a compliment) Indirect messages: messages that suggest you should get or deserve a reward or that suggest you a future reward (if you hold on).
Inclusion	Compliments/Rewards after being physical active; Compliments after good results; Positive; Advice with long term reward.
Exclusion	Almost no messages when it is obvious the individual has not started yet, because there is nothing to 'reinforce'/reward.
Example incl.	"You look great!!", "Well done!", "You are making great improvements", "Keep up the good work"
Example excl.	"I'm so proud! Keep it up." - Because the first part is quite personal, this would fit more in HR

Process	Stimulus Control/Reminding yourself — Code: SC
Brief definition	Control of situations and other causes that trigger for not doing activities
Full definition	SC is a process whereby the individual needs to remove or add stimuli for doing physical activities.
Practical def.	Tell the subject to set up reminders to be active, or be the reminder (behavioral process)
Perspectives	Direct messages: messages that instigate you to do some exercise in the very near future. ("now", "today", "tomorrow": the message serves as a cue from the 'environment') Indirect messages: messages that help and remind to reshape your environment so as to provide positive cues or remove negative cues.
Inclusion	Reminders for doing activities; Usually no arguments; Imperatives for doing activities
Exclusion	Arguments why they should do activities, with less/no focus on the near-future reminder. A commitment or promise (this fit more in SEL).
Example incl.	"Just do something! Go for a run! Go for a walk! Put your shoes on!" "Get moving! Don't let no exercise be your downfall!" - (there is enough reminder that it fits SC, and not enough goal/commit that it fits SEL)
Example excl.	"You don't need to be able to run a marathon today, but you do need to exercise daily" - It is not a clear reminder, and the focus is on long-term behavior, so it fits better in CC

B Fifty peer-designed motivational messages used for evaluation survey

Motivational Message	Cat.	α
If you exercise, your quality of life will be much higher than it is now.	CR	
Exercise helps lower your risk of heart attack and stoke.	CR	
Exercise helps keep blood sugar, and blood pressure under control. Those who regularly work out are 3 times less likely to develop these problems.	CR	
Exercise will help clear your mind and reduce stress.	CR	
Regular exercise will keep you strong and raise your stamina levels.	CR	0.76
Unless some changes are made in your weight you could risk getting a heart attack,stroke or something else that could effect your life.	DR	
You need to exercise before it takes a toll on your body!	DR	
Start working out before its too late, you're not getting any younger!	DR	
Don't let yourself get old and regret not taking care of your body and your health. Keep at it every day and remember what you're working for.	DR	
It's easier to wake up early in the morning and workout, than it is to look in the mirror and not like what you see.	DR	0.82
You owe it to your family and friends to take care of yourself.	ER	
Take care of yourself so your health doesn't become a burden on other people.	ER	
Your loved ones want you to be around so you should exercise to get healthy.	ER	
Your friends and family are counting on you!	ER	
Remember, you are doing this for your friends and family as much as yourself.	ER	0.88
The local gym has lots of fun classes you can check out; you'll get fit, and meet new people!	SOL	
Try some social exercising. Take a yoga or pilates class. Make some friends who tempt you to have good habits this week.	SOL	
Look online for a good beginner workout.	SOL	
You're not alone! Tons of people are working to exercise more frequently!	SOL	
If you want to start exercising, you could always go to a gym; you can preserve your routine regardless of the weather.	SOL	0.73
Think about all of the benefits of becoming healthy.	SR	
Imagine being in the best shape of your life with a long future ahead of you.	SR	
In 10 years, will you be glad you watched television, or upset that you didn't exercise?	SR	
You will be amazed at the changes you will see in yourself after exercising regularly!	SR	
Imagine what you'll look like next year if you start now	SR	0.73
Today is the day to get moving and get healthy	SEL	
Decide today to do some exercise.	SEL	

This will get easier the more you do it. Keep at it!	SEL	
The hardest thing is always starting, you can do this!	SEL	
You can and will achieve your fitness goals!	SEL	0.72
I know how hard it is to find time to exercise, but think of how much time you will save by being stronger and more efficient.	HR	
You should exercise more. I know it's hard, but it's worth it to keep going. You'll feel so much better in the long term is you don't give up now.	HR	
I believe in you; you can incorporate exercise into your daily life.	HR	
I'm proud of you for wanting to be healthy, keep at it.	HR	
I'm proud of you for wanting to get in shape.	HR	0.74
Activity doesn't have to be a chore, find something fun to do!	CC	
Find something active yet fun to do today, so that you'll enjoy your exercise more.	CC	
Don't forget to exercise often, it's more rewarding than watching TV.	CC	
Think of exercise as a way to let go and relax your mind.	CC	
Try a new location to jog in. Keep your mind active while you exercise.	CC	0.68
You are doing a good job, keep up the good lifestyle.	RM	
You're doing great- you should be proud of yourself!	RM	
Congratulations on working towards a healthier you!	RM	
Be proud of yourself for sticking with the plan.	RM	
Don't forget to stop and appreciate your own hard work today.	RM	0.85
Find at least 30 minutes in your schedule today for some exercise.	SC	
Get out there and exercise!	SC	
Always plan your daily activities allowing room for your exercise.	SC	
Schedule your workouts on your calendar.	SC	
Have you worked out yet today? Now's the perfect time! Get up, get going!	SC	0.73

Table B.1: The fifty peer-designed messages evaluated and their coding and alpha.

C | Fifty expert-designed motivational messages used for evaluation survey

Motivational Message	Cat.	α
Regular exercise is important for you to stay healthy.	CR	
Regular exercise will help you to better handle all the stress factors in your life.	CR	
After a longer time of regular exercising you will see how your body changes in a positive way.	CR	
You will feel so much better with some regular exercise in your life.	CR	
Regular exercise will make you less vulnerable for diseases.	CR	0.73
Think about the physical disadvantages that result from your inactivity.	DR	
How would you feel in a year if you do not start to exercise now?	DR	
Are you done doing nothing?	DR	
Are you still hesitating? Pretty soon time will run against you!	DR	
The longer you push exercise aside, the longer you are denying yourself a better quality of life.	DR	0.78
How about joing your friends that are already exercising? It will not only improve your own health, but also your social relationships.	ER	
Aren't your friends already envying you because of your discipline?	ER	
Regular exercise with friends does not only improve your health, but also deepens your social relationships.	ER	
Work out with friends. You will not only stay healthy, but also improve your social life.	ER	
Have you talked to a friend about this decision?	ER	0.75
Look around in your neighbourhood if you can find any facilities to do some sports.	SOL	
Keep an eye out for some nice areas in your district where you can go jogging.	SOL	
Keep an eye out for a gym or jogging route that is close by.	SOL	
Try out different sports, and then think about which one you liked the most. Is there any club nearby which offers this sport?	SOL	
Think of a sport that you personally enjoy the most - running, swimming, tennis...what do you like?	SOL	0.82
Wouldn't it be worth it for you own well-being to start exercising?	SR	
Imagine how much better you would feel with some regular exercise in your life!	SR	
Think about all the positive consequences that will result from your exercise.	SR	
Think of all the great results you could achieve within a year if you start now to implement fitness in your daily life.	SR	
If you start working out today you will see results within two weeks! And wouldn't that be a great feeling?	SR	0.75
You can do it - start your journey into a more active and healthy life.	SEL	

Try to set daily goals that you can reach. A bit of exercising (e.g. going for a strong walk) is always better than none.	SEL	
Keep on exercising until it will become a steady aspect of your life!	SEL	
Start now! The beginning is always the most difficult part of the whole journey!	SEL	
It's time to finally change your thoughts into action! Start with some easy and slow exercises.	SEL	0.69
I promise you that you will love working out!	HR	
I promise it's going to pay off afterwards!	HR	
I have trust in you - you can do it!	HR	
I am sure that you can do it, but to get there you have to overcome yourself and start with the first steps.	HR	
I am sure you would feel much better with a little exercise in your life.	HR	0.72
Think of something that could make exercising even more fun for you! Try it out the next time!	CC	
Don't quit! Starting again from the beginning is more difficult than continuing.	CC	
If you keep on going, those exercises will get much easier for you in the future.	CC	
Make exercise your hobby! If you have fun while doing it, it will get easier to start.	CC	
Don't forget to enjoy working out.	CC	0.68
Wow, you are doing very well! Hold on!	RM	
Good work, you already managed the most important part. Now keep on going.	RM	
You can be proud of yourself.	RM	
It is impressive how good you are doing with incorporating regular exercise into your daily life.	RM	
You have come this far, you should be proud of your achievements.	RM	0.83
Set fixed times in the week for physical exercise.	SC	
Have you already worked out today? No? - Then what are you waiting for?	SC	
Start to make a weekly agenda to set fixed times for your exercises.	SC	
What are you doing right now? What about going for a run?	SC	
Have you already planned your next exercise session? Planning helps to keep on track.	SC	0.70

Table C.1: The fifty expert-designed messages evaluated and their coding and alpha.

D | Interaction effects for peer- and expert-designed messages comparison

Fixed effects	Estimate	Std. Error	t-value
(Intercept)	**3.436**^{3}	0.152	22.622
Contemplation:Designer (peer)	0.106	0.227	0.466
Preparation:Designer (peer)	0.299	0.220	1.357
Action:Designer (peer)	0.326	0.230	1.420
Maintenance:Designer (peer)	0.306	0.213	1.434

Table D.1: The estimates and standard error of the two way interaction effects between: designer of the message and stage of the participant. Significant effects reported for $^{1}p < 0.05$, $^{2}p < 0.01$, $^{3}p < 0.001$.

Fixed effects	Estimate	Std. Error	t-value
(Intercept)	**3.436**^{3}	0.152	22.622
DR:Designer (peer)	0.298	0.177	1.678
ER:Designer (peer)	**-0.416**^{1}	0.177	-2.344
SOL:Designer (peer)	0.010	0.177	0.057
SR:Designer (peer)	0.047	0.177	0.262
SEL:Designer (peer)	-0.117	0.177	-0.657
HR:Designer (peer)	0.246	0.177	1.386
CC:Designer (peer)	-0.075	0.177	-0.422
RM:Designer (peer)	**-0.384**^{1}	0.177	-2.163
SC:Designer (peer)	-0.149	0.177	-0.838

Table D.2: The estimates and standard error of the two way interaction effects between: process-categories and designer of the message. Significant effects reported for $^{1}p < 0.05$, $^{2}p < 0.01$, $^{3}p < 0.001$.

Fixed effects	Estimate	Std. Error	t-value
(Intercept)	**3.436**^{3}	0.152	22.622
DR:Contemplation	0.073	0.139	0.521
ER:Contemplation	**-0.296**^{1}	0.139	-2.126
SOL:Contemplation	**-0.317**^{1}	0.139	-2.273
SR:Contemplation	0.131	0.139	0.939
SEL:Contemplation	0.021	0.139	0.153
HR:Contemplation	0.005	0.139	0.037
CC:Contemplation	-0.040	0.139	-0.285
RM:Contemplation	-0.092	0.139	-0.662
SC:Contemplation	-0.130	0.139	-0.933
DR:Preparation	-0.009	0.134	-0.070
ER:Preparation	**-0.374**^{2}	0.134	-2.786
SOL:Preparation	**-0.340**^{1}	0.134	-2.539
SR:Preparation	0.161	0.134	1.198
SEL:Preparation	-0.071	0.134	-0.527
HR:Preparation	-0.039	0.134	-0.292
CC:Preparation	-0.072	0.134	-0.538
RM:Preparation	-0.105	0.134	-0.785
SC:Preparation	-0.209	0.134	-1.562
DR:Action	-0.084	0.134	-0.626
ER:Action	**-0.309**^{1}	0.134	-2.311
SOL:Action	**-0.430**^{2}	0.134	-3.208
SR:Action	0.135	0.134	1.005
SEL:Action	-0.075	0.134	-0.558
HR:Action	-0.056	0.134	-0.417
CC:Action	-0.015	0.134	-0.111
RM:Action	-0.168	0.134	-1.258
SC:Action	-0.180	0.134	-1.341
DR:Maintenance	0.096	0.127	0.751
ER:Maintenance	**-0.279**^{1}	0.127	-2.191
SOL:Maintenance	-0.228	0.127	-1.790
SR:Maintenance	0.038	0.127	0.298
SEL:Maintenance	-0.157	0.127	-1.235
HR:Maintenance	-0.092	0.127	-0.720
CC:Maintenance	-0.066	0.127	-0.517
RM:Maintenance	-0.248	0.127	1.950
SC:Maintenance	-0.159	0.127	-1.247

Table D.3: The estimates and standard error of the two way interaction effects between: process-categories and stage of the participant. Significant effects reported for $^{1}p < 0.05$, $^{2}p < 0.01$, $^{3}p < 0.001$.

Fixed effects	Estimate	Std. Error	t-value
(Intercept)	**3.436**^{3}	0.152	22.622
DR:Contemplation:Designer (peer)	-0.143	0.203	-0.704
ER:Contemplation:Designer (peer)	**0.452**^{1}	0.203	2.223
SOL:Contemplation:Designer (peer)	-0.048	0.203	-0.236
SR:Contemplation:Designer (peer)	-0.126	0.203	-0.618
SEL:Contemplation:Designer (peer)	0.014	0.203	0.068
HR:Contemplation:Designer (peer)	-0.041	0.203	-0.201
CC:Contemplation:Designer (peer)	-0.044	0.203	-0.219
RM:Contemplation:Designer (peer)	0.301	0.203	1.481
SC:Contemplation:Designer (peer)	0.084	0.203	0.415
DR:Preparation:Designer (peer)	-0.036	0.197	-0.181
ER:Preparation:Designer (peer)	0.246	0.197	1.248
SOL:Preparation:Designer (peer)	-0.169	0.197	-0.856
SR:Preparation:Designer (peer)	-0.340	0.197	-1.724
SEL:Preparation:Designer (peer)	-0.087	0.197	-0.441
HR:Preparation:Designer (peer)	-0.162	0.197	-0.822
CC:Preparation:Designer (peer)	-0.171	0.197	-0.865
RM:Preparation:Designer (peer)	0.041	0.197	0.208
SC:Preparation:Designer (peer)	0.027	0.197	0.137
DR:Action:Designer (peer)	0.187	0.206	0.910
ER:Action:Designer (peer)	**0.463**^{1}	0.206	2.250
SOL:Action:Designer (peer)	0.121	0.206	0.590
SR:Action:Designer (peer)	-0.112	0.206	-0.547
SEL:Action:Designer (peer)	0.006	0.206	0.029
HR:Action:Designer (peer)	-0.153	0.206	-0.742
CC:Action:Designer (peer)	-0.108	0.206	-0.525
RM:Action:Designer (peer)	0.180	0.206	0.878
SC:Action:Designer (peer)	0.226	0.206	1.099
DR:Maintenance:Designer (peer)	-0.074	0.191	-0.390
ER:Maintenance:Designer (peer)	0.373	0.191	1.952
SOL:Maintenance:Designer (peer)	-0.112	0.191	-0.586
SR:Maintenance:Designer (peer)	-0.038	0.191	-0.197
SEL:Maintenance:Designer (peer)	0.128	0.191	0.672
HR:Maintenance:Designer (peer)	-0.066	0.191	-0.344
CC:Maintenance:Designer (peer)	0.020	0.191	0.106
RM:Maintenance:Designer (peer)	0.269	0.191	1.409
SC:Maintenance:Designer (peer)	0.188	0.191	0.983

Table D.4: The estimates and standard error of the three way interaction effects between: process-categories, designer of the message, and stage of the participant. Significant effects reported for $^1p < 0.05$, $^2p < 0.01$, $^3p < 0.001$.

E | Details on messages used for in-the-wild experiment

Category	PC	C	P	A	M	Total
Random condition						
CR	11 (8/3)	28 (24/4)	29 (15/14)	69 (48/21)	61 (34/27)	198 (129/69)
DR	6 (6/0)	35 (22/13)	24 (13/11)	50 (17/33)	60 (25/35)	175 (83/92)
ER	14 (1/13)	41 (23/18)	30 (15/15)	57 (17/40)	72 (21/51)	214 (77/137)
SOL	4 (1/3)	17 (8/9)	9 (7/2)	18 (7/11)	22 (9/13)	70 (32/38)
SR	12 (8/4)	32 (23/9)	32 (24/8)	62 (39/23)	62 (39/23)	200 (133/67)
SEL	15 (5/10)	33 (27/6)	28 (19/9)	51 (24/27)	49 (34/15)	176 (109/67)
HR	13 (0/13)	42 (19/23)	34 (14/20)	59 (28/31)	53 (28/25)	201 (89/112)
CC	11 (6/5)	37 (26/11)	31 (20/11)	69 (35/34)	65 (37/28)	213 (124/89)
RM	9 (3/6)	29 (26/3)	48 (31/17)	61 (38/23)	62 (49/13)	209 (147/62)
SC	4 (2/2)	20 (10/10)	27 (13/14)	36 (16/20)	20 (10/10)	107 (51/56)
Total	99 (40/59)	314 (208/106)	292 (171/121)	532 (269/263)	526 (286/240)	1763 (974/789)
Tailored condition						
CR	5 (3/2)	15 (5/10)	11 (4/7)	12 (5/7)	18 (8/10)	61 (25/36)
DR	30 (10/20)	66 (15/51)	5 (1/4)	7 (2/5)	10 (4/6)	118 (32/86)
ER	30 (4/26)	70 (10/60)	13 (6/7)	10 (3/7)	21 (11/10)	144 (34/110)
SOL	27 (7/20)	57 (11/46)	4 (1/3)	1 (1/0)	3 (0/3)	92 (20/72)
SR	4 (2/2)	71 (40/31)	19 (11/8)	16 (8/8)	25 (13/12)	135 (74/61)
SEL	2 (2/0)	6 (2/4)	110 (66/44)	22 (13/9)	12 (6/6)	152 (89/63)
HR	2 (2/0)	7 (3/4)	138 (76/62)	115 (60/55)	19 (12/7)	281 (153/128)
CC	2 (2/0)	5 (2/3)	150 (80/70)	129 (66/63)	28 (15/13)	314 (165/149)
RM	4 (4/0)	4 (0/4)	12 (8/4)	112 (64/48)	187 (160/27)	319 (236/83)
SC	2 (1/1)	3 (1/2)	7 (1/6)	111 (53/58)	222 (90/132)	345 (146/199)
Total	108 (37/71)	304 (89/215)	469 (254/215)	535 (275/260)	545 (319/226)	1961 (974/987)
Both conditions						
CR	16 (11/5)	43 (29/14)	40 (19/21)	81 (53/28)	79 (42/37)	259 (154/105)
DR	36 (16/20)	101 (37/64)	29 (14/15)	57 (19/38)	70 (29/41)	293 (115/178)
ER	44 (5/39)	111 (33/78)	43 (21/22)	67 (20/47)	93 (32/61)	358 (111/247)
SOL	31 (8/23)	74 (19/55)	13 (8/5)	19 (8/11)	25 (9/16)	162 (52/110)
SR	16 (10/6)	103 (63/40)	51 (35/16)	78 (47/31)	87 (52/35)	335 (207/128)
SEL	17 (7/10)	39 (29/10)	138 (85/53)	73 (37/36)	61 (40/21)	328 (198/130)
HR	15 (2/13)	49 (22/27)	172 (90/82)	174 (88/86)	72 (40/32)	482 (242/240)
CC	13 (8/5)	42 (28/14)	181 (100/81)	198 (101/97)	93 (52/41)	527 (289/238)
RM	13 (7/6)	33 (26/7)	60 (39/21)	173 (102/71)	249 (209/40)	528 (383/145)
SC	6 (3/3)	23 (11/12)	34 (14/20)	147 (69/78)	242 (100/142)	452 (197/255)
Total	207 (77/130)	618 (297/321)	761 (425/336)	1067 (544/523)	1071 (605/466)	3724 (1948/1776)

Table E.1: The number of messages and how they are rated (motivating/not motivating), per category, per stage, and per condition.

Bibliography

Adams, J. and White, M. (2004). Why don't stage-based activity promotion interventions work? *Health Education Research*, 20(2):237–243. (Cited on page 111.)

Adnan, M., Mukhtar, H., and Naveed, M. (2012). Persuading students for behavior change by determining their personality type. *15th International Multitopic Conference*, pages 439–449. (Cited on page 23.)

Ajzen, I. (2002). Perceived behavioral control, self-efficacy, locus of control, and the theory of planned behavior. *Journal of Applied Social Psychology*, 32(4):665–683. (Cited on page 3.)

Ajzen, I. (2005). *Attitudes, personality, and behavior*. McGraw-Hill Education (UK). (Cited on page 22.)

Albaina, I. M., Visser, T., van der Mast, C. A., and Vastenburg, M. H. (2009). Flowie: A persuasive virtual coach to motivate elderly individuals to walk. In *3rd International Conference on Pervasive Computing Technologies for Healthcare*, pages 1–7. IEEE. (Cited on page 15.)

Alkış, N. and Temizel, T. T. (2015). The impact of individual differences on influence strategies. *Personality and Individual Differences*, 87:147–152. (Cited on pages 23, 26, and 61.)

Anderson, I., Maitland, J., Sherwood, S., Barkhuus, L., Chalmers, M., Hall, M., Brown, B., and Muller, H. (2007). Shakra: tracking and sharing daily activity levels with unaugmented mobile phones. *Mobile Networks and Applications*, 12(2-3):185–199. (Cited on page 18.)

Arteaga, S. M., Kudeki, M., Woodworth, A., and Kurniawan, S. (2010). Mobile system to motivate teenagers' physical activity. In *Proceedings of the 9th International Conference on Interaction Design and Children*, pages 1–10. ACM. (Cited on page 23.)

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48. (Cited on pages 52, 75, and 98.)

Bauman, A. E., Reis, R. S., Sallis, J. F., Wells, J. C., Loos, R. J. F., and Martin, B. W. (2012). Correlates of physical activity: why are some people physically active and others not? *Lancet*, 380(9838):258–71. (Cited on page 22.)

Bauman, A. E., Sallis, J. F., Dzewaltowski, D. A., and Owen, N. (2002). Toward a better understanding of the influences on physical activity: the role of determinants, correlates, causal variables, mediators, moderators, and confounders. *American Journal of Preventive Medicine*, 23(2):5–14. (Cited on page 22.)

Bell, M., Chalmers, M., Barkhuus, L., Hall, M., Sherwood, S., Tennent, P., Brown, B., Rowland, D., Benford, S., Capra, M., et al. (2006). Interweaving mobile games with everyday life. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 417–426. ACM. (Cited on page 12.)

Benisovich, S. V., Rossi, J. S., Norman, G. J., and Nigg, C. R. (1998). A multidimensional approach to exercise self-efficacy: Relationship with exercise behaviour and attitudes towards exercise. In *Annual Meeting of the New England Psychology Association*. (Cited on pages 35, 51, and 90.)

Berkovsky, S., Coombe, M., Freyne, J., Bhandari, D., and Baghaei, N. (2010). Physical activity motivating games: virtual rewards for real activity. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 243–252. ACM. (Cited on page 14.)

Bernard, P., Carayol, M., Gourlan, M., Boiché, J., Romain, A. J., Bortolon, C., Lareyre, O., and Ninot, G. (2017). Moderators of theory-based interventions to promote physical activity in 77 randomized controlled trials. *Health Education & Behavior*, 44(2):227–235. (Cited on page 87.)

Biddle, S. J. H. and Mutrie, N. (2007). *Psychology of physical activity: Determinants, well-being and interventions*. Routledge. (Cited on page 3.)

Blair, S. N. and Brodney, S. (1999). Effects of physical inactivity and obesity on morbidity and mortality: current evidence and research issues. *Medicine and Science in Sports and Exercise*, 31:S646–S662. (Cited on page 1.)

Blair, S. N., Kohl, H., Barlow, C. E., Paffenbarger, R., Gibbons, L. W., and Macera, C. A. (1995). Changes in physical fitness and all-cause mortality. *Journal of the American Medical Association*, 273(14):1093–1098. (Cited on page 1.)

Bogg, T. (2008). Conscientiousness, the transtheoretical model of change, and exercise: a neo-socioanalytic integration of trait and social-cognitive frameworks in the prediction of behavior. *Journal of Personality*, 76(4):775–802. (Cited on page 44.)

Bridle, C., Riemsma, R. P., Pattenden, J., Sowden, A. J., Mather, L., Watt, I. S., and Walker, A. (2005). Systematic review of the effectiveness of health behavior interventions based on the transtheoretical model. *Psychology & Health*, 20(3):283–301. (Cited on page 111.)

Brug, J., Conner, M., Harre, N., Kremers, S., McKellar, S., and Whitelaw, S. (2005). The transtheoretical model and stages of change: a critique observations by five commentators on the paper by adams, j and white, m (2004) why don't stage-based activity promotion interventions work? *Health Education Research*, 20(2):244–258. (Cited on page 111.)

Burnham, K. P. and Anderson, D. R. (2004). Multimodel inference understanding aic and bic in model selection. *Sociological Methods & Research*, 33(2):261–304. (Cited on page 53.)

Busch, M., Mattheiss, E., Reisinger, M., Orji, R., Fröhlich, P., and Tscheligi, M. (2016). More than sex: The role of femininity and masculinity in the design of personalized persuasive games. In *Proceedings of the 11th International Conference on Persuasive Technology*, pages 219–229. Springer. (Cited on pages 24 and 62.)

Callison-Burch, C. and Dredze, M. (2010). Creating speech and language data with amazon's mechanical turk. In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*, pages 1–12. (Cited on page 30.)

Campbell, T., Ngo, B., and Fogarty, J. (2008). Game design principles in everyday fitness applications. In *Proceedings of the 2008 ACM conference on Computer Supported Cooperative Work*, pages 249–252. ACM. (Cited on page 14.)

Cheng, J., Teevan, J., Iqbal, S. T., and Bernstein, M. S. (2015). Break it down: A comparison of macro-and microtasks. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 4061–4064. ACM. (Cited on page 8.)

Cialdini, R. B. (2001). Influence: Science and practice. *Boston: Allyn & Bacon*. (Cited on pages 15 and 23.)

Cole-Lewis, H. and Kershaw, T. (2010). Text messaging as a tool for behavior change in disease prevention and management. *Epidemiologic Reviews*, 32(1):56–69. (Cited on pages 4 and 24.)

Coley, H. L., Sadasivam, R. S., Williams, J. H., Volkman, J. E., Schoenberger, Y.-M., Kohler, C. L., Sobko, H., Ray, M. N., Allison, J. J., Ford, D. E., Gilbert, G. H., and Houston, T. K. (2013). Crowdsourced peer- versus expert-written smoking-cessation messages. *American Journal of Preventive Medicine*, 45(5):543–50. (Cited on pages 6, 30, 32, 68, and 81.)

Consolvo, S., Everitt, K., Smith, I., and Landay, J. A. (2006). Design requirements for technologies that encourage physical activity. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 457–466. ACM. (Cited on page 17.)

Consolvo, S., Klasnja, P., McDonald, D. W., Avrahami, D., Froehlich, J., LeGrand, L., Libby, R., Mosher, K., and Landay, J. A. (2008a). Flowers or a robot army?: encouraging awareness & activity with personal, mobile displays. In *Proceedings of the 10th International Conference on Ubiquitous Computing*, pages 54–63. ACM. (Cited on page 17.)

Consolvo, S., Klasnja, P., McDonald, D. W., and Landay, J. A. (2009a). Goal-setting considerations for persuasive technologies that encourage physical activity. In *Proceedings of the 4th international Conference on Persuasive Technology*, page 8. ACM. (Cited on pages 17 and 18.)

Consolvo, S., Landay, J. A., and McDonald, D. W. (2009b). Designing for behavior change in everyday life. *focus*, 405:414. (Cited on page 17.)

Consolvo, S., McDonald, D. W., and Landay, J. A. (2009c). Theory-driven design strategies for technologies that support behavior change in everyday life. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 405–414. ACM. (Cited on page 17.)

Consolvo, S., McDonald, D. W., Toscos, T., Chen, M. Y., Froehlich, J., Harrison, B., Klasnja, P., LaMarca, A., LeGrand, L., Libby, R., et al. (2008b). Activity sensing in the wild: a field trial of ubifit garden. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1797–1806. ACM. (Cited on page 17.)

Costa, P. T. and McCrae, R. R. (1992). Normal personality assessment in clinical practice: The neo personality inventory. *Psychological Assessment*, 4(1):5. (Cited on pages 35, 51, 74, and 90.)

Courneya, K. S. and Hellsten, L.-A. M. (1998). Personality correlates of exercise behavior, motives, barriers and preferences: An application of the five-factor model. *Personality and Individual Differences*, 24(5):625–633. (Cited on page 22.)

Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., and Petticrew, M. (2008). Developing and evaluating complex interventions: the new medical research council guidance. *BMJ*, 337:a1655. (Cited on page 63.)

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches*. SAGE Publications, Incorporated. (Cited on page 87.)

Dahl, D. B. (2016). *xtable: Export Tables to LaTeX or HTML*. R package version 1.8-2. (Cited on pages 52 and 75.)

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, pages 319–340. (Cited on page 90.)

de Vries, R. A. J., Truong, K. P., and Evers, V. (2016a). Crowd-designed motivation: Combining personality and the transtheoretical model. In *Proceedings of the 11th International Conference on Persuasive Technology*, pages 41–52. Springer. (Cited on pages 29 and 153.)

de Vries, R. A. J., Truong, K. P., Kwint, S., Drossaert, C. H. C., and Evers, V. (2016b). Crowd-designed motivation: Motivational messages for exercise adherence based on behavior change theory. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 297–308. ACM. (Cited on pages 29, 47, and 153.)

de Vries, R. A. J., Truong, K. P., Zaga, C., Li, J., and Evers, V. (2017a). A word of advice: how to tailor motivational text messages based on behavior change theory to personality and gender. *Personal and Ubiquitous Computing*, 21(4):675–687. (Cited on pages 47 and 153.)

de Vries, R. A. J., Zaga, C., Bayer, F., Drossaert, C. H. C., Truong, K. P., and Evers, V. (2017b). Experts get me started, peers keep me going: Comparing crowd-versus expert-designed motivational text messages for exercise behavior change. In *11th EAI International Conference on Pervasive Computing Technologies for Healthcare*, pages 155–162. ACM. (Cited on pages 67 and 153.)

Dishman, R. K. and Sallis, J. F. (1994). Determinants and interventions for physical activity and exercise. *Physical Activity, Fitness, and Health*. (Cited on page 21.)

Dow, S. P., Glassco, A., Kass, J., Schwarz, M., Schwartz, D. L., and Klemmer, S. R. (2010). Parallel prototyping leads to better design results, more divergence, and increased self-efficacy. *ACM Transactions on Computer-Human Interaction*, 17(4):18. (Cited on page 30.)

Dusseldorp, E., Van Genugten, L., van Buuren, S., Verheijden, M. W., and van Empelen, P. (2014). Combinations of techniques that effectively change health behavior: Evidence from meta-cart analysis. *Health Psychology*, 33(12):1530. (Cited on page 63.)

Ferron, M. and Massa, P. (2013). Transtheoretical model for designing technologies supporting an active lifestyle. In *Proceedings of the Biannual Conference of the Italian Chapter of SIGCHI*, page 7. ACM. (Cited on pages 26 and 31.)

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Sage. (Cited on pages 39 and 71.)

Fiuza-Luces, C., Garatachea, N., Berger, N. A., and Lucia, A. (2013). Exercise is the real polypill. *Physiology*, 28(5):330–358. (Cited on page 1.)

Fogg, B. J. (2003). *Persuasive technology: using computers to change what we think and do*. San Francisco: Morgan Kaufmann Publishers. (Cited on pages 4, 15, and 24.)

Fogg, B. J. (2009). A behavior model for persuasive design. In *Proceedings of the 4th International Conference on Persuasive Technology*, page 40. ACM. (Cited on page 2.)

Fujiki, Y., Kazakos, K., Puri, C., Buddharaju, P., Pavlidis, I., and Levine, J. (2008). Neat-o-games: blending physical activity and fun in the daily routine. *Computers in Entertainment (CIE)*, 6(2):21. (Cited on page 14.)

Gabbiadini, A., Sagioglou, C., and Greitemeyer, T. (2018). Does pokémon go lead to a more physically active life style? *Computers in Human Behavior*, 84:258–263. (Cited on page 3.)

Gallagher, P., Yancy, W. S., Denissen, J. J. A., Kühnel, A., and Voils, C. I. (2013). Correlates of daily leisure-time physical activity in a community sample: Narrow personality traits and practical barriers. *Health Psychology*, 32(12):1227–35. (Cited on pages 22 and 32.)

Geller, K. S., Nigg, C. R., Motl, R. W., Horwath, C., and Dishman, R. K. (2012). Transtheoretical model constructs for physical activity behavior are invariant across time among ethnically diverse adults in hawaii. *Psychology of Sport and Exercise*, 13(5):606–613. (Cited on page 87.)

Gil-Castiñeira, F., Fernández-López, A., Bravo, C. L., Cid-Vieytes, N., Conde-Lagoa, D., Costa-Montenegro, E., and González-Castaño, F. J. (2011). Runwithus: a social sports application in the ubiquitous oulu environment. In *Proceedings of the 10th International Conference on Mobile and Ubiquitous Multimedia*, pages 195–204. ACM. (Cited on page 12.)

Glanz, K. (2015). *Health behavior: Theory, research, and practice*. John Wiley & Sons. (Cited on page 24.)

Gneezy, U., Meier, S., and Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *Journal of Economic Perspectives*, 25(4):191–210. (Cited on page 3.)

Godin, G. and Shephard, R. J. (1997). Godin leisure-time exercise questionnaire. *Medicine & Science in Sports & Exercise*, 29(6s):S36. (Cited on pages 35 and 90.)

Goldberg, L. R. (1992). The development of markers for the big-five factor structure. *Psychological Assessment*, 4(1):26. (Cited on page 22.)

Graether, E. and Mueller, F. F. (2012). Joggobot: a flying robot as jogging companion. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems*, pages 1063–1066. ACM. (Cited on pages 10 and 11.)

Green, L. W. (1970). Manual for scoring socioeconomic status for research on health behavior. *Public Health Reports*, 85(9):815. (Cited on pages 35, 50, and 74.)

Guest, G. and MacQueen, K. M. (2007). *Handbook for team-based qualitative research*. Rowman Altamira. (Cited on pages 36, 62, and 71.)

Halko, S. and Kientz, J. A. (2010). Personality and persuasive technology: an exploratory study on health-promoting mobile applications. pages 150–161. (Cited on pages 23 and 44.)

Heath, G. W., Parra, D. C., Sarmiento, O. L., Andersen, L. B., Owen, N., Goenka, S., Montes, F., and Brownson, R. C. (2012). Evidence-based intervention in physical activity: lessons from around the world. *Lancet*, 380(9838):272–81. (Cited on page 87.)

Heer, J. and Bostock, M. (2010). Crowdsourcing graphical perception: using mechanical turk to assess visualization design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 203–212. ACM. (Cited on page 30.)

Hekler, E. B., Klasnja, P., Froehlich, J. E., and Buman, M. P. (2013). Mind the theoretical gap: interpreting, using, and developing behavioral theory in hci research. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3307–3316. ACM. (Cited on pages 2 and 4.)

Hekler, E. B., Michie, S., Pavel, M., Rivera, D. E., Collins, L. M., Jimison, H. B., Garnett, C., Parral, S., and Spruijt-Metz, D. (2016). Advancing models and theories for digital behavior change interventions. *American Journal of Preventive Medicine*, 51(5):825–832. (Cited on page 116.)

Hermawati, S. and Lawson, G. (2014). Managing obesity through mobile phone applications: a state-of-the-art review from a user-centred design perspective. *Personal and Ubiquitous Computing*, pages 1–21. (Cited on page 9.)

Hirsh, J. B., Kang, S. K., and Bodenhausen, G. V. (2012). Personalized persuasion: tailoring persuasive appeals to recipients' personality traits. *Psychological Science*, 23(6):578–81. (Cited on page 23.)

Horsch, C., Spruit, S., Lancee, J., van Eijk, R., Beun, R. J., Neerincx, M., and Brinkman, W.-P. (2016). Reminders make people adhere better to a self-help sleep intervention. *Health and Technology*, pages 1–16. (Cited on page 63.)

Hoyt, A. L., Rhodes, R. E., Hausenblas, H. A., and Giacobbi, P. R. (2009). Integrating five-factor model facet-level traits with the theory of planned behavior and exercise. *Psychology of Sport and Exercise*, 10(5):565–572. (Cited on pages 22 and 32.)

Hsueh, P.-Y., Melville, P., and Sindhwani, V. (2009). Data quality from crowdsourcing: A study of annotation selection criteria. In *Proceedings of the NAACL HLT 2009 Workshop on Active Learning for Natural Language Processing*, HLT '09, pages 27–35. (Cited on page 30.)

Ingledew, D. K. and Markland, D. (2008). The role of motives in exercise participation. *Psychology & Health*, 23(7):807–828. (Cited on page 23.)

John, O. P., Naumann, L. P., and Soto, C. J. (2008). Paradigm shift to the integrative big five trait taxonomy. *Handbook of Personality: Theory and Research*, 3:114–158. (Cited on page 62.)

Joseph, J., Breslin, C., and Skinner, H. (1999). *Critical perspectives on the transtheoretical model and stages of change*. Guilford Press. (Cited on page 111.)

Kaptein, M. C. (2015). Formalizing customization in persuasive technologies. In *Proceedings of the 10th International Conference on Persuasive Technology*, pages 27–38. Springer. (Cited on page 63.)

Kaptein, M. C., De Ruyter, B., Markopoulos, P., and Aarts, E. (2012). Adaptive persuasive systems: a study of tailored persuasive text messages to reduce snacking. *ACM Transactions on Interactive Intelligent Systems*, 2(2):10. (Cited on pages 23, 30, 50, 105, and 118.)

Kaptein, M. C. and Kruijswijk, J. (2016). Streamingbandit: Developing adaptive persuasive systems. *CoRR*, abs/1602.06700. (Cited on page 119.)

Kelly, M. P. (2016). Digital technologies and disease prevention. *American Journal of Preventive Medicine*, 51(5):861–863. (Cited on page 116.)

Kim, K., Bae, J., Nho, M.-W., and Lee, C. H. (2011). How do experts and novices differ? relation versus attribute and thinking versus feeling in language use. *Psychology of Aesthetics, Creativity, and the Arts*, 5(4):379. (Cited on page 77.)

Kimura, H., Ebisui, J., Funabashi, Y., Yoshii, A., and Nakajima, T. (2011). idetective: a persuasive application to motivate healthier behavior using smart phone. In *Proceedings of the 2011 ACM Symposium on Applied Computing*, pages 399–404. ACM. (Cited on page 12.)

King, A. C., Rejeski, W. J., and Buchner, D. M. (1998). A critical review and recommendations. 15(4). (Cited on page 87.)

Kittur, A., Chi, E. H., and Suh, B. (2008). Crowdsourcing user studies with mechanical turk. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 453–456. ACM. (Cited on page 30.)

Klasnja, P., Consolvo, S., McDonald, D. W., Landay, J. A., and Pratt, W. (2009). Using mobile & personal sensing technologies to support health behavior change in everyday life: lessons learned. In *AMIA Annual Symposium Proceedings*, volume 2009, page 338. American Medical Informatics Association. (Cited on pages 17 and 18.)

Klasnja, P., Consolvo, S., and Pratt, W. (2011). How to evaluate technologies for health behavior change in hci research. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3063–3072. ACM. (Cited on pages 17, 19, 87, and 118.)

Klasnja, P. and Pratt, W. (2012). Healthcare in the pocket: Mapping the space of mobile-phone health interventions. *Journal of Biomedical Informatics*, 45(1):184–198. (Cited on page 9.)

Kreuter, M., Farrell, D., Olevitch, L., and Brennan, L. (2000). Tailoring health messages: Customizing communication with computer technology. (Cited on page 5.)

Kristan, J. and Suffoletto, B. (2015). Using online crowdsourcing to understand young adult attitudes toward expert-authored messages aimed at reducing hazardous alcohol consumption and to collect peer-authored messages. *Translational Behavioral Medicine*, 5(1):45–52. (Cited on pages 23, 30, 62, 68, and 79.)

Kruijswijk, J., van Emden, R., Parvinen, P., and Kaptein, M. C. (2016). Streamingbandit; experimenting with bandit policies. (Cited on page 119.)

Kuznetsova, A., Bruun Brockhoff, P., and Haubo Bojesen Christensen, R. (2016). *lmerTest: Tests in Linear Mixed Effects Models*. R package version 2.0-32. (Cited on pages 52, 75, and 98.)

Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, pages 159–174. (Cited on pages 37 and 71.)

Latimer, A. E., Brawley, L. R., and Bassett, R. L. (2010). A systematic review of three approaches for constructing physical activity messages: What messages work and what improvements are needed? *The International Journal of Behavioral Nutrition and Physical Activity*, 7:36. (Cited on pages 5, 6, 23, 27, and 30.)

LeBlanc, A. G. and Chaput, J.-P. (2017). Pokémon go: A game changer for the physical inactivity crisis? *Preventive Medicine*, 101:235–237. (Cited on page 3.)

Lin, J. J., Mamykina, L., Lindtner, S., Delajoux, G., and Strub, H. B. (2006). Fish'n'steps: Encouraging physical activity with an interactive computer game. In *Proceedings of the 8th International Conference on Ubiquitous Computing*, pages 261–278. Springer. (Cited on page 18.)

Lister, C., West, J. H., Cannon, B., Sax, T., and Brodegard, D. (2014). Just a fad? gamification in health and fitness apps. *JMIR Serious Games*, 2(2). (Cited on page 2.)

Littell, J. H. and Girvin, H. (2002). Stages of change a critique. *Behavior Modification*, 26(2):223–273. (Cited on pages 104, 109, and 111.)

Locke, E. A. and Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, 57(9):705. (Cited on pages 4 and 13.)

MacQueen, K. M., McLellan, E., Kay, K., and Milstein, B. (1998). Codebook development for team-based qualitative analysis. *Cultural Anthropology Methods*, 10(2):31–36. (Cited on pages 36 and 62.)

Maitland, J., Sherwood, S., Barkhuus, L., Anderson, I., Hall, M., Brown, B., Chalmers, M., and Muller, H. (2006). Increasing the awareness of daily activity levels with pervasive computing. In *Pervasive Health Conference and Workshops*, pages 1–9. IEEE. (Cited on page 18.)

Marcus, B. H., Bock, B. C., Pinto, B. M., Forsyth, L. A. H., Roberts, M. B., and Traficante, R. M. (1998a). Efficacy of an individualized, motivationally-tailored physical activity intervention. *Annals of Behavioral Medicine*, 20(3):174–180. (Cited on pages 26 and 87.)

Marcus, B. H., Emmons, K. M., Simkin-Silverman, L. R., Linnan, L. A., Taylor, E. R., Bock, B. C., Roberts, M. B., Rossi, J. S., and Abrams, D. B. (1998b). Evaluation of motivationally tailored vs. standard self-help physical activity interventions at the workplace. *American Journal of Health Promotion*, 12(4):246–253. (Cited on page 87.)

Marcus, B. H., Rossi, J. S., Selby, V. C., Niaura, R. S., and Abrams, D. B. (1992). The stages and processes of exercise adoption and maintenance in a worksite sample. *Health Psychology*, 11(6):386. (Cited on pages 49, 51, and 53.)

Marge, M., Banerjee, S., and Rudnicky, A. I. (2010). Using the amazon mechanical turk for transcription of spoken language. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, pages 5270–5273. (Cited on page 30.)

Mason, W. and Suri, S. (2012). Conducting behavioral research on amazon's mechanical turk. *Behavior Research Methods*, 44(1):1–23. (Cited on pages 33, 64, and 112.)

McNamee, P., Murray, E., Kelly, M. P., Bojke, L., Chilcott, J., Fischer, A., West, R., and Yardley, L. (2016). Designing and undertaking a health economics study of digital health interventions. *American Journal of Preventive Medicine*, 51(5):852–860. (Cited on page 116.)

Michie, S., Carey, R. N., Johnston, M., Rothman, A. J., De Bruin, M., Kelly, M. P., and Connell, L. E. (2017). From theory-inspired to theory-based interventions: A protocol for developing and testing a methodology for linking behaviour change techniques to theoretical mechanisms of action. *Annals of Behavioral Medicine*, 52(6):501–512. (Cited on page 4.)

Michie, S., Johnston, M., Francis, J., Hardeman, W., and Eccles, M. (2008). From theory to intervention: Mapping theoretically derived behavioural determinants to behaviour change techniques. *Applied Psychology*, 57(4):660–680. (Cited on pages 4, 5, 24, and 27.)

Miller, W. R. and Rollnick, S. (2002). *Motivational interviewing: Preparing people for change*. Guilford press. (Cited on page 4.)

Mueller, F. F., Agamanolis, S., and Picard, R. (2003). Exertion interfaces: sports over a distance for social bonding and fun. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 561–568. ACM. (Cited on page 10.)

Mueller, F. F., Gibbs, M. R., and Vetere, F. (2008). Taxonomy of exertion games. In *Proceedings of the 20th Australasian Conference on Computer-Human Interaction: Designing for Habitus and Habitat*, pages 263–266. ACM. (Cited on page 10.)

Mueller, F. F., Gibbs, M. R., and Vetere, F. (2009). Design influence on social play in distributed exertion games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1539–1548. ACM. (Cited on page 10.)

Mueller, F. F., O'Brien, S., and Thorogood, A. (2007). Jogging over a distance: supporting a jogging together experience although being apart. In *CHI'07 Extended Abstracts on Human Factors in Computing Systems*, pages 1989–1994. ACM. (Cited on page 10.)

Mueller, F. F., Vetere, F., Gibbs, M. R., Agamanolis, S., and Sheridan, J. (2010). Jogging over a distance: the influence of design in parallel exertion games. In *Proceedings of the 5th ACM SIGGRAPH Symposium on Video Games*, pages 63–68. ACM. (Cited on pages 10 and 11.)

Munson, S. A. and Consolvo, S. (2012). Exploring goal-setting, rewards, self-monitoring, and sharing to motivate physical activity. In *6th International Conference on Pervasive Computing*

Technologies for Healthcare, pages 25–32. IEEE. (Cited on page 17.)

Murray, E., Hekler, E. B., Andersson, G., Collins, L. M., Doherty, A., Hollis, C., Rivera, D. E., West, R., and Wyatt, J. C. (2016). Evaluating digital health interventions: key questions and approaches. (Cited on pages 116 and 117.)

Mutsuddi, A. U. and Connelly, K. (2012). Text messages for encouraging physical activity are they effective after the novelty effect wears off? In *6th International Conference on Pervasive Computing Technologies for Healthcare*, pages 33–40. IEEE. (Cited on pages 5 and 29.)

Nigg, C. R., Geller, K. S., Motl, R. W., Horwath, C. C., Wertin, K. K., and Dishman, R. K. (2011). A research agenda to examine the efficacy and relevance of the transtheoretical model for physical activity behavior. *Psychology of Sport and Exercise*, 12(1):7–12. (Cited on pages 24, 25, and 31.)

Nigg, C. R., Norman, G. J., Rossi, J. S., and Benisovich, S. V. (1999). Processes of exercise behavior change: Redeveloping the scale. *Annals of Behavioral Medicine*, 21:S79. (Cited on pages 35, 50, and 74.)

Nigg, C. R., Rossi, J. S., Norman, G. J., and Benisovich, S. V. (1998). Structure of decisional balance for exercise adoption. *Annals of Behavioral Medicine*, 20:S211. (Cited on pages 35, 51, and 90.)

Noar, S. M., Benac, C. N., and Harris, M. S. (2007). Does tailoring matter? meta-analytic review of tailored print health behavior change interventions. *Psychological Bulletin*, 133(4):673–693. (Cited on pages 4, 5, and 6.)

Norman, G. J., Benisovich, S. V., Nigg, C. R., and Rossi, J. S. (1998). Examining three exercise staging algorithms in two samples. In *19th Annual Meeting of the Society of Behavioral Medicine*. (Cited on pages 35, 50, 74, and 90.)

O'Brien, S. and Mueller, F. F. (2007). Jogging the distance. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 523–526. ACM. (Cited on page 10.)

Oinas-Kukkonen, H. (2010). Behavior change support systems: A research model and agenda. In *Proceedings of the 5th International Conference on Persuasive Technology*, pages 4–14. Springer. (Cited on page 16.)

Oinas-Kukkonen, H. and Harjumaa, M. (2008). A systematic framework for designing and evaluating persuasive systems. In *Proceedings of the 3th International Conference on Persuasive Technology*, pages 164–176. Springer. (Cited on page 16.)

Patrick, K., Hekler, E. B., Estrin, D., Mohr, D. C., Riper, H., Crane, D., Godino, J., and Riley, W. T. (2016). The pace of technologic change: implications for digital health behavior intervention research. (Cited on page 116.)

Patrick, K., Raab, F., Adams, M. A., Dillon, L., Zabinski, M., Rock, C. L., Griswold, W. G., and Norman, G. J. (2009). A text message–based intervention for weight loss: randomized controlled trial. *Journal of Medical Internet Research*, 11(1). (Cited on page 30.)

Pennebaker, J. W., Booth, R. J., Boyd, R. L., and Francis, M. (2015a). Linguistic inquiry and word count: Liwc2015. (Cited on page 77.)

Pennebaker, J. W., Boyd, R. L., Jordan, K., and Blackburn, K. (2015b). The development and psychometric properties of liwc2015. Technical report. (Cited on page 77.)

Pescatello, L. S. and American College of Sports Medicine (2014). *ACSM's guidelines for exercise testing and prescription*. Lippincott Williams & Wilkins. (Cited on page 2.)

Prochaska, J. O. and DiClemente, C. C. (1983). Stages and processes of self-change of smoking: toward an integrative model of change. *Journal of Consulting and Clinical Psychology*, 51(3):390. (Cited on pages 13 and 25.)

Prochaska, J. O., DiClemente, C. C., and Norcross, J. C. (1993). In search of how people change: Applications to addictive behaviors. *Journal of Addictions Nursing*, 5(1):2–16. (Cited on page 25.)

Prochaska, J. O. and Velicer, W. F. (1997). The transtheoretical model of health behavior change. *American Journal of Health Promotion*, 12(1):38–48. (Cited on pages 1, 2, 3, 25, and 87.)

Redfern, J., Thiagalingam, A., Jan, S., Whittaker, R., Hackett, M., Mooney, J., De Keizer, L., Hillis, G. S., and Chow, C. K. (2014). Development of a set of mobile phone text messages designed for prevention of recurrent cardiovascular events. *European Journal of Preventive Cardiology*, 21(4):492–499. (Cited on page 30.)

Renfree, I., Harrison, D., Marshall, P., Stawarz, K., and Cox, A. (2016). Don't kick the habit: The role of dependency in habit formation apps. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 2932–2939. ACM. (Cited on page 1.)

Rhoads, K. (2007). How many influence, persuasion, compliance tactics & strategies are there. *http://www.workingpsychology.com/numbertactics.html*. (Cited on page 15.)

Rhodes, R. E., Courneya, K. S., and Jones, L. W. (2004). Personality and social cognitive influences on exercise behavior: adding the activity trait to the theory of planned behavior. *Psychology of Sport and Exercise*, 5(3):243–254. (Cited on pages 22, 23, and 44.)

Rhodes, R. E. and Smith, N. E. I. (2006). Personality correlates of physical activity: a review and meta-analysis. *British Journal of Sports Medicine*, 40(12):958–65. (Cited on pages 22, 23, 32, and 44.)

Rimer, B. K. and Kreuter, M. W. (2006). Advancing tailored health communication: A persuasion and message effects perspective. *Journal of Communication*, 56(s1):S184–S201. (Cited on pages 5 and 29.)

Sherwood, N. E. and Jeffery, R. W. (2000). The behavioral determinants of exercise: implications for physical activity interventions. *Annual Review of Nutrition*, 20(1):21–44. (Cited on page 21.)

Sohn, M. and Lee, J. (2007). Up health: ubiquitously persuasive health promotion with an instant messaging system. In *CHI'07 Extended Abstracts on Human Factors in Computing Systems*, pages 2663–2668. ACM. (Cited on page 16.)

Spencer, L., Adams, T. B., Malone, S., Roy, L., and Yost, E. (2006). Applying the transtheoretical model to exercise: a systematic and comprehensive review of the literature. *Health Promotion Practice*, 7(4):428–43. (Cited on page 26.)

Stanley, K. G., Pinelle, D., Bandurka, A., McDine, D., and Mandryk, R. L. (2008). Integrating cumulative context into computer games. In *Proceedings of the 2008 Conference on Future Play: Research, Play, Share*, pages 248–251. ACM. (Cited on page 14.)

Stokols, D. (1996). Translating social ecological theory into guidelines for community health promotion. *American Journal of Health Promotion*, 10(4):282–298. (Cited on pages 4 and 24.)

Tausczik, Y. R. and Pennebaker, J. W. (2010). The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1):24–54. (Cited on page 77.)

Team, R. C. (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. (Cited on pages 52, 75, 91, and 98.)

Toomim, M., Kriplean, T., Pörtner, C., and Landay, J. (2011). Utility of human-computer interactions: Toward a science of preference measurement. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2275–2284. ACM. (Cited on page 30.)

Toscos, T., Faber, A., An, S., and Gandhi, M. P. (2006). Chick clique: persuasive technology to motivate teenage girls to exercise. In *CHI'06 Extended Abstracts on Human Factors in Computing Systems*, pages 1873–1878. ACM. (Cited on page 16.)

Warburton, D. E., Nicol, C. W., and Bredin, S. S. (2006). Health benefits of physical activity: the evidence. *Canadian Medical Association Journal*, 174(6):801–809. (Cited on page 1.)

WHO (2010). *Global recommendations on physical activity for health*. World Health Organization. (Cited on pages 1 and 2.)

Yan, A. F., Stevens, P., Wang, Y., Weinhardt, L., Holt, C. L., O'Connor, C., Feller, T., Xie, H., and Luelloff, S. (2015). mhealth text messaging for physical activity promotion in college students: A formative participatory approach. *American Journal of Health Behavior*, 39(3):395–408. (Cited on page 23.)

Yardley, L., Choudhury, T., Patrick, K., and Michie, S. (2016a). Current issues and future directions for research into digital behavior change interventions. *American Journal of Preventive Medicine*, 51(5):814–815. (Cited on page 116.)

Yardley, L., Spring, B. J., Riper, H., Morrison, L. G., Crane, D. H., Curtis, K., Merchant, G. C., Naughton, F., and Blandford, A. (2016b). Understanding and promoting effective engagement with digital behavior change interventions. *American Journal of Preventive Medicine*, 51(5):833–842. (Cited on pages 116 and 117.)

Zaidan, O. F. and Callison-Burch, C. (2011). Crowdsourcing translation: Professional quality from non-professionals. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1*, HLT '11, pages 1220–1229. (Cited on page 30.)

List of publications

This dissertation is mainly based on four full-paper publications:

de Vries, R. A. J., Truong, K. P., & Evers, V. Crowd-Designed Motivation: Combining Personality and the Transtheoretical Model. *In International Conference on Persuasive Technology*. (pp. 41-52). Springer International Publishing. [de Vries et al., 2016a]

de Vries, R. A. J., Truong, K. P., Kwint, S., Drossaert, C. H. C., & Evers, V. Crowd-Designed Motivation: Motivational Messages for Exercise Adherence Based on Behavior Change Theory. *In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. (pp. 297-308). ACM. [de Vries et al., 2016b]

de Vries, R. A. J., Truong, K. P., Zaga, C., Li, J., & Evers, V. A Word of Advice: How to Tailor Motivational Text Messages Based on Behavior Change Theory to Personality and Gender. *Theme Issue: Supporting a Healthier Lifestyle with e-Coaching systems of the journal Personal and Ubiquitous Computing*. (pp. 675-687). Springer International Publishing. [de Vries et al., 2017a]

de Vries, R. A. J., Zaga, C., Bayer, F., Drossaert, C. H. C., Truong, K. P., & Evers, V. Experts Get Me Started, Peers Keep Me Going: Comparing Crowd- versus Expert-Designed Motivational Text Messages for Exercise Behavior Change. *In Proceedings of the EAI International Conference on Pervasive Computing Technologies for Healthcare.* (pp. 155-162). ACM. [de Vries et al., 2017b]

SIKS Dissertation Series

2011

01 Botond Cseke (RUN), Variational Algorithms for Bayesian Inference in Latent Gaussian Models

02 Nick Tinnemeier (UU), Organizing Agent Organizations. Syntax and Operational Semantics of an Organization-Oriented Programming Language

03 Jan Martijn van der Werf (TUE), Compositional Design and Verification of Component-Based Information Systems

04 Hado van Hasselt (UU), Insights in Reinforcement Learning; Formal analysis and empirical evaluation of temporal-difference

05 Bas van der Raadt (VU), Enterprise Architecture Coming of Age - Increasing the Performance of an Emerging Discipline.

06 Yiwen Wang (TUE), Semantically-Enhanced Recommendations in Cultural Heritage

07 Yujia Cao (UT), Multimodal Information Presentation for High Load Human Computer Interaction

08 Nieske Vergunst (UU), BDI-based Generation of Robust Task-Oriented Dialogues

09 Tim de Jong (OU), Contextualised Mobile Media for Learning

10 Bart Bogaert (UvT), Cloud Content Contention

11 Dhaval Vyas (UT), Designing for Awareness: An Experience-focused HCI Perspective

12 Carmen Bratosin (TUE), Grid Architecture for Distributed Process Mining

13 Xiaoyu Mao (UvT), Airport under Control. Multiagent Scheduling for Airport Ground Handling

14 Milan Lovric (EUR), Behavioral Finance and Agent-Based Artificial Markets

15 Marijn Koolen (UvA), The Meaning of Structure: the Value of Link Evidence for Information Retrieval

16 Maarten Schadd (UM), Selective Search in Games of Different Complexity

17 Jiyin He (UVA), Exploring Topic Structure: Coherence, Diversity and Relatedness

18 Mark Ponsen (UM), Strategic Decision-Making in complex games

19 Ellen Rusman (OU), The Mind's Eye on Personal Profiles

20 Qing Gu (VU), Guiding service-oriented software engineering - A view-based approach

21 Linda Terlouw (TUD), Modularization and Specification of Service-Oriented Systems

22 Junte Zhang (UVA), System Evaluation of Archival Description and Access

23 Wouter Weerkamp (UVA), Finding People and their Utterances in Social Media

24 Herwin van Welbergen (UT), Behavior Generation for Interpersonal Coordination with Virtual Humans On Specifying, Scheduling and Realizing Multimodal Virtual Human Behavior

25 Syed Waqar ul Qounain Jaffry (VU), Analysis and Validation of Models for Trust Dynamics

26 Matthijs Aart Pontier (VU), Virtual Agents for Human Communication - Emotion Regulation and Involvement-Distance Trade-Offs in Embodied Conversational Agents and Robots

27 Aniel Bhulai (VU), Dynamic website optimization through autonomous management of design patterns

28 Rianne Kaptein (UVA), Effective Focused Retrieval by Exploiting Query Context and Document Structure

29 Faisal Kamiran (TUE), Discrimination-aware Classification

30 Egon van den Broek (UT), Affective Signal Processing (ASP): Unraveling the mystery of emotions

31 Ludo Waltman (EUR), Computational and Game-Theoretic Approaches for Modeling Bounded Rationality

32 Nees-Jan van Eck (EUR), Methodological Advances in Bibliometric Mapping of Science

33 Tom van der Weide (UU), Arguing to Motivate Decisions

34 Paolo Turrini (UU), Strategic Reasoning in Interdependence: Logical and Game-theoretical Investigations

35 Maaike Harbers (UU), Explaining Agent Behavior in Virtual Training

36 Erik van der Spek (UU), Experiments in serious game design: a cognitive approach

37 Adriana Burlutiu (RUN), Machine Learning for Pairwise Data, Applications for Preference Learning and Supervised Network Inference

38 Nyree Lemmens (UM), Bee-inspired Distributed Optimization

39 Joost Westra (UU), Organizing Adaptation using Agents in Serious Games

40 Viktor Clerc (VU), Architectural Knowledge Management in Global Software Development

41 Luan Ibraimi (UT), Cryptographically Enforced Distributed Data Access Control

42 Michal Sindlar (UU), Explaining Behavior through Mental State Attribution

43 Henk van der Schuur (UU), Process Improvement through Software Operation Knowledge

44 Boris Reuderink (UT), Robust Brain-Computer Interfaces

45 Herman Stehouwer (UvT), Statistical Language Models for Alternative Sequence Selection

46 Beibei Hu (TUD), Towards Contextualized Information Delivery: A Rule-based Architecture for the Domain of Mobile Police Work

47 Azizi Bin Ab Aziz (VU), Exploring Computational Models for Intelligent Support of Persons with Depression

48 Mark Ter Maat (UT), Response Selection and Turn-taking for a Sensitive Artificial Listening Agent

49 Andreea 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

02 Muhammad Umair (VU), Adaptivity, emotion, and Rationality in Human and Ambient Agent Models

03 Adam Vanya (VU), Supporting Architecture Evolution by Mining Software Repositories

04 Jurriaan Souer (UU), Development of Content Management System-based Web Applications

05 Marijn Plomp (UU), Maturing Interorganisational Information Systems

06 Wolfgang Reinhardt (OU), Awareness Support for Knowledge Workers in Research Networks

07 Rianne van Lambalgen (VU), When the Going Gets Tough: Exploring Agent-based Models of Human Performance under Demanding Conditions

08 Gerben de Vries ◡̈ (UVA), Kernel Methods for Vessel Trajectories

09 Ricardo Neisse (UT), Trust and Privacy Management Support for Context-Aware Service Platforms

10 David Smits (TUE), Towards a Generic Distributed Adaptive Hypermedia Environment

11 J.C.B. Rantham Prabhakara (TUE), Process Mining in the Large: Preprocessing, Discovery, and Diagnostics

12 Kees van der Sluijs (TUE), Model Driven Design and Data Integration in Semantic Web Information Systems

13 Suleman Shahid (UvT), Fun and Face: Exploring non-verbal expressions of emotion during playful interactions

14 Evgeny Knutov (TUE), Generic Adaptation Framework for Unifying Adaptive Web-based Systems

15 Natalie van der Wal (VU), Social Agents. Agent-Based Modelling of Integrated Internal and Social Dynamics of Cognitive and Affective Processes.

16 Fiemke Both (VU), Helping people by understanding them - Ambient Agents supporting task execution and depression treatment

17 Amal Elgammal (UvT), Towards a Comprehensive Framework for Business Process Compliance

18 Eltjo Poort (VU), Improving Solution Architecting Practices

19 Helen Schonenberg (TUE), What's Next? Operational Support for Business Process Execution

20 Ali Bahramisharif (RUN), Covert Visual Spatial Attention, a Robust Paradigm for Brain-Computer Interfacing

21 Roberto Cornacchia (TUD), Querying Sparse Matrices for Information Retrieval

22 Thijs Vis (UvT), Intelligence, politie en veiligheidsdienst: verenigbare grootheden?

23 Christian Muehl (UT), Toward Affective Brain-Computer Interfaces: Exploring the Neurophysiology of Affect during Human Media Interaction

24 Laurens van der Werff (UT), Evaluation of Noisy Transcripts for Spoken Document Retrieval

25 Silja Eckartz (UT), Managing the Business Case Development in Inter-Organizational IT Projects: A Methodology and its Application

26 Emile de Maat (UVA), Making Sense of Legal Text

27 Hayrettin Gurkok (UT), Mind the Sheep! User Experience Evaluation & Brain-Computer Interface Games

28 Nancy Pascall (UvT), Engendering Technology Empowering Women

29 Almer Tigelaar (UT), Peer-to-Peer Information Retrieval

30 Alina Pommeranz (TUD), Designing Human-Centered Systems for Reflective Decision Making

31 Emily Bagarukayo (RUN), A Learning by Construction Approach for Higher Order Cognitive Skills Improvement, Building Capacity and Infrastructure

32 Wietske Visser (TUD), Qualitative multi-criteria preference representation and reasoning

33 Rory Sie (OUN), Coalitions in Cooperation Networks (COCOON)

34 Pavol Jancura (RUN), Evolutionary analysis in PPI networks and applications

35 Evert Haasdijk (VU), Never Too Old To Learn – On-line Evolution of Controllers in Swarm- and Modular Robotics

36 Denis Ssebugwawo (RUN), Analysis and Evaluation of Collaborative Modeling Processes

37 Agnes Nakakawa (RUN), A Collaboration Process for Enterprise Architecture Creation

38 Selmar Smit (VU), Parameter Tuning and Scientific Testing in Evolutionary Algorithms

39 Hassan Fatemi (UT), Risk-aware design of value and coordination networks

40 Agus Gunawan (UvT), Information Access for SMEs in Indonesia

41 Sebastian Kelle (OU), Game Design Patterns for Learning

42 Dominique Verpoorten (OU), Reflection Amplifiers in self-regulated Learning

43 Withdrawn

44 Anna Tordai (VU), On Combining Alignment Techniques

45 Benedikt Kratz (UvT), A Model and Language for Business-aware Transactions

46 Simon Carter (UVA), Exploration and Exploitation of Multilingual Data for Statistical Machine Translation

47 Manos Tsagkias (UVA), Mining Social Media: Tracking Content and Predicting Behavior

48 Jorn Bakker (TUE), Handling Abrupt Changes in Evolving Time-series Data

49 Michael Kaisers (UM), Learning against Learning - Evolutionary dynamics of reinforcement learning algorithms in strategic interactions

50 Steven van Kervel (TUD), Ontologogy driven Enterprise Information Systems Engineering

51 Jeroen de Jong (TUD), Heuristics in Dynamic Sceduling; a practical framework with a case study in elevator dispatching

2013

01 Viorel Milea (EUR), News Analytics for Financial Decision Support

02 Erietta Liarou (CWI), MonetDB/DataCell: Leveraging the Column-store Database Technology for Efficient and Scalable Stream Processing

03 Szymon Klarman (VU), Reasoning with Contexts in Description Logics

04 Chetan Yadati (TUD), Coordinating autonomous planning and scheduling

05 Dulce Pumareja (UT), Groupware Requirements Evolutions Patterns

06 Romulo Goncalves (CWI), The Data Cyclotron: Juggling Data and Queries for a Data Warehouse Audience

07	Giel van Lankveld (UvT), Quantifying Individual Player Differences
08	Robbert-Jan Merk (VU), Making enemies: cognitive modeling for opponent agents in fighter pilot simulators
09	Fabio Gori (RUN), Metagenomic Data Analysis: Computational Methods and Applications
10	Jeewanie Jayasinghe Arachchige (UvT), A Unified Modeling Framework for Service Design.
11	Evangelos Pournaras (TUD), Multi-level Reconfigurable Self-organization in Overlay Services
12	Marian Razavian (VU), Knowledge-driven Migration to Services
13	Mohammad Safiri (UT), Service Tailoring: User-centric creation of integrated IT-based homecare services to support independent living of elderly
14	Jafar Tanha (UVA), Ensemble Approaches to Semi-Supervised Learning Learning
15	Daniel Hennes (UM), Multiagent Learning - Dynamic Games and Applications
16	Eric Kok (UU), Exploring the practical benefits of argumentation in multi-agent deliberation
17	Koen Kok (VU), The PowerMatcher: Smart Coordination for the Smart Electricity Grid
18	Jeroen Janssens (UvT), Outlier Selection and One-Class Classification
19	Renze Steenhuizen (TUD), Coordinated Multi-Agent Planning and Scheduling
20	Katja Hofmann (UvA), Fast and Reliable Online Learning to Rank for Information Retrieval
21	Sander Wubben (UvT), Text-to-text generation by monolingual machine translation
22	Tom Claassen (RUN), Causal Discovery and Logic
23	Patricio de Alencar Silva (UvT), Value Activity Monitoring
24	Haitham Bou Ammar (UM), Automated Transfer in Reinforcement Learning
25	Agnieszka Anna Latoszek-Berendsen (UM), Intention-based Decision Support. A new way of representing and implementing clinical guidelines in a Decision Support System
26	Alireza Zarghami (UT), Architectural Support for Dynamic Homecare Service Provisioning
27	Mohammad Huq (UT), Inference-based Framework Managing Data Provenance
28	Frans van der Sluis (UT), When Complexity becomes Interesting: An Inquiry into the Information eXperience
29	Iwan de Kok (UT), Listening Heads
30	Joyce Nakatumba (TUE), Resource-Aware Business Process Management: Analysis and Support
31	Dinh Khoa Nguyen (UvT), Blueprint Model and Language for Engineering Cloud Applications
32	Kamakshi Rajagopal (OUN), Networking For Learning; The role of Networking in a Lifelong Learner's Professional Development
33	Qi Gao (TUD), User Modeling and Personalization in the Microblogging Sphere
34	Kien Tjin-Kam-Jet (UT), Distributed Deep Web Search
35	Abdallah El Ali (UvA), Minimal Mobile Human Computer Interaction
36	Than Lam Hoang (TUe), Pattern Mining in Data Streams
37	Dirk Börner (OUN), Ambient Learning Displays
38	Eelco den Heijer (VU), Autonomous Evolutionary Art
39	Joop de Jong (TUD), A Method for Enterprise Ontology based Design of Enterprise Information Systems
40	Pim Nijssen (UM), Monte-Carlo Tree Search for Multi-Player Games
41	Jochem Liem (UVA), Supporting the Conceptual Modelling of Dynamic Systems: A Knowledge Engineering Perspective on Qualitative Reasoning
42	Léon Planken (TUD), Algorithms for Simple Temporal Reasoning
43	Marc Bron (UVA), Exploration and Contextualization through Interaction and Concepts
2014	
01	Nicola Barile (UU), Studies in Learning Monotone Models from Data
02	Fiona Tuliyano (RUN), Combining System Dynamics with a Domain Modeling Method

03 Sergio Raul Duarte Torres (UT), Information Retrieval for Children: Search Behavior and Solutions

04 Hanna Jochmann-Mannak (UT), Websites for children: search strategies and interface design - Three studies on children's search performance and evaluation

05 Jurriaan van Reijsen (UU), Knowledge Perspectives on Advancing Dynamic Capability

06 Damian Tamburri (VU), Supporting Networked Software Development

07 Arya Adriansyah (TUE), Aligning Observed and Modeled Behavior

08 Samur Araujo (TUD), Data Integration over Distributed and Heterogeneous Data Endpoints

09 Philip Jackson (UvT), Toward Human-Level Artificial Intelligence: Representation and Computation of Meaning in Natural Language

10 Ivan Salvador Razo Zapata (VU), Service Value Networks

11 Janneke van der Zwaan (TUD), An Empathic Virtual Buddy for Social Support

12 Willem van Willigen (VU), Look Ma, No Hands: Aspects of Autonomous Vehicle Control

13 Arlette van Wissen (VU), Agent-Based Support for Behavior Change: Models and Applications in Health and Safety Domains

14 Yangyang Shi (TUD), Language Models With Meta-information

15 Natalya Mogles (VU), Agent-Based Analysis and Support of Human Functioning in Complex Socio-Technical Systems: Applications in Safety and Healthcare

16 Krystyna Milian (VU), Supporting trial recruitment and design by automatically interpreting eligibility criteria

17 Kathrin Dentler (VU), Computing healthcare quality indicators automatically: Secondary Use of Patient Data and Semantic Interoperability

18 Mattijs Ghijsen (UVA), Methods and Models for the Design and Study of Dynamic Agent Organizations

19 Vinicius Ramos (TUE), Adaptive Hypermedia Courses: Qualitative and Quantitative Evaluation and Tool Support

20 Mena Habib (UT), Named Entity Extraction and Disambiguation for Informal Text: The Missing Link

21 Kassidy Clark (TUD), Negotiation and Monitoring in Open Environments

22 Marieke Peeters (UU), Personalized Educational Games - Developing agent-supported scenario-based training

23 Eleftherios Sidirourgos (UvA/CWI), Space Efficient Indexes for the Big Data Era

24 Davide Ceolin (VU), Trusting Semi-structured Web Data

25 Martijn Lappenschaar (RUN), New network models for the analysis of disease interaction

26 Tim Baarslag (TUD), What to Bid and When to Stop

27 Rui Jorge Almeida (EUR), Conditional Density Models Integrating Fuzzy and Probabilistic Representations of Uncertainty

28 Anna Chmielowiec (VU), Decentralized k-Clique Matching

29 Jaap Kabbedijk (UU), Variability in Multi-Tenant Enterprise Software

30 Peter de Cock (UvT), Anticipating Criminal Behaviour

31 Leo van Moergestel (UU), Agent Technology in Agile Multiparallel Manufacturing and Product Support

32 Naser Ayat (UvA), On Entity Resolution in Probabilistic Data

33 Tesfa Tegegne (RUN), Service Discovery in eHealth

34 Christina Manteli (VU), The Effect of Governance in Global Software Development: Analyzing Transactive Memory Systems.

35 Joost van Ooijen (UU), Cognitive Agents in Virtual Worlds: A Middleware Design Approach

36 Joos Buijs (TUE), Flexible Evolutionary Algorithms for Mining Structured Process Models

37 Maral Dadvar (UT), Experts and Machines United Against Cyberbullying

38 Danny Plass-Oude Bos (UT), Making brain-computer interfaces better: improving usability through post-processing.

39 Jasmina Maric (UvT), Web Communities, Immigration, and Social Capital

40 Walter Omona (RUN), A Framework for Knowledge Management Using ICT in Higher Education

41 Frederic Hogenboom (EUR), Automated Detection of Financial Events in News Text

42 Carsten Eijckhof (CWI/TUD), Contextual Multidimensional Relevance Models

43 Kevin Vlaanderen (UU), Supporting Process Improvement using Method Increments

44 Paulien Meesters (UvT), Intelligent Blauw. Met als ondertitel: Intelligence-gestuurde politiezorg in gebiedsgebonden eenheden.

45 Birgit Schmitz (OUN), Mobile Games for Learning: A Pattern-Based Approach

46 Ke Tao (TUD), Social Web Data Analytics: Relevance, Redundancy, Diversity

47 Shangsong Liang (UVA), Fusion and Diversification in Information Retrieval

2015

01 Niels Netten (UvA), Machine Learning for Relevance of Information in Crisis Response

02 Faiza Bukhsh (UvT), Smart auditing: Innovative Compliance Checking in Customs Controls

03 Twan van Laarhoven (RUN), Machine learning for network data

04 Howard Spoelstra (OUN), Collaborations in Open Learning Environments

05 Christoph Bösch (UT), Cryptographically Enforced Search Pattern Hiding

06 Farideh Heidari (TUD), Business Process Quality Computation - Computing Non-Functional Requirements to Improve Business Processes

07 Maria-Hendrike Peetz (UvA), Time-Aware Online Reputation Analysis

08 Jie Jiang (TUD), Organizational Compliance: An agent-based model for designing and evaluating organizational interactions

09 Randy Klaassen (UT), HCI Perspectives on Behavior Change Support Systems

10 Henry Hermans (OUN), OpenU: design of an integrated system to support lifelong learning

11 Yongming Luo (TUE), Designing algorithms for big graph datasets: A study of computing bisimulation and joins

12 Julie M. Birkholz (VU), Modi Operandi of Social Network Dynamics: The Effect of Context on Scientific Collaboration Networks

13 Giuseppe Procaccianti (VU), Energy-Efficient Software

14 Bart van Straalen (UT), A cognitive approach to modeling bad news conversations

15 Klaas Andries de Graaf (VU), Ontology-based Software Architecture Documentation

16 Changyun Wei (UT), Cognitive Coordination for Cooperative Multi-Robot Teamwork

17 André van Cleeff (UT), Physical and Digital Security Mechanisms: Properties, Combinations and Trade-offs

18 Holger Pirk (CWI), Waste Not, Want Not! - Managing Relational Data in Asymmetric Memories

19 Bernardo Tabuenca (OUN), Ubiquitous Technology for Lifelong Learners

20 Lois Vanhée (UU), Using Culture and Values to Support Flexible Coordination

21 Sibren Fetter (OUN), Using Peer-Support to Expand and Stabilize Online Learning

22 Zhemin Zhu (UT), Co-occurrence Rate Networks

23 Luit Gazendam (VU), Cataloguer Support in Cultural Heritage

24 Richard Berendsen (UVA), Finding People, Papers, and Posts: Vertical Search Algorithms and Evaluation

25 Steven Woudenberg (UU), Bayesian Tools for Early Disease Detection

26 Alexander Hogenboom (EUR), Sentiment Analysis of Text Guided by Semantics and Structure

27 Sándor Héman (CWI), Updating compressed colomn stores

28 Janet Bagorogoza (TiU), Knowledge Management and High Performance; The Uganda Financial Institutions Model for HPO

29 Hendrik Baier (UM), Monte-Carlo Tree Search Enhancements for One-Player and Two-Player Domains

30 Kiavash Bahreini (OU), Real-time Multimodal Emotion Recognition in E-Learning

31 Yakup Koç (TUD), On the robustness of Power Grids

32 Jerome Gard (UL), Corporate Venture Management in SMEs

33 Frederik Schadd (TUD), Ontology Mapping with Auxiliary Resources

34 Victor de Graaf (UT), Gesocial Recommender Systems

35 Jungxao Xu (TUD), Affective Body Language of Humanoid Robots: Perception and Effects in Human Robot Interaction

2016

01 Syed Saiden Abbas (RUN), Recognition of Shapes by Humans and Machines

02 Michiel Christiaan Meulendijk (UU), Optimizing medication reviews through decision support: prescribing a better pill to swallow

03 Maya Sappelli (RUN), Knowledge Work in Context: User Centered Knowledge Worker Support

04 Laurens Rietveld (VU), Publishing and Consuming Linked Data

05 Evgeny Sherkhonov (UVA), Expanded Acyclic Queries: Containment and an Application in Explaining Missing Answers

06 Michel Wilson (TUD), Robust scheduling in an uncertain environment

07 Jeroen de Man (VU), Measuring and modeling negative emotions for virtual training

08 Matje van de Camp (TiU), A Link to the Past: Constructing Historical Social Networks from Unstructured Data

09 Archana Nottamkandath (VU), Trusting Crowdsourced Information on Cultural Artefacts

10 George Karafotias (VUA), Parameter Control for Evolutionary Algorithms

11 Anne Schuth (UVA), Search Engines that Learn from Their Users

12 Max Knobbout (UU), Logics for Modelling and Verifying Normative Multi-Agent Systems

13 Nana Baah Gyan (VU), The Web, Speech Technologies and Rural Development in West Africa - An ICT4D Approach

14 Ravi Khadka (UU), Revisiting Legacy Software System Modernization

15 Steffen Michels (RUN), Hybrid Probabilistic Logics - Theoretical Aspects, Algorithms and Experiments

16 Guangliang Li (UVA), Socially Intelligent Autonomous Agents that Learn from Human Reward

17 Berend Weel (VU), Towards Embodied Evolution of Robot Organisms

18 Albert Meroño Peñuela (VU), Refining Statistical Data on the Web

19 Julia Efremova (Tu/e), Mining Social Structures from Genealogical Data

20 Daan Odijk (UVA), Context & Semantics in News & Web Search

21 Alejandro Moreno Célleri (UT), From Traditional to Interactive Playspaces: Automatic Analysis of Player Behavior in the Interactive Tag Playground

22 Grace Lewis (VU), Software Architecture Strategies for Cyber-Foraging Systems

23 Fei Cai (UVA), Query Auto Completion in Information Retrieval

24 Brend Wanders (UT), Repurposing and Probabilistic Integration of Data; An Iterative and data model independent approach

25 Julia Kiseleva (TU/e), Using Contextual Information to Understand Searching and Browsing Behavior

26 Dilhan Thilakarathne (VU), 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

27 Wen Li (TUD), Understanding Geo-spatial Information on Social Media

28 Mingxin Zhang (TUD), Large-scale Agent-based Social Simulation - A study on epidemic prediction and control

29 Nicolas Höning (TUD), Peak reduction in decentralised electricity systems - Markets and prices for flexible planning

30 Ruud Mattheij (UvT), The Eyes Have It

31 Mohammad Khelghati (UT), Deep web content monitoring

32 Eelco Vriezekolk (UT), Assessing Telecommunication Service Availability Risks for Crisis Organisations

33 Peter Bloem (UVA), Single Sample Statistics, exercises in learning from just one example

34 Dennis Schunselaar (TUE), Configurable Process Trees: Elicitation, Analysis, and Enactment

35 Zhaochun Ren (UVA), Monitoring Social Media: Summarization, Classification and Recommendation

36 Daphne Karreman (UT), Beyond R2D2: The design of nonverbal interaction behavior optimized for robot-specific morphologies

37 Giovanni Sileno (UvA), Aligning Law and Action - a conceptual and computational inquiry

38 Andrea Minuto (UT), Materials that Matter - Smart Materials meet Art & Interaction Design

39 Merijn Bruijnes (UT), Believable Suspect Agents; Response and Interpersonal Style Selection for an Artificial Suspect

40 Christian Detweiler (TUD), Accounting for Values in Design

41 Thomas King (TUD), Governing Governance: A Formal Framework for Analysing Institutional Design and Enactment Governance

42 Spyros Martzoukos (UVA), Combinatorial and Compositional Aspects of Bilingual Aligned Corpora

43 Saskia Koldijk (RUN), Context-Aware Support for Stress Self-Management: From Theory to Practice

44 Thibault Sellam (UVA), Automatic Assistants for Database Exploration

45 Bram van de Laar (UT), Experiencing Brain-Computer Interface Control

46 Jorge Gallego Perez (UT), Robots to Make you Happy

47 Christina Weber (UL), Real-time foresight - Preparedness for dynamic innovation networks

48 Tanja Buttler (TUD), Collecting Lessons Learned

49 Gleb Polevoy (TUD), Participation and Interaction in Projects. A Game-Theoretic Analysis

50 Yan Wang (UVT), The Bridge of Dreams: Towards a Method for Operational Performance Alignment in IT-enabled Service Supply Chains

2017

01 Jan-Jaap Oerlemans (UL), Investigating Cybercrime

02 Sjoerd Timmer (UU), Designing and Understanding Forensic Bayesian Networks using Argumentation

03 Daniël Harold Telgen (UU), Grid Manufacturing; A Cyber-Physical Approach with Autonomous Products and Reconfigurable Manufacturing Machines

04 Mrunal Gawade (CWI), Multi-core Parallelism in a Column-store

05 Mahdieh Shadi (UVA), Collaboration Behavior

06 Damir Vandic (EUR), Intelligent Information Systems for Web Product Search

07 Roel Bertens (UU), Insight in Information: from Abstract to Anomaly

08 Rob Konijn (VU) , Detecting Interesting Differences:Data Mining in Health Insurance Data using Outlier Detection and Subgroup Discovery

09 Dong Nguyen (UT), Text as Social and Cultural Data: A Computational Perspective on Variation in Text

10 Robby van Delden (UT), (Steering) Interactive Play Behavior

11 Florian Kunneman (RUN), Modelling patterns of time and emotion in Twitter #anticipointment

12 Sander Leemans (TUE), Robust Process Mining with Guarantees

13 Gijs Huisman (UT), Social Touch Technology - Extending the reach of social touch through haptic technology

14 Shoshannah Tekofsky (UvT), You Are Who You Play You Are: Modelling Player Traits from Video Game Behavior

15 Peter Berck (RUN), Memory-Based Text Correction

16 Aleksandr Chuklin (UVA), Understanding and Modeling Users of Modern Search Engines

17 Daniel Dimov (UL), Crowdsourced Online Dispute Resolution

18 Ridho Reinanda (UVA), Entity Associations for Search

19 Jeroen Vuurens (UT), Proximity of Terms, Texts and Semantic Vectors in Information Retrieval

20 Mohammadbashir Sedighi (TUD), Fostering Engagement in Knowledge Sharing: The Role of Perceived Benefits, Costs and Visibility

21 Jeroen Linssen (UT), Meta Matters in Interactive Storytelling and Serious Gaming (A Play on Worlds)

22 Sara Magliacane (VU), Logics for causal inference under uncertainty

23 David Graus (UVA), Entities of Interest — Discovery in Digital Traces

24 Chang Wang (TUD), Use of Affordances for Efficient Robot Learning

25 Veruska Zamborlini (VU), Knowledge Representation for Clinical Guidelines, with applications to Multimorbidity Analysis and Literature Search

26 Merel Jung (UT), Socially intelligent robots that understand and respond to human touch

27 Michiel Joosse (UT), Investigating Positioning and Gaze Behaviors of Social Robots: People's Preferences, Perceptions and Behaviors

28 John Klein (VU), Architecture Practices for Complex Contexts

29 Adel Alhuraibi (UvT), From IT-BusinessStrategic Alignment to Performance: A Moderated Mediation Model of Social Innovation, and Enterprise Governance of IT"

30 Wilma Latuny (UvT), The Power of Facial Expressions

31 Ben Ruijl (UL), Advances in computational methods for QFT calculations

32 Thaer Samar (RUN), Access to and Retrievability of Content in Web Archives

33 Brigit van Loggem (OU), Towards a Design Rationale for Software Documentation: A Model of Computer-Mediated Activity

34 Maren Scheffel (OU), The Evaluation Framework for Learning Analytics

35 Martine de Vos (VU), Interpreting natural science spreadsheets

36 Yuanhao Guo (UL), Shape Analysis for Phenotype Characterisation from High-throughput Imaging

37 Alejandro Montes Garcia (TUE), WiBAF: A Within Browser Adaptation Framework that Enables Control over Privacy

38 Alex Kayal (TUD), Normative Social Applications

39 Sara Ahmadi (RUN), Exploiting properties of the human auditory system and compressive sensing methods to increase noise robustness in ASR

40 Altaf Hussain Abro (VUA), Steer your Mind: Computational Exploration of Human Control in Relation to Emotions, Desires and Social Support For applications in human-aware support systems

41 Adnan Manzoor (VUA), Minding a Healthy Lifestyle: An Exploration of Mental Processes and a Smart Environment to Provide Support for a Healthy Lifestyle

42 Elena Sokolova (RUN), Causal discovery from mixed and missing data with applications on ADHD datasets

43 Maaike de Boer (RUN), Semantic Mapping in Video Retrieval

44 Garm Lucassen (UU), Understanding User Stories - Computational Linguistics in Agile Requirements Engineering

45 Bas Testerink (UU), Decentralized Runtime Norm Enforcement

46 Jan Schneider (OU), Sensor-based Learning Support

47 Jie Yang (TUD), Crowd Knowledge Creation Acceleration

48 Angel Suarez (OU), Collaborative inquiry-based learning

2018

01 Han van der Aa (VUA), Comparing and Aligning Process Representations

02 Felix Mannhardt (TUE), Multi-perspective Process Mining

03 Steven Bosems (UT), Causal Models For Well-Being: Knowledge Modeling, Model-Driven Development of Context-Aware Applications, and Behavior Prediction

04 Jordan Janeiro (TUD), Flexible Coordination Support for Diagnosis Teams in Data-Centric Engineering Tasks

05 Hugo Huurdeman (UVA), Supporting the Complex Dynamics of the Information Seeking Process

06 Dan Ionita (UT), Model-Driven Information Security Risk Assessment of Socio-Technical Systems

07 Jieting Luo (UU), A formal account of opportunism in multi-agent systems

08 Rick Smetsers (RUN), Advances in Model Learning for Software Systems

09 Xu Xie (TUD), Data Assimilation in Discrete Event Simulations

10 Julienka Mollee (VUA), Moving forward: supporting physical activity behavior change through intelligent technology

11	Mahdi Sargolzaei (UVA), Enabling Framework for Service-oriented Collaborative Networks
12	Xixi Lu (TUE), Using behavioral context in process mining
13	Seyed Amin Tabatabaei (VUA), Computing a Sustainable Future
14	Bart Joosten (UVT), Detecting Social Signals with Spatiotemporal Gabor Filters
15	Naser Davarzani (UM), Biomarker discovery in heart failure
16	Jaebok Kim (UT), Automatic recognition of engagement and emotion in a group of children
17	Jianpeng Zhang (TUE), On Graph Sample Clustering
18	Henriette Nakad (UL), De Notaris en Private Rechtspraak
19	Minh Duc Pham (VUA), Emergent relational schemas for RDF
20	Manxia Liu (RUN), Time and Bayesian Networks
21	Aad Slootmaker (OUN), EMERGO: a generic platform for authoring and playing scenario-based serious games
22	Eric Fernandes de Mello Araujo (VUA), Contagious: Modeling the Spread of Behaviours, Perceptions and Emotions in Social Networks
23	Kim Schouten (EUR), Semantics-driven Aspect-Based Sentiment Analysis
24	Jered Vroon (UT), Responsive Social Positioning Behaviour for Semi-Autonomous Telepresence Robots
25	Riste Gligorov (VUA), Serious Games in Audio-Visual Collections

Regular exercise is important for you to stay healthy. Regular exercise will help you to better handle all the stress factors in your life. After a longer time of regular exercising you will see how your body changes in a positive way. You will feel so much better with some regular exercise in your life. Regular exercise will make you less vulnerable for diseases. Think about the physical disadvantages that result from your inactivity. How would you feel in a year if you do not start to exercise now? Are you done doing nothing? Are you still hesitating? Pretty soon time will run against you! The longer you push exercise aside, the longer you are denying yourself a better quality of life. How about joining your friends that are already exercising? It will not only improve your own health, but also your social relationships. Aren't your friends already envying you because of your discipline? Regular exercise with friends does not only improve your health, but also deepens your social relationships. Work out with friends. You will not only stay healthy, but also improve your social life. Have you talked to a friend about this decision? Look around in your neighbourhood if you can find any facilities to do some sports. Keep an eye out for some nice areas in your district where you can go jogging. Keep an eye out for a gym or jogging route that is close by. Try out different sports, and then think about which one you liked the most. Is there any club nearby which offers this sport? Think of a sport that you personally enjoy the most - running, swimming, tennis...what do you like? Wouldn't it be worth it for you own well-being to start exercising? Imagine how much better you would feel with some regular exercise in your life! Think about all the positive consequences that will result from your exercise. Think of all the great results you could achieve within a year if you start now to implement fitness in your daily life. If you start working out today you will see results within two weeks! And wouldn't that be a great feeling? You can do it - start your journey into a more active and healthy life. Try to set daily goals that you can reach. A bit of exercising (e.g. going for a strong walk) is always better than none. Keep on exercising until it will become a steady aspect of your life! Start now! The beginning is always the most difficult part of the whole journey! It's time to finally change your thoughts into action! Start with some easy and slow exercises. I promise you that you will love working out! I promise it's going to pay off afterwards! I have trust in you - you can do it! I am sure that you can do it, but to get there you have to overcome yourself and start with the first steps. I am sure you would feel much better with a little exercise in your life. Think of something that could make exercising even more fun for you! Try it out the next time! Don't quit! Starting again from the beginning is more difficult than continuing. If you keep on going, those exercises will get much easier for you in the future. Make exercise your hobby! If you have fun while doing it, it will get easier to start. Don't forget to enjoy working out. Wow, you are doing very well! Hold on! Good work, you already managed the most important part. Now keep on going. You can be proud of yourself. It is impressive how good you are doing with incorporating regular exercise into your daily life. You have come this far, you should be proud of your achievements. Set fixed times in the week for physical exercise. Have you already worked out today? No? - Then what are you waiting for? Start to make a weekly agenda to set fixed times for your exercises. What are you doing right now? What about going for a run? Have you already planned your next exercise session? Planning helps to keep on track.