Year : 2022

In the Quest for the Motivation App: Designing Effective Behavior Change Apps Through the Lens of the Self-Determination Theory

Villalobos Zúñiga María Gabriela

UNIL | Université de Lausanne

FACULTÉ DES HAUTES ÉTUDES COMMERCIALES

DÉPARTEMENT DES SYSTÈMES D'INFORMATION

In the Quest for the Motivation App: Designing Effective Behavior Change Apps Through the Lens of the Self-Determination Theory

THÈSE DE DOCTORAT

présentée à la

Faculté des Hautes Études Commerciales
de l'Université de Lausanne

pour l'obtention du grade de
Docteure ès Sciences en systèmes d'information

par

María Gabriela VILLALOBOS ZÚÑIGA

Directeur de thèse
Prof. Mauro Cherubini

Jury

Prof. Rafael Lalive, président
Prof. Kévin Huguenin, expert interne
Prof. Rita Orji, experte externe
Prof. Max Birk, expert externe

LAUSANNE
2022

UNIL | Université de Lausanne

FACULTÉ DES HAUTES ÉTUDES COMMERCIALES

DÉPARTEMENT DES SYSTÈMES D'INFORMATION

In the Quest for the Motivation App: Designing Effective Behavior Change Apps Through the Lens of the Self-Determination Theory

THÈSE DE DOCTORAT

présentée à la

Faculté des Hautes Études Commerciales
de l'Université de Lausanne

pour l'obtention du grade de
Docteure ès Sciences en systèmes d'information

par

María Gabriela VILLALOBOS ZÚÑIGA

Directeur de thèse
Prof. Mauro Cherubini

Jury

Prof. Rafael Lalive, président
Prof. Kévin Huguenin, expert interne
Prof. Rita Orji, experte externe
Prof. Max Birk, expert externe

LAUSANNE
2022

UNIL | Université de Lausanne
HEC Lausanne
Le Décanat
Bâtiment Internef
CH-1015 Lausanne

IMPRIMATUR

Sans se prononcer sur les opinions de l'autrice, la Faculté des Hautes Etudes Commerciales de l'Université de Lausanne autorise l'impression de la thèse de Madame María Gabriela VILLALOBOS ZÚÑIGA, titulaire d’un bachelor in Ingeniería en Computación de l’Institut de Technologie du Costa Rica, d’un Master in Science (Technology) de l’Université de Aalto et d’un Master en Sciences, Technologies, Santé de l’Université de Paris-Sud XI, en vue de l'obtention du grade de docteure ès Sciences en systèmes d’information.

La thèse est intitulée :

IN THE QUEST FOR THE MOTIVATION APP: DESIGNING EFFECTIVE BEHAVIOR CHANGE APPS THROUGH THE LENS OF THE SELF-DETERMINATION THEORY

Lausanne, le 14 janvier 2022

La Doyenne

Marianne SCHMID MAST

Members of the thesis committee

Prof. Mauro Cherubini
University of Lausanne
Thesis supervisor

Prof. Kévin Huguenin
University of Lausanne
Internal member of the thesis committee

Prof. Rita Orji
Dalhousie University
External member of the thesis committee

Prof. Max Birk
Eindhoven University of Technology
External member of the thesis committee

University of Lausanne
Faculty of Business and Economics

PhD in Information Systems

I hereby certify that I have examined the doctoral thesis of

Maria Gabriela VILLALOBOS ZUNIGA

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: ______________________ Date: 7.1.2022

Prof. Mauro CHERUBINI
Thesis supervisor

University of Lausanne
Faculty of Business and Economics

PhD in Information Systems

I hereby certify that I have examined the doctoral thesis of

Maria Gabriela VILLALOBOS ZUNIGA

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: _______________________________ Date: 23/12/2021____________

Prof. Kévin HUGUENIN
Internal member of the doctoral committee

University of Lausanne
Faculty of Business and Economics

PhD in Information Systems

I hereby certify that I have examined the doctoral thesis of

Maria Gabriela VILLALOBOS ZUNIGA

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: Rita Orji Date: January 6, 2022

Prof. Rita ORJI
External member of the doctoral committee

University of Lausanne
Faculty of Business and Economics

PhD in Information Systems

I hereby certify that I have examined the doctoral thesis of

Maria Gabriela VILLALOBOS ZUNIGA

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: Date: 20/12/2021

Prof. Max BIRK
External member of the doctoral committee

In the Quest for the Motivation App: Designing Effective Behavior Change Apps Through the Lens of the Self-Determination Theory

Gabriela Villalobos-Zúñiga

A thesis submitted to the University of Lausanne
for the degree of Ph.D. in Information Systems
Lausanne - December, 2021

Abstract

Mobile apps are used to support behavior change goals (e.g., stopping a lousy habit, increasing the physical activity frequency, or learning a new skill). Because these apps are pervasive, they are great tools to reach and help people in their self-improvement path towards better habits and well-being. However, users of these apps fail to reach their objectives because they lack the motivation to attain their goals. App creators might help to support motivation by integrating human motivation theories into their designs. Still, not many apps are theory-grounded. To address this issue, we organized our research around a comprehensive human motivation theory: The Self-Determination Theory (SDT). We explored three research streams: (1) developing an artifact that maps market app features to support the SDT Basic Psychological Needs (BPNs); (2) creating an SDT inspired physical activity app that provided empirical evidence that its design supported the BPNs; (3) providing empirical evidence of an SDT inspired app design that contributes to increasing the physical activity and motivation of individuals. The results of our studies show that the SDT can inform the design of behavior change app features. Moreover, these SDT inspired features can be used to create a physical activity app that improves individuals' intrinsic motivation and physical activity level. This thesis results have practical implications for app designers, policymakers, and health practitioners whose interest lies in creating theory-informed and effective behavior change apps.

Abstract

Les applications mobiles sont utilisées pour les objectifs de changement de comportements (par exemple, arrêter une mauvaise habitude, augmenter la fréquence d'activités physiques ou apprendre une nouvelle compétence). Comme ces applications sont omniprésentes, elles constituent d'excellents outils pour atteindre et aider les gens dans leur chemin vers de meilleures habitudes et bien-être. Cependant, les utilisateurs de ces applications ne parviennent pas à atteindre leurs objectifs car ils manquent de motivation pour y parvenir. Les créateurs d'applications pourraient contribuer à soutenir leur motivation en intégrant les théories basées sur la motivation humaine dans leurs conceptions. Pourtant, peu d'applications sont fondées sur ces théories. Pour résoudre ce problème, nous avons organisé notre recherche autour d'une théorie globale de la motivation humaine : La théorie de l'autodétermination (SDT). Nous avons exploré trois axes de recherche : (1) le développement d'un artefact qui met en correspondance les caractéristiques des applications du marché avec les besoins psychologiques fondamentaux (BPN) de SDT ; (2) la création d'une application d'activité physique basée sur SDT qui fournit des preuves empiriques que sa conception soutient les BPN ; (3) donner des preuves empiriques d'une conception d'application inspirée de SDT qui contribue à augmenter l'activité physique et la motivation des individus. Les résultats de nos études montrent que SDT peut servir de base à la conception de fonctionnalités d'application pour les changements comportementaux. De plus, ces caractéristiques inspirées de SDT peuvent être utilisées pour créer une application d'activité physique qui améliore la motivation personnelle et le niveau d'activité physique des individus. Les résultats de cette thèse ont des implications pratiques pour les créateurs d'applications, les respon-

sables politiques et les médecins de la santé qui s'intéressent à la création d'applications efficaces et fondées sur cette théorie de la modification du comportement.

Published Works

At the time of submission, the chapters of this thesis work have previously appeared in peer-reviewed publications. The following are the full references for these publications.

1. M. Cherubini, G. Villalobos-Zuñiga, M.-O. Boldi, and R. Bonazzi. 2020. The Unexpected Downside of Paying or Sending Messages to People to Make Them Walk: Comparing Tangible Rewards and Motivational Messages to Improve Physical Activity. ACM Trans. Comput.-Hum. Interact (ToCHI). 27, 2, Article 8 (March 2020), 44 pages. DOI: 10.1145/3365665
2. G. Villalobos-Zuñiga, M. Cherubini. 2020. Apps That Motivate: a Taxonomy of App Features Based on Self-Determination Theory. International Journal of Human-Computer Studies (IJHCS). Volume 140, August 2020, 102449. DOI: 10.1016/j.ijhcs.2020.102449
3. G. Villalobos-Zúñig. Designing for Gender Role Differences Through the Lens of Self-Determination Theory. Proceedings of User Modeling, Adaptation and Personalization (UMAP) 2021. DOI: 10.1145/3450614.3464622
4. G. Villalobos-Zúñiga, I. Rodríguez, A. Fedosov, M. Cherubin. Informed Choices, Progress Monitoring and Comparison with Peers: Features to Support the Autonomy, Competence and Relatedness Needs, as suggested by the Self-Determination Theory. Proceedings of The ACM International Conference on Mobile Human-Computer Interaction (MobileHCI) 2021. DOI: 10.1145/3447526.3472039

Declaration

This thesis is the result of my own work, except where explicit reference is made to the work of others. Chapters 1, 2, and 4 corresponded to previous publications and were modified by request of the Thesis Jury.

Gabriela Villalobos-Zúñiga

Acknowledgements

Thank you | Merci | Gracias

Figure 1
http://www.gvillaloboz.com/phd-thesis

Contents

"Don't let the noise of others' opinions drown out your own inner voice. And most important, have the courage to follow your heart and intuition. They somehow already know what you truly want to become."

— Steve Jobs

Introduction

"Tous les jours, à tous points de vue, je vais de mieux en mieux."
— Émile Coué, 1857–1926

0.1 Context and Relevance

Progress towards *the unique self we want to become*, this is what the Swiss psychologist Carl Gustav Jung believed to be the natural state of human beings. He coined this reflection in the early 20th century when he explained that our natural tendency is to come to *self-realization*[1] through the process he titled *individuation* [Jung, 1928, p.184]:

> "Individuation means to become a single, discrete being, and, inasmuch as the concept individuality embraces that innermost, last, and incomparable uniqueness of our being, it also includes the idea of becoming one's own real self. Hence individuation could also be translated as 'coming to self-hood,' or 'self-realization."'

Jung was not the only psychologist who studied the area of *self-realization*, some other prominent psychologists have also referred to this sense of constantly pursuing what we *feel* we want to become. For instance, Abraham Maslow, who in 1943 examined the concept of *self-actualization*. He introduced this term as one of the basic needs of human beings in his famous work a "Theory of Motivation" [Maslow, 1943]. Specifically, Maslow defined a hierarchy of needs and situated at the base of the pyramid the psychological needs, followed by the safety needs, a level up he placed the need for love and belonging, followed by the need for esteem and on the top he introduced the need for *self-actualization*.

[1]The Merriam-Webster Diccionary defines self-realization as: fulfillment by oneself of the possibilities of one's character or personality.

More recently, in 1994, the Canadian psychologist, Albert Bandura described the term *self-efficacy* referring to people's belief about their capabilities to produce effects and how this belief contributes to support personal development [Bandura and Wessels, 1994].

In essence, all the above thoughts from these renowned psychologists converge in one aspect: we humans have a *natural need* to become better versions of ourselves and, in this way, attain *self-realization*.

Today, in the 21th century, these concepts of *self-realization*, *self-actualization*, *self-efficacy*, and constant improvement direct us to what modern psychologists call *well-being*: "The level to which one finds meaning in life and fulfills one's greatest potential." [Calvo and Peters, 2014, p.14].

Why is well-being important? Because having high well-being indicates that our lives are progressing favorably, and we feel satisfied and fulfilled. There are multiple ways to measure well-being, and keeping an eye on these metrics results useful not only for the general population but also for policymakers. These metrics serve to evaluate how effective their policies have been. Measuring well-being is also instrumental for multiple stakeholders involved in disease prevention and health promotion.

Who benefits from higher states of well-being? The overall population receives a direct benefit because higher levels of well-being are associated with decreased risk of disease, illness, injury, and increased longevity [Pressman and Cohen, 2005]; [Ostir et al., 2000]; [Ostir et al., 2001]; [Diener and Biswas-Diener, 2011]; [L. Fredrickson and Levenson, 1998]. In turn, all these benefits contribute to the governments and their health systems because their populations will be healthier and fall less sick. Furthermore, people with high levels of well-being are more productive at work and are more likely to contribute to their communities [Frey and Stutzer, 2010]; [Tov and Diener, 2009].

We might find ourselves asking **how to achieve higher states of well-being?** A promising approach is by focusing on physical activity improvement. A vast body of research supports the positive relationship between physical activity and well-being. Therefore, by defining and pursuing fitness goals, we can pave the way for promoting well-being. Unfortunately, multiple situations prevent us from reaching our physical activity goals, such as making bad decisions in the present, like eating non-healthy food because it is tasty and producing adverse effects on our fitness level. Another example could be surrendering to the lack of motivation for the gym training session and staying at home watching Netflix instead. Possible reasons for these behaviors can be explained by the lack of *good habits* or acquiring positive behavior patterns.

Fortunately, we have at our disposal tools, methods, and books to support reaching our physical activity objectives (e.g., [Miller and Frisch, 2009]; [Doran et al., 1981]; [Hyatt, 2018]; [Grove, 2015]). These are inspiring options for setting and reaching our goals. However, these solutions require some effort from the individual to assimilate the information and to put it into practice. Likewise, books will require time to read and digest. Most recently, fitness coaches have become popular in supporting individuals to reach their goals. However, availability to such professional is not widely spread, and accessible to everyone.

Luckily the technology field offers another approach that supports us in reaching our goals, which is envisioned in the concept of *behavior change apps*. These are intentionally designed "to foster and assist behavior change, and sustainment" [Hekler et al., 2013, p. 3308]. *Behavior change apps* are easy to access (i.e., main app markets `App Store` and `Google Play`), have zero or low cost (e.g., 1 USD to 10 USD), and many present goal-setting frameworks in an understandable and practical manner. Examples of these include: `MyFitnessPal` [Under Armour Inc., 2019] and `Yoga-Go` [A. L. Amazing Apps Ltd., 2019].

In the last years, the domain of *behavior change* apps has received tremendous interest and support because of the pervasiveness of mobile phones. Users worldwide have more than 100 thousand health apps at their disposal, and about 500 million people use mobile health applications [Edwards et al., 2016]; [Fox and Duggan, 2012]. More recently, in the past year, the COVID-19 breakthrough has contributed to the increased usage of health apps due to people's concerns about their mental health [Computerweekly.com, 2021].

Behavior change apps result in a convenient way for people aiming to reach health goals. These could be healthy eating, keeping a diet, losing/gaining/maintain weight, practice some sport, exercise or improving their health status [Bhuyan et al., 2016]; [Elavsky et al., 2017]. Nowadays, there is an app for almost any new habit or behavior a person wants to develop or sustain. However, despite all these promising characteristics and benefits, users of these behavior change apps fail to reach their goals because they *lack motivation* to pursue the intended activity.

My research lies precisely at the overlap between *human motivation* and *technology*. As mentioned earlier, there exist multiple behavior change apps available through the app markets. In some of these apps, their creators have received inspiration and support from psychological theories (e.g., Fabulous app, Happify app), but also a large majority lacks theoretical support. **Is it relevant to design theory informed behavior change apps?** Yes!, because the theoretical foundation provides guidance on what behavioral outcomes expect in response to a particular stimuli, this in consequence leads to more robust design and more effective apps.

In my research, I take as a foundation the human motivation theory named Self Determination Theory (SDT) [Deci and Ryan, 2000]. What makes this theory stand from other human motivation theories is that it does not consider motivation as a unitary concept; namely, it varies just in *amount*. Instead, it states that humans have different *types* of motivation. The main distinction is that humans have autonomous motivation and controlled motivation. The former occurs when we experience willingness, volition, or choice or, in other words, when we perform an activity with a sense of interest, enjoyment, and value. The latter refers to moments when we perform an activity to obtain a reward or avoid punishment; in other words, we do something because we feel pressured, demanded, or obliged. What is significant about this distinction is that authors have found that when people are more autonomously motivated, the performance, wellness and engagement are more significant than when experiencing controlled motivation.

Furthermore, the SDT posits that human beings have three Basic Psychological Needs (BPN). These are *competence* or the feeling of being confident and effective at the target activity. *Relatedness*, to feel cared for others or to feel care by others; to have a sense of belonging to groups that are important to the individual. And *autonomy* the feeling to experience self-endorsement and ownership of their actions. When these psychological needs are satisfied the individual experiences states of high-quality motivation, performance, and well-being.

Another essential remark concerning autonomous motivation is that SDT authors distinguish two types of autonomous motivation: *intrinsic* and *extrinsic*. The former is doing an activity because we find it interesting and enjoyable (e.g., reading a book about cooking because we enjoy gastronomy). The later concerns doing something because it leads to separable outcomes (e.g., studying a gastronomy book to obtain a cook certificate). Extrinsic motivation can become autonomous motivation when the individual can internalize the value of the activity and make it feel their own. Fostering intrinsic motivation is more appealing for sustained behavior change because the moving force of an individual to perform an activity originates from their internal self and not from external elements that might not be consistently available.

With the theoretical ground of this thesis introduced, I continue by stating my general research aim:

To understand how human motivation can be supported through technology. Precisely, determine how we can design behavior change apps to support the Basic Psychological Needs aiming to promote higher states of well-being and motivation in the context of physical activity.

From a personal perspective, what passions me about this research topic is identifying how behavior change apps could be made more effective in changing people's behavior and helping them attain their goals. This constant goal achievement will lead to happier and more fulfilled individuals who positively influence their close ones. Finally, having communities with more fulfilled individuals will set the foundations for flourishing and thriving societies.

0.2 Thesis Scope

The concept of *well-being* frames this thesis, and for experimental and exploration purposes, I chose physical activity as the application domain. As mentioned earlier, a large body of research demonstrates the positive association between physical activity and well-being. Further, this thesis focuses on the psychological aspects that influence the physical activity of individuals. It does not intend to study the physiological changes (e.g., changes in heart rate or body mass index) individuals might incur due to modifications in their behavior. We acknowledge that physiological changes might occur through the interventions presented in this thesis, which might be of interest for future studies. Additionally, it is essential to clarify that the focus of this thesis is on the study and design of theory-informed behavior change technology. Therefore, all the interventions presented in this thesis are mediated by technology and differ from other non-technology-based behavior change interventions (e.g., [Austin et al., 2013], [Slovinec D'Angelo et al., 2014]).

From a general perspective, this thesis introduces one particular path (of many) in which researchers and app designers can translate the Self-Determination Theory constructs into the design of physical activity mobile apps. Moreover, this thesis guides researchers and designers in understanding some of the characteristics apps should have to provoke behavioral changes. Further, this thesis presents steps to address the societal challenge of supporting individuals interested in improving their physical activity but lacking motivation.

0.3 Research Objectives

To begin exploring the research topic, I proposed the following research question:

RQ1: Can the Self-Determination Theory suggest characteristics that behavior change apps would need to possess to support behavior change interventions? Do behavior change apps on the market possess these?

To answer this question, I analyzed 208 apps from the Apple App Store, identified 12 design features present in these apps, and classified them according to the SDT. This work contributes with an *artifact* in the form of a simple design tool to evaluate how the app features and combination of these may support the motivational process of individuals towards reaching their goals. In chapter 1, I present the development of this research stream. The research output is materialized in a publication in the International Journal of Human-Computer Studies.

I leveraged the artifact mentioned above and designed an SDT-based steps tracking app to develop the next research stream and move towards my research aim. Precisely, the app comprises three main elements or features, each intentionally designed to support the Basic Psychological Needs. During four weeks, 49 participants interacted with the app. I chose *physical activity* as the application domain as it is one of the most reliable paths to achieve higher well-being states. The research question I formulated was:

RQ2: What are the individuals' perceptions about the hypothesized app features that aim to support autonomy, competence, and relatedness when it comes to improving their physical activity?

The findings of this study showed that the hypothesized app features included in the design contributed *empirical* evidence that they correctly mapped to the Basic Psychological Needs. This contribution gave us insights that moved us closer to understanding how to support the BPNs through technology in the physical activity domain. In chapter 2, I develop this research stream. The output of this research project is published on the ACM MobileHCI'21 Conference.

The next area to explore and move closer to the general research aim was to generalize the design of BPNs supporting features, and understand how the combination of these in the same interactive artifact can influence the overall behavior of its user. The research question I formulated was:

RQ3: What are the BPNs feature combinations that produce the most positive effects on the individuals' behavior when it comes to improving their physical activity?

To answer this question, I designed an experiment testing seven different feature combinations, with the objective of evaluating the behavioral outcomes for each feature combination. I evaluated the participant's intrinsic motivation, BPNs support, engagement, and physical activity level for each combination. At the moment of writing this thesis, this project is in an ongoing status; therefore, I present preliminary results in chapter 3. Our main finding shows that combining our *competence* and *relatedness* features yielded significant positive effects on the participants' physical activity and intrinsic motivation. The results of this study contribute *empirical* evidence that shows the positive effect of the *competence* and *relatedness* BPNs feature combination on the physical activity level, intrinsic motivation and engagement of our participants.

0.4 Complementary Research Streams

During my research, I also investigated the effects of tangible rewards in the form of monetary incentives and motivational messages on the motivation of people to walk. This study was conducted during a 10-month period in which participants used a mobile steps-tracking app and were requested to achieve 10K daily steps. When reaching the steps goal, they were rewarded. This research contributed *empirical* evidence showing the detrimental effect that these types of rewards have on the individuals' motivation. This research stream is developed in Chapter 4. The output of this research project is materialized in a publication on the ACM Transactions of Computer-Human Interaction.

Furthermore, I explored how to design technology for elderly at the light of the Persuasive System Design [Oinas-Kukkonen and Harjumaa, 2009], the results of this effort were presented at the The ACM Conference on Human Factors in Computing Systems CHI'2017 Workshop on Designing Mobile Interactions for the Ageing Populations [Villalobos-Zuñiga and Cherubini, 2017]. Another topic of exploration distilled from the longitudinal data analysis of the experiment described in chapter 4. I presented this research effort at CHI'2018 Workshop on Long Term Self Tracking [Villalobos-Zúñiga and Cherubini, 2018]. Finally and more recently, inspired by my interest in inclusive technology, I leveraged the Self-Determination Theory to suggest ways to design inclusive-behavior-change technology. This work is part of the Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization [Villalobos-Zúñiga, 2021].

To conclude these introductory paragraphs, I summarize the contribution of this thesis which is four-fold: (1) It introduces a taxonomy of behavior change app features that serves as an *artifact* to evaluate behavior change apps. This tool also guides designers and researchers to analyze various app features combinations that have not been fully explored. (2) It provides *empirical* evidence of physical activity app features that correctly mapped the Basic Psychological Needs. (3) This thesis provides *empirical* evidence on the Basic Psychological Needs app feature combinations that support an increase in intrinsic motivation and physical activity. (4) Finally, it provides *empirical* evidence on the harmful effects of tangible rewards and motivational messages on the motivation and physical activity of individuals.

Bibliography

[A. L. Amazing Apps Ltd., 2019] A. L. Amazing Apps Ltd. (2019). Yoga-Go: Weight loss workouts. Last accessed August 22, 2019.

[Austin et al., 2013] Austin, S., Guay, F., Senécal, C., Fernet, C., and Nouwen, A. (2013). Longitudinal testing of a dietary self-care motivational model in adolescents with diabetes. *Journal of Psychosomatic Research*, 75(2):153–159.

[Bandura and Wessels, 1994] Bandura, A. and Wessels, S. (1994). *Self-efficacy. Encyclopedia of human behavior*. New York: Academic Press.

[Bhuyan et al., 2016] Bhuyan, S. S., Lu, N., Chandak, A., Kim, H., Wyant, D., Bhatt, J., Kedia, S., and Chang, C. F. (2016). Use of Mobile Health Applications for Health-Seeking Behavior Among US Adults. *Journal of Medical Systems*, 40(6).

[Calvo and Peters, 2014] Calvo, R. A. and Peters, D. (2014). *Positive computing: technology for wellbeing and human potential*. MIT Press.

[Computerweekly.com, 2021] Computerweekly.com (2021). Covid-19 has led to a 25% increase in health app downloads, research shows. Last accessed July 29, 2021.

[Deci and Ryan, 2000] Deci, E. L. and Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological inquiry*, 11(4):227–268.

[Diener and Biswas-Diener, 2011] Diener, E. and Biswas-Diener, R. (2011). *Happiness: Unlocking the mysteries of psychological wealth*. John Wiley & Sons.

[Doran et al., 1981] Doran, G. T. et al. (1981). There'sa smart way to write management's goals and objectives. *Management review*, 70(11):35–36.

[Edwards et al., 2016] Edwards, E. A., Lumsden, J., Rivas, C., Steed, L., Edwards, L. A., Thiyagarajan, A., Sohanpal, R., Caton, H., Griffiths, C. J., Munafò, M. R., Taylor, S., and Walton, R. T. (2016). Gamification for health promotion: systematic review of behaviour change techniques in smartphone apps. *BMJ Open*, 6(10):6–0.

[Elavsky et al., 2017] Elavsky, S., Smahel, D., and Machackova, H. (2017). Who are mobile app users from healthy lifestyle websites? Analysis of patterns of app use and user characteristics. *Translational Behavioral Medicine*, 7(4):891–901.

[Fox and Duggan, 2012] Fox, S. and Duggan, M. (2012). Mobile Health 2012. Technical report, Pew Research Center.

[Frey and Stutzer, 2010] Frey, B. S. and Stutzer, A. (2010). *Happiness and economics*. Princeton University Press.

[Grove, 2015] Grove, A. S. (2015). *High output management*. Vintage.

[Hekler et al., 2013] 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 CHI'13*, page 3307, New York, New York, USA. ACM Press.

[Hyatt, 2018] Hyatt, M. (2018). *Your Best Year Ever: A 5-step Plan for Achieving Your Most Important Goals*. Baker Books.

[Jung, 1928] Jung, C. G. (1928). *Two Essays On Analytical Psychology*.

[L. Fredrickson and Levenson, 1998] L. Fredrickson, B. and Levenson, R. W. (1998). Positive emotions speed recovery from the cardiovascular sequelae of negative emotions. *Cognition & emotion*, 12(2):191–220.

[Maslow, 1943] Maslow, A. H. (1943). a Theory of Human Motivation. *Psychological Review*, 50(4):370–396.

[Miller and Frisch, 2009] Miller, C. A. and Frisch, M. B. (2009). *Creating your best life: The ultimate life list guide*. Sterling Publishing Company, Inc.

[Oinas-Kukkonen and Harjumaa, 2009] Oinas-Kukkonen, H. and Harjumaa, M. (2009). Persuasive Systems Design: Key Issues, Process Model, and System Features. *Communications of the Association for Information Systems*, 24(28):485–500.

[Ostir et al., 2000] Ostir, G. V., Markides, K. S., Black, S. A., and Goodwin, J. S. (2000). Emotional well-being predicts subsequent functional independence and survival. *Journal of the American Geriatrics Society*, 48(5):473–478.

[Ostir et al., 2001] Ostir, G. V., Markides, K. S., Peek, M. K., and Goodwin, J. S. (2001).

The association between emotional well-being and the incidence of stroke in older adults. *Psychosomatic medicine*, 63(2):210–215.

[Pressman and Cohen, 2005] Pressman, S. D. and Cohen, S. (2005). Does positive affect influence health? *Psychological bulletin*, 131(6):925.

[Ryan and Deci, 2000] Ryan, R. M. and Deci, E. L. (2000). When rewards compete with nature: The undermining. *Intrinsic and extrinsic motivation: The search for optimal motivation and performance*, page 13.

[Slovinec D'Angelo et al., 2014] Slovinec D'Angelo, M. E., Pelletier, L. G., Reid, R. D., and Huta, V. (2014). The roles of self-efficacy and motivation in the prediction of short- and long-term adherence to exercise among patients with coronary heart disease. *Health Psychology*, 33(11):1344–1353.

[Tov and Diener, 2009] Tov, W. and Diener, E. (2009). The well-being of nations: Linking together trust, cooperation, and democracy. In *The Science of Well-Being*, pages 155–173. Springer.

[Under Armour Inc., 2019] Under Armour Inc. (2019). MyFitnessPal: Fitness starts with what you eat. Last accessed August 22, 2019.

[Villalobos-Zúñiga, 2021] Villalobos-Zúñiga, G. (2021). Designing for gender-role differences through the lens of self-determination theory. In *Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*, pages 299–302.

[Villalobos-Zuñiga and Cherubini, 2017] Villalobos-Zuñiga, G. and Cherubini, M. (2017). Not a technology person: Motivating older adults toward the use of mobile technology. *arXiv e-prints*, pages arXiv–1707.

[Villalobos-Zúñiga and Cherubini, 2018] Villalobos-Zúñiga, G. and Cherubini, M. (2018). Activity self-tracking with smart phones: How to approach odd measurements? *arXiv preprint arXiv:1804.04855*.

[Zhu et al., 2019] Zhu, Z., Feng, T., Huang, Y., Liu, X., Lei, H., Li, G., Deng, D., Zhang, N., and Huang, W. (2019). Excessive physical activity duration may be a risk factor for hypertension in young and middle-aged populations. *Medicine*, 98(18).

Chapter 1

Apps That Motivate: a Taxonomy of App Features

"Outer changes always begin with an inner change of attitude."
— Albert Einstein, 1879–1955

1.1 Introduction to the Chapter

In this chapter, we introduce this thesis's first research stream. Exploring the field of behavior change applications requires having an understanding of the state-of-the-art of mobile market apps. This exploration frames this research effort where at the light of the Self-Determination Theory, we reviewed 208 apps from the APPLE APP STORE by systematically testing each app feature and classifying them according to the Basic Psychological Needs (BPNs) posited by the Theory.

1.2 Introduction

In the last few years, there has been a dramatic increase of apps that are intended to bring about positive behavioural change (e.g. losing weight, quitting cigarettes, learning a new language or reducing waste). We will refer to this group of applications as *behavior change apps* because these have been explicitly designed "to foster and assist behaviour change and sustainment" [Hekler et al., 2013, p. 3308].

Recent surveys reveal that over 100K health applications (apps) are available worldwide for smartphones; the most popular apps are for exercise, diet, and weight management, and 500M users use mobile health applications [Edwards et al., 2016, Fox and Duggan, 2012]. The reason for this growth is apparent: smartphones are pervasive and provide a unique opportunity to reach a broad audience of users. Also, behavior change apps have been increasingly used in app-based health promotion programs, which might have contributed to their diffusion [Lee et al., 2018].
Behaviour change apps use numerous strategies to modify the behaviour of the user. For instance, some apps pay the user a *reward* for completing the specified activity (e.g., `Clue` app [Clue by Biowink GmbH, 2019]). Others provide accurate *feedback* about the user's performance and how it evolves over time (e.g., `Goalify` app [GmbH, 2019]). Recently, the usefulness of some of these apps has come under scrutiny [Ferrara, 2013, Skarecki, 2015, Jebelle and Burrows, 2019]. Despite the wealth of Internet resources on behavior change, designers often face scarcity of professional guidelines or industry standard [Lister et al., 2014]. In some cases, apps fall flat in producing changes, specifically when we look at the long-term effects of these interventions [Jeffery et al., 2000, Harrison et al., 2015], or even backfire, which is what happens when an intervention triggers audiences to adopt the opposite target behavior [Erskine et al., 2010, Stibe and Cugelman, 2016].

In this work, we look at how psychological theory can inform design. By taking this approach we do not imply that all design is (or must be) informed by theory. Also, often the mapping between theory and practice is mediated by "real-world necessities, complexities, budget limits, stakeholder feedback, market testing and politics" [Stibe and Cugelman, 2016, p. 3]. However, we argue that psychological theories can suggest new avenues to designers and reveal new areas of inquiry to HCI researchers.
Therefore, it is relevant to ask: *Can psychological theories suggest characteristics that behaviour change apps would need to possess to support behaviour change interventions? Do behaviour change apps on the market possess these?* To this end, several behavioral theories exist, and many have been used extensively in the field of HCI. A non-exhaustive list includes: the Social Cognitive Theory (or SCT) [Bandura, 1986], Theory of Planned Behavior (or TPB) [Ajzen, 1985], the Trans-Theoretical Model (or TTM) [Prochaska and Di Clemente, 1983], the Health Belief Model (or HBM) [Rosenstock, 1974], and

the Goal-Setting Theory [Locke and Latham, 2002]. These theories focus on *observable behaviour*: they predict whether people might enact target actions (e.g., perform physical activity, hydrate regularly) based on various constructs (constructs are the basic determinants or mechanisms that a theory postulates to influence behaviour [Hekler et al., 2013, p. 3309]).

Many constructs of the theories above relate to *motivation*, which concerns what moves people to action. Within these theories, the concept of motivation is a unitary concept: it is typically undifferentiated for types, qualities, or orientations [Ryan and Deci, 2017]. For instance, theories such as the SCT or the HBM predict motivation from the "strength of one's beliefs about being able to achieve outcomes"
[Ryan and Deci, 2017, p. 13].Taking a different stance, Self Determination Theory (or SDT) [Ryan and Deci, 2000a] is especially different from other approaches to motivation because it emphasizes the different types and sources of motivation that impact the quality and dynamics of behaviour. SDT suggests that some forms of motivation are entirely volitional (i.e., they reflect one's interests and values) whereas others can be wholly external (i.e., when paid, coerced or otherwise pressured into doing something).

Several studies informed by the SDT show the effects of incentives on motivation [Deci and Ryan, 1985, Deci, 1971, Deci, 1972, Deci and Ryan, 1980]. These incentives, can be used to bootstrap the internalization process[1], however these can harm the motivation of people who are already intrinsically motivated.

While the theories above attempt to explain how internal antecedents to action (i.e., knowledge, beliefs, or attitudes) influence behaviour, SDT aims at explaining how external conditions hinder support to the internal processes of change that might lead people to adopt the target behaviour.

SDT has been shown to have applicability across multiple life domains [Deci and Ryan, 2008], and it has been used to describe the development of causal action and self-determined behaviour [Wehmeyer et al., 2017]. Furthermore, SDT-based interventions have been shown to have long-term benefits [Friederichs et al., 2015].

Scholars began systematically reviewing the design of apps that support behaviour change in order to categorize their features. Few of these efforts have attempted to relate behaviour change theories with the functionalities of these applications. A recent study classified exercise apps according to HBM, TTM, TPB, and SCT [Cowan et al., 2012]. More recent classifications also covered several behaviour change theories but did not look at SDT (cf. [Lister et al., 2014, Michie et al., 2013]).

[1]Namely, making attitudes or behavior part of one's own nature by assimilation.

In this paper, we review 208 behaviour change apps and perform a functional decomposition to identify the basic features of these apps that support behavior change. We will explain in detail the method in Sec. 3.5. As a theoretical lens to organize and evaluate these tools in their ability to support behaviour-change interventions, we apply the SDT.
From the analysis of the apps, we identified 12 distinctive features that support the constructs of SDT. Only 25.5% of the reviewed apps provide full support for all the constructs required by SDT. We find that certain mechanisms are widely supported in current applications (e.g., Reminders), and that there are design possibilities aligned with the theory which are under-explored (e.g., Intergroup Competition).
In Sec. 4.6 we demonstrate the value of the findings by discussing how the taxonomy suggests how behavior change apps features should be designed. Furthermore, we discuss relevant research gaps suggested by the taxonomy. Next, we review prior work.

1.3 Related Work

Human behavior is defined by Davis et al. as "anything a person does in response to internal or external events" [Davis et al., 2015]. These responses are often recurring (e.g., I am bored therefore I eat). Modifying the typical responses one gives to a situation (or set of stimuli) might prove hard, as these often provide gratification, safety, comfort and other forms of satisfaction that might not be available otherwise. Behavior theories employ a set of concepts, definitions and propositions that explain or predict responses to events or situations [Glanz et al., 2008]. Here we highlight a few points of distinction among the theories that brought us to focus on Self-Determination Theory for the study reported in this paper.

1.3.1 Behavior Theories and Human Motivation

Behavior theories are models, a simplified representation of reality. Every model has points of strength and weaknesses. Social Cognitive Theory is defined as an *ecological* theory as it focuses on the importance of context (i.e., the social and physical environment) as a determinant of health behavior [McLeroy et al., 1988]. Social Cognitive Theory offers interventionists clear targets to minimize external barriers to behavior change. However, it falls short when describing internal stages of change and the processes that determine this change [Rejeski and Fanning, 2019]. Similarly, other behavior theories such as the Health Belief Model or the Theory of Planned Behavior have an *extrinsic focus*: they are concerned with how specific belief-based antecedents determine behavior [Hagger and Chatzisarantis, 2014, Leavell, 2017]. These models emphasize decisional balance: the relative weight of perceived benefits as compared to perceived barriers to engage in a target behavior. Another theory that focuses on external determinants to action is the Goal-Setting Theory [Locke and Latham, 2002]. The theory involves the development of an *action plan* (an external artifact) designed to motivate and guide a person in attaining behavior change [Grant, 2012]. Although these theories support many behavior-change programs and they have lots of merits, recent research urged the community to complement this view by looking also at *internal* aspects of change (i.e., how individuals live, account for and cope with life changes) [Rapp et al., 2019, p. 2].

Therefore in this work we decided to focus on the level of analysis encompassing inner psychological changes. Here, the qualifier inner or internal has *not* to be mistaken by the term 'unconscious'.[2] When we use the adjective *inner* or *internal*, we specifically refer to psychological processes that lead individuals to recognize specific behaviors as part of one's world, the fabric of our intentions. Extrinsic incentives, external barriers and facilitators could be very important in a behavior-change intervention. However, in this work we focus on the internal aspects of change because these can produce long-term benefits to the individuals (we will come back this this point in Sec. 1.3.2 and in Sec. 1.3.3). Two prominent behavior theories provide constructs that can explain different types of motivation and the internal processes that can lead people to move across them: the Trans-Theoretical Model and SDT. TTM consists of five interrelated stages of change that are delineated with a time frame and tasks associated with movement through that stage [Kennedy and Gregoire, 2009]. Recent critics to the theory do not identify the qualitative differences between each stage [Davidson, 1998]. Other researchers question whether the stages should be ordered in a specific way – that each stage is linked integrally to instances of those following it [DiClemente, 2003]. More importantly, TTM does not distinguish between internal and external sources of motivation with respect to decisional balance [Kennedy and Gregoire, 2009].

[2]SDT posits that there are two types of motivated behaviors: those that are consciously chosen in the service of intrinsic or extrinsic needs (i.e., the self-determined behaviors) and the 'mindless' or automated behaviors (i.e., the non consciously chosen). We we will discuss this point more in detail in Sec. 1.3.3.

Regulatory style	Amotivation	External regulation	Introjection	Identification	Integration	Intrinsic motivation
Perceived locus of causality	Impersonal	External	Somewhat external	Somewhat internal	Internal	Internal
Processes	*Non relevance non intentionality*	*Compliance Reactance*	*Approval from others*	*Self-endorsement of goals*	*Congruence*	*Inherent satisfaction*

Figure 1.1
Various levels of human motivation postulated by SDT (adapted from [Ryan and Deci, 2000a]).

Conversely, SDT makes of the distinctions between exogenous vs. endogenous sources of motivation one of its core constructs. Next, we review SDT in detail.

1.3.2 Self-Determination Theory

SDT postulates that people have not only different amounts of motivation towards a certain activity but also –and more importantly– different *types* of motivation, specifically different orientations with underlying attitudes and goals that give rise to action [Deci and Ryan, 1985]. The most basic distinction is between *intrinsic* and *extrinsic* motivation. The former refers to doing something because a person finds it inherently interesting or enjoyable (e.g., reading a book), whereas the latter refers to doing something because it leads to a separate outcome (e.g., preparing for an exam). Furthermore, SDT proposes that there are several types of extrinsic motivation that differ in the degree of internalization (i.e., the degree to which the behavioral regulation is autonomous versus controlled). As described by Ryan and Deci "behaviors can be externally regulated, meaning they are directly controlled by external and self-alien forces; or they can be controlled through introjection, in which case the person has taken but not fully accepted external controls" [Ryan and Deci, 2017, p. 14]. This source of the regulation of the autonomous behavior plays a very important role in moderating the basic need of autonomy and is often referred to as *perceived locus of causality* [DeCharms, 1968]. Therefore, SDT organizes the different types along this *control–autonomy continuum*: *amotivation* (or absence of intention to act), *external regulation* (to obtain a reward), *introjected regulation* (to avoid guilt), *identification* (accepted external regulation), *integration* (self-determined action). Figure 4.1 presents the various levels of human motivation postulated by SDT. On this last stage the individual has acquired autonomous motivation towards the target activity.

According to SDT, extrinsic motivation types can encourage a person to behave a certain way in the short-term, but fail to maintain the behavior over time [Deci and Ryan, 1985]. Behavior-change interventions that are designed around extrinsic motivation types might not produce modifications of behavior that last after the intervention has ended. Conversely, when individuals reach intrinsic levels of motivation, they develop self-determined action towards the target activity. When this state is reached, interventions are no longer needed, and the changes in behavior become consolidated and persistent through time.

The theory posits that people have natural tendencies toward self-determined action. However, in order for this to happen people require contextual conditions to satisfy three basic psychological needs (or BPNs): *autonomy*, *competence*, and *relatedness* [Ryan and Deci, 2017]: autonomy refers to feeling willingness and volition with respect to one's behavior; competence refers to feeling effective in one-s interaction with the social environment; and relatedness refers to both experiencing others as responsive and sensitive and being able to be responsive and sensitive to them. When their basic needs are satisfied, people experience growth, integrity, and well-being. Conversely, when their psychological needs are not met, there could be psychological harm [Deci and Ryan, 2000]. Satisfying the BPNs pushes the individuals to move along the control–autonomy continuum to reach autonomous motivation towards the target activity. One point that is often debated in the literature which deserves further discussion is the role of *implicit processes*–processes for which the individual might not be consciously aware.

1.3.3 Conscious vs. Unconscious Influences to Human Behavior

In the last few years, research on psychology has moved away from models that focus exclusively on deliberative, intentional and explicit influences on behavior and towards theories that also account for the non-conscious, impulsive and implicit influences on behavior [Strack and Deutsch, 2004, Jonathan and Keith, 2009, Stanovich, 2010]. These approaches are referred to as *dual systems models* of motivation [Hagger and Chatzisarantis, 2015, p.20]. These recognize that behavior is a function of deliberative, volitional and planned inferences as well as those that are automatic, non-conscious, and unplanned. Recent research has demonstrated that automatic processes are key to habit formation, which in turn can be used for behavior change interventions [Pinder et al., 2018]. This line of research also pointed out that many interventions fail to

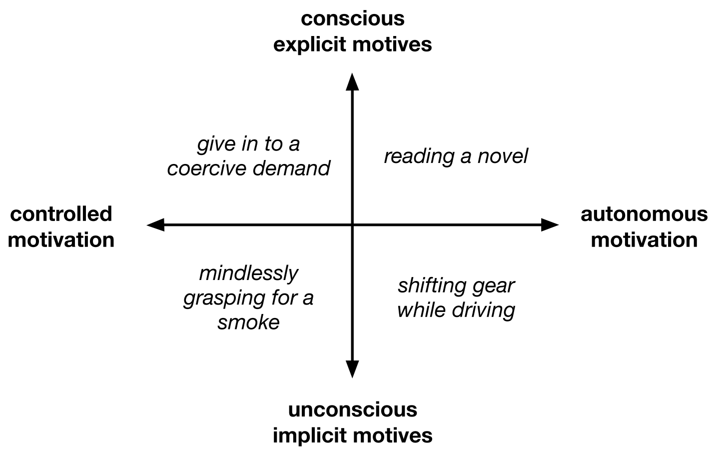

Figure 1.2: Contrasting graph of various types of motivational processes.

achieve behavior change because they neglect the role of automatic, non-conscious behavior [Verplanken and Wood, 2006]. Within the conceptual framework of SDT, the issue of conscious and unconscious motives needs to be distinguished from the issue of autonomous versus heteronomous motivation. These are orthogonal concepts, and present some interesting interfaces: implicit or unconscious events may prompt either autonomous or controlled behaviors, just as behaviors that are conscious may be regulated by either autonomous or controlled motivations [Ryan and Deci, 2017, adapted from p. 77]. Figure 1.2 presents examples of situations that are driven by different degrees of autonomous motivation and cognitive awareness. SDT also cautions that instigating behaviors *exclusively* through interventions that leverage unconscious mechanisms (for instance, *subliminal priming* [Pinder et al., 2017]) are at risk of making the person feel controlled as the locus of control is likely to be perceived external [Niemiec et al., 2010a, Schultz and Ryan, 2015]. While implicit and explicit influences are distinct, a number of experiments have revealed that when people are self-determined in their values and commitments, they also show congruence between their implicit and explicit motives and attitudes [Legault et al., 2007, Radel et al., 2017]. In other words, when target activities are intrinsically motivating, self-regulation is not needed to perform such tasks as those are inherently pleasurable: the reflective and impulsive systems of the mind are aligned [van Hooft, 2018]. Next, we looked for studies that classified the behavior change apps according to the support these provided to the BPNs.

1.3.4 Classifications of Behavior Change Apps

A large body of HCI research exists on the effectiveness of behavior change apps (see for instance [Consolvo et al., 2008,Fritz et al., 2014,Hsu et al., 2014,Purpura et al., 2011]). Most of this research focuses on the study of app design, as a whole, and on the effects that the resulting designs yield on behavior change. For example in relation to how behavior change apps can help users eat more healthily [Coughlin et al., 2015,Okumus et al., 2016], quit smoking [Abroms et al., 2011], exercise more [Consolvo et al., 2008], or cope with stress [Gimpel et al., 2015,Konrad et al., 2015], to list a few. More recently researchers started focusing on the distinct *features* of apps, because each feature could provide support to distinct cognitive processes [Heffner et al., 2015,Stawarz et al., 2014]. Therefore, to understand which specific aspect of app design relates to a particular change in behavior, it is necessary to decompose the app into its constituting functionalities.

In the last few years, two taxonomies were proposed to classify behavior change strategies and techniques. Oinas-Kukkonen et al. [Oinas-Kukkonen and Harjumaa, 2009] developed the Persuasive Systems Design framework that proposes 28 behavior change design principles. More recently, Michie et al. [Michie et al., 2013] presented a comprehensive 93-item taxonomy of theory-supported behavior-change techniques. More recently Caraban et al. [Caraban et al., 2019] conducted a systematic review of papers published in the last years in the domain of HCI and identified 23 strategies of behavior change. Unfortunately, for our purpose these classifications are of little use since they presented categories that are not specific to software features, and more importantly, they were not derived around the internal processes of change we discussed before. More recently scholars classified behavior change applications using a variety of strategies. Edwards et al. [Edwards et al., 2016] reviewed 64 apps from the health domain and classified their behavior change principles around 16 categories. They found no correlation between user rating (a possible proxy for health benefits) and game content or price. Similarly, Geuens et al. [Geuens et al., 2016] reviewed mobile apps designed for chronic-arthritis patients and derived 37 behavior change principles.

Unfortunately, these classifications were not based on theories of human motivation. More relevant for this research is the work of Lister et al. [Lister et al., 2014], who conducted an analysis of of *gamification* constructs in 132 apps that support individuals in their of physical activity and healthy dieting. They identified 13 behavior-change

constructs. Similarly, Stawarz et al. [Stawarz et al., 2015] conducted a review of 115 habit-formation apps and found 10 behavior-change technique. Cowan et al. [Cowan et al., 2012] performed content analysis on 127 apps from the 'Health & Fitness' category. Apps were generally observed to be lacking in theoretical content. Although these studies looked at the constructs from the angle of several psychological theories, they missed coverage of SDT. Some researchers in the digital-games domain have mapped game characteristics to SDT [Birk et al., 2016, Deterding, 2016, Kappen and Nacke, 2013, Ryan et al., 2006]. However, their work focuses on games exclusively and does not include a systematic evaluation of design features in relation to the basic psychological needs. Our specific interest is to categorize design features according to the BPNs specified by SDT. By using the three BPNs as drivers orienting our review we are implicitly focusing on those features that support individuals towards the autonomy end of the motivation continuum. We do this because we are interested in interventions that can produce *long-term benefits* to the user. Persistent changes are in fact required for many behaviors, such as diet and exercise, to reduce long-term risks (e.g., hearth disease, cancer) [Haskell et al., 2007]. We therefore pose our research question as follows:
RQ: *What features of behavior change apps support the BPNs for human motivation?*

Next, we describe our research methodology.

1.4 Methodology

We conducted a systematic review and analysis of apps on the `Apple App Store`. We chose to focus on the Apple App Store exclusively because, as we will detail next, the review process required several weeks to complete the analysis of each app. As we lacked resources, we could not extend the review to other stores. Additionally, we selected the Apple store because, at the time of the study, this had the largest market share in Switzerland [Statcounter, 2020, Minutes, 2020]. We identified apps aiming at supporting users who were willing to change their behavior. We then performed a functional decomposition to extract the main features of the apps that were relevant for behavior change, and coded the app features. Finally, we mapped them to the SDT BPNs. A flowchart detailing the features analysis is available online.[3]

[3]See https://osf.io/zy78r/, last visited June 2021.

This method was derived from previous studies (cf. [Alharthi et al., 2018, Edwards et al., 2016, Stawarz et al., 2015, Stawarz et al., 2018]) and adapted to our specific research goals. Following the approach used in previous work, we decided to not use sampling as a mechanism of app selection. Instead, we collected the "top-rated" apps, guaranteeing a cross-section of popular apps. However, using this approach alone might miss out interesting examples of behavior change apps that were not in the top 100 charts at the time of the review. For this reason, we complemented it with a keyword-based search following also similar studies (cf. [Lister et al., 2014, Stawarz et al., 2014, Lyngs et al., 2019]). We note, that any keyword selection is arbitrary to a lesser or larger extent. However, if the same set of keywords is used at a given time point and from the same place, the selection process is perfectly reproducible[4].

1.4.1 Data Collection Procedure

Top-Down Search. We began our data collection process by looking at the Apple App Store top 100 charts[5] in all 23 categories[6] as reported by `AppAnnie` [App Annie Inc., 2019], a business intelligence provider, on April 5^{th} 2019. Taking our definition from Hekler et al. [Hekler et al., 2013], we established that the app description in the App Store should report that the app was designed *purposely* to foster and assist behavior change and sustainment (e.g., stop smoking, sustain an active lifestyle), to form good habits (e.g., meditate everyday, drink 2 liters of water per day), or to improve skills (e.g., learn to cook vegan, learn a new language). For example, in `Asana Rebel` [Asana Rebel GmbH, 2019], the App Store description includes the following text: "Get motivated and build lifelong habits with proven, unique, modern methods". The previous text matched our inclusion criteria as it contained the keywords *motivation* and *habit*. Whereas, the following example did *not* match our inclusion criteria as it describes an app that only provides cooking recipes: "Get 40+ free healthy recipes and kitchen hacks! Your complete healthy recipe book ..." `Runtasty` [Runtastic Inc., 2019]. Using this criteria, we reviewed 2300 apps, and we selected and downloaded 244 apps for the next step. See Figure 1.3 for a diagram of the data collection process.

[4] Available apps often change in the online stores. If the selection is repeated months apart the results might differ. Similarly, querying the app stores from different places in the world might yield different results because these are customized based on the place from where the query is issued.

[5] The Top Charts represent the rankings on the official App Store.

[6] i.e., Books, Business, Education, Entertainment, Finance, Food & Drink, Health & Fitness, Kids, Lifestyle, Magazines & Newspapers, Medical, Music, Navigation, News, Photo & Video, Prod., Ref., Shopping, Soc. Network., Sports, Travel, Utilities and Weather.

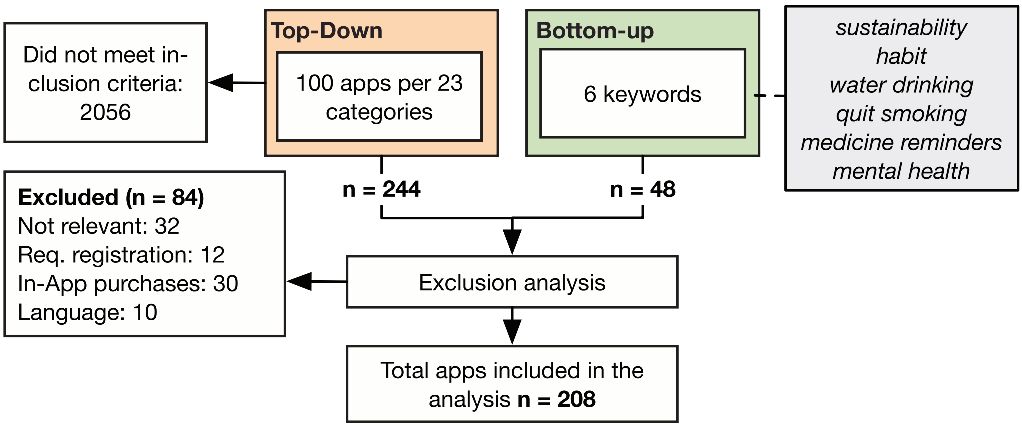

Figure 1.3
Selection process of the behavior change apps.

Bottom-Up Search. To increase the diversity of our sample, we decided to employ also a *bottom-up* approach searching for behavior change apps by keywords. We did this because the top 100 charts we used for app selection included exclusively the top user-rated apps, hence increasing the chances that less popular behavior change apps (i.e., the long tail) could be excluded from our classification. For instance, this was the case of *Simple Habit* [Simple Habits Inc., 2019] that includes several behavior change features but is ranked 158 in category `Health & Fitness`, as of May 2019. As a first step, we compiled a list of *excluded keywords* (e.g., *health*, *fitness*, *education*) from the apps we had already selected through the top-down approach. We then extracted search keywords from recent research on behavior change applications. The keywords used were: *sustainability* [Brynjarsdottir et al., 2012, Midden and Ham, 2018], *habit* [Purpura et al., 2011, Renfree et al., 2016], *water drinking* [Lally et al., 2010], *quit smoking* [Graham et al., 2006, Khaled et al., 2009], *medicine reminders* [de Oliveira et al., 2010, Stawarz et al., 2014], and *mental health* [Grist et al., 2017] (see Table 1.2 for the specific references). The choice of this particular set of papers was based on the related work review, references on social media, and the authors' experience. We then extracted the list of keywords from each of the papers and we matched those against the excluded keywords. The keywords that were uncommon between the two sets were used as *search keywords* for the bottom-up approach. In short, we specifically avoided selecting keywords that were already covered extensively by the top-down search. The final list of selected keywords (and their link to related literature) is reported on Table 1.2. We input each keyword in the search bar of the iPhone App Store app. Then, following the top-down approach described above, we reviewed the description of each app that appeared on the search results. The apps that matched the selection criteria were selected for the next step. As the App Store continued to load search results as we scrolled, we establish the saturation point when the last 10 apps do not meet our inclusion criteria. Finally, we included a total of 48 apps in the sample after performing the 6 keyword-based searches.

Exclusion Criteria. Before moving to the next phase, we had to exclude a 84 apps for one or several of the following reasons: i. Tools intended for general productivity that were selected because of misleading descriptions (e.g., `Snap Calc` [IAC Search & Media Technologies Limited, 2019]); ii. Tools that required corporate subscriptions or other types of registration (e.g., `Success Factors` [inc., 2019b]); iii. Apps that hid features behind in-app purchases (e.g., `Sworkit Fitness` [Nexercise Inc., 2019]); iv. Apps whose locale was not English, French, or Spanish as these are the languages authors are familiar with (e.g., `Chefkoch SmartList` [Chefkoch GmbH, 2019]). This resulted in 208 relevant apps. Figure 1.3 shows the process with which apps were selected and excluded from the sample.

1.4.2 Apps testing

The test of the apps was completed during the eight months that followed the apps selection. Each app was installed on a device and tested for several hours across multiple weeks by the first author of this paper. Multiple apps were tested in parallel.

Keyword	Selection Rationale	Ref.
sustainability	Develop environmental-friendly behavior.	[Brynjarsdottir et al., 2012, Midden and Ham, 2018]
habit	Quit/curb bad habits or start good habits.	[Purpura et al., 2011, Renfree et al., 2016]
water drinking	Encourage staying hydrated.	[Lally et al., 2010]
quit smoking	Smoking cessation support.	[Graham et al., 2006, Khaled et al., 2009]
medicine reminders	Medication compliance and adherence.	[de Oliveira et al., 2010, Stawarz et al., 2014]
mental health	Promotes mental well-being.	[Grist et al., 2017]

Table 1.1
Keywords extracted from literature on behavior change applications. These were used in the bottom-up search. The last column indicates the reference to the literature from which the concept was taken.

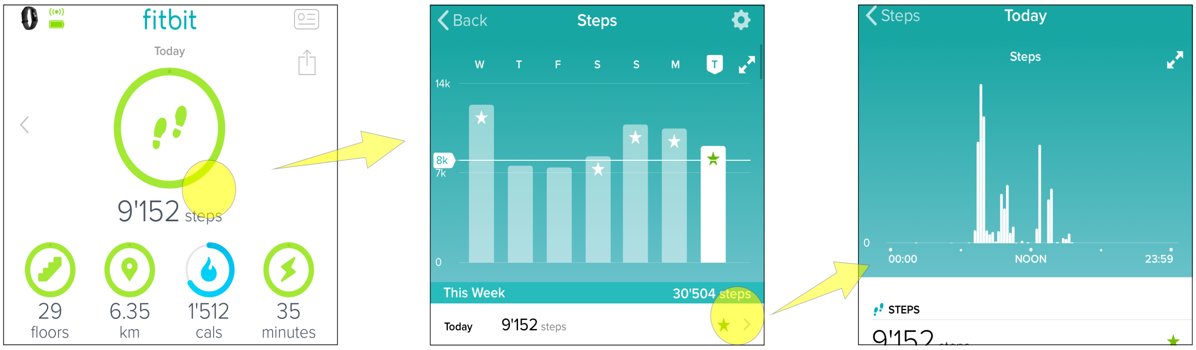

Figure 1.4
(left) Dashboard of the Fitbit App (main sub-system, cropped). Tapping the number of steps switches the view to the number of steps sub-system (center). Tapping on a given day brings the user to the daily number of steps functionality (right). At this level, there are no further sub-systems that can be decomposed.

Approximately 36 apps per month were installed during the course of 8 months. More in details, we followed these steps: i. completed the warm welcome in each app (if available); ii. created accounts whenever this was required or suggested by the app; iii. performed the actions suggested by the apps whenever possible (see details below); iv. reacted to app notifications. Target actions were performed when suggestions of the apps fit the personal schedule of the researcher and when she felt motivated to perform them (e.g., reading a book, drinking a glass of water, taking a language module, going for a run). In some cases, target actions had to be simulated as not applicable in the life of the researcher (e.g., tracking intake of birth control pills). Performing the target actions ensured familiarization with the data that had to be coded [Braun and Clarke, 2006] (i.e., the features of the app). This step is required by the methodology we used to code the features (see Sec. 1.4.4). During this time, unstructured notes and thoughts were captured in the researchers' diary. Particularly, the researcher documented whether she felt pressured by the app to perform the target activity, whether she felt competent to perform the target activity, and whether she felt supported by others. These three elements corresponded to the BPNs of the SDT and informed the subsequent coding process. Furthermore, testing the features across several weeks provided us with a longitudinal exposure to the app features that could reveal adaptability of functionalities over time (e.g., delivery of reminders at particular times of the week based on the user's activity). The features that tailored interventions during this time frame were analyzed. Next we describe the functional decomposition process.

1.4.3 Functional Decomposition

To extract the main functionalities of each app, we followed the guidelines of *functional decomposition* proposed by Chiriac et al. [Chiriac et al., 2011]. The first step of functional decomposition, involves dividing the system (or app) into *self-controlled* sub-systems (or functions). The second step requires identifying how these sub-systems interact with each other. These two steps define the first level of decomposition. On the second level of decomposition, each of the sub-systems are decomposed into other subsystems. We stop the process when the next decomposition level reaches the level of basic UI components (e.g., buttons, labels, icons, sliders). As an example, while performing the decomposition process for the `Fitbit` app [Fitbit Inc., 2019], we first identified the *dashboard*, *challenges*, *community*, and *notifications* as self-controlled sub-systems (see Figure 1.4). By tapping on the steps indicator (top of the screen), the system presents the *list of steps* sub-system (see Figure 1.4, center). Then, when tapping on "Today", the system showed the *steps number* sub-system that can not be further decomposed (see Figure 1.4, right). Next, we classified the sub-systems.

1.4.4 Features Coding

We coded the sub-systems identified during the functional decomposition process using a procedure adapted from *thematic analysis* [Braun and Clarke, 2006]. The analysis process involves two stages. The first stage consists in grouping together sub-systems that despite visual design differences provide the same basic functionality. These groups are formed and labeled using an inductive approach that started from the analysis of the sub-system derived from the first 30 apps. These formed the initial codebook. Then, both codes and their definitions are updated as new apps are analyzed. We stopped refining the codebook around the hundredth app, because we kept seeing recurring functionalities and no significant changes occurred. Once the codebook was stable, two coders (the first author and a master's student) coded independently all the identified sub-systems of the apps. In the second stage, overarching themes are evolved from the more granular data. For example, the functions that provide statistics concerning the number of activities performed or the task performance formed the theme: 'Activity Feedback'. We clustered codes where we had evidence from the literature that these were providing the same support to the behavior change process. This phase was done collaboratively by the first and second author of the paper. These themes are the patterns under which we organized the results section and we will refer to them as *Behavior Change Features*. Next, we classified the Behavior Change Features according to which Basic Psychological Needs they cater to.

1.4.5 Theory-driven coding

As discussed, the codebook we applied during this last step was theory-driven and derived from the SDT [Ryan and Deci, 2017]. Basically, we applied one of three labels: Autonomy, Competence, or Relatedness. The link between each behavior change feature and the BPN it caters to was conceptualized applying one or multiple of the following heuristics: i. studies that demonstrated effects of the feature on the BPNs (the reader will find the relevant references in the next section); ii. perception of the effects of the feature on the BPNs, as noted by the first author who tested the feature on herself, and as discussed with the second author during the analysis process; iii. comparative analysis of the variation of designs identified in the sample of behavior change apps. In the majority of cases, applying the three heuristics above led the authors to associate each behavior change feature to only one BPN. This was also facilitated by the fact that BPNs are in their definition orthogonal to one another. However, we found three features which cater to two BPNs (as it emerged from the comparative analysis of apps). In these three cases, we classified the feature according to the BPN with the highest relevance (i.e., primary classification), but also to the additional BPN that could receive support depending on how the feature was implemented (i.e., secondary classification). We will discuss these specific cases in the next section. In addition, one might expect that each feature, depending on its particular implementation, to be influencing a BPN to a lesser or higher degree. This is possibly true, but assessing the exact extent to which a particular implementation influences a BPN goes beyond the scope of the present study, and it likely requires controlled experiments with full factorial designs. Figure 1.5 presents the coding process. As for the previous coding step, the first two authors of the paper independently assigned the features to the BPNs.

1.4.6 Inter-rater Reliability

We used Cohen's κ to assess inter-rater reliability for the thematic analysis (data-driven coding) and for the BPNs coding activity (theory-driven). For the data-driven coding,

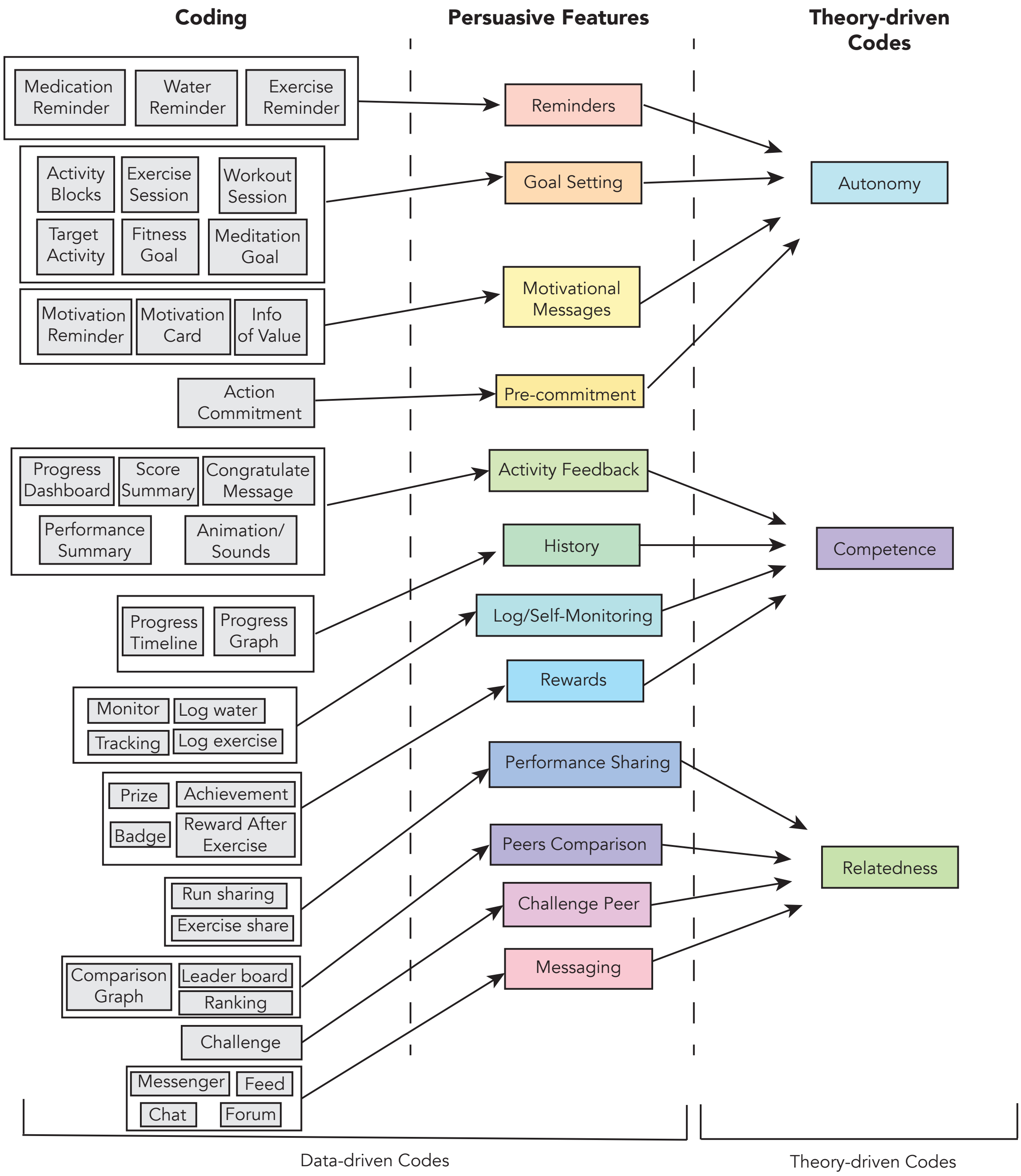

Figure 1.5
Diagram showing the taxonomy creation process. From left to right the Data-driven Codes which includes the coding and the clustering of codes to generate the themes/features. Next, the Theory-Driven Codes which include the three SDT Basic Psychological Needs as categories.

we measured an agreement of $\kappa\omega = 0.85$ (with 95% CI 0.81 to 0.89). For the theory-driven coding, we measured an agreement of $\kappa\omega = 0.91$ (with 95% CI 0.87 to 0.95). The level of agreement was considered sufficient to warrant the subsequent analysis of the data. All discrepancies were then resolved through discussion with a third trained reviewer.

1.5 Taxonomy

After the functional decomposition and coding of the features of the behavior change apps, we identified 12 features that cater to the BPNs defined by the Self-Determination Theory [7], last visited June 2021. In the following paragraphs, we introduce each feature and detail which characteristics was related to the satisfaction of the basic psychological needs. For each feature, we provide examples taken from the functional decomposition. See Table 1.5 for a summary of the findings.

1.5.1 Autonomy-Supportive Features

"Autonomy refers to feeling willingness and volition with respect to one's behavior. The need for autonomy refers to the need of an individual to experience self-endorsement and ownership of their actions" [Ryan and Deci, 2017, p. 86].

Reminders. This feature is often implemented as a message delivered around the time the user should perform a specific activity. The times of delivery of these

[7] The complete list of apps coded in this research is available at https://osf.io/zy78r/

Keyword	Selection Rationale	Ref.
sustainability	Develop environmental-friendly behavior.	[Brynjarsdottir et al., 2012, Midden and Ham, 2018]
habit	Quit/curb bad habits or start good habits.	[Purpura et al., 2011, Renfree et al., 2016]
water drinking	Encourage staying hydrated.	[Lally et al., 2010]
quit smoking	Smoking cessation support.	[Graham et al., 2006, Khaled et al., 2009]
medicine reminders	Medication compliance and adherence.	[de Oliveira et al., 2010, Stawarz et al., 2014]
mental health	Promotes mental well-being.	[Grist et al., 2017]

Table 1.2
Keywords extracted from literature on behavior change applications. These were used in the bottom-up search. The last column indicates the reference to the literature from which the concept was taken.

Feature	A	C	R	Coverage %	Example of App	Figure no.
Reminders	•			71.8	30 Day Fitness Challenge	1.6
Goal Setting	•			55.9	Feastr	1.9
Motivational Messages	•			9.1	Kwit	1.13
Pre-commitments	•		○	0.9	Stick	1.15
Activity Feedback		•		40.9	Babbel	1.17
History		•		30.9	Peak	1.20
Log/Self-Monitoring	○	•		29.6	Round	1.21
Rewards		•		19.1	GreenApes	1.23
Performance Sharing			•	18.2	MindShift	1.26
Peer Comparison		○	•	11.8	Duolingo	1.29
Challenge Peer			•	11.4	Fitbit	1.31
Messaging			•	6.4	Goalify	1.32

Table 1.3
Behavior change features and % of coverage in the sample, including examples of apps containing the given feature. A: autonomy, C: competence, R: relatedness, •: primary classification, ○: secondary

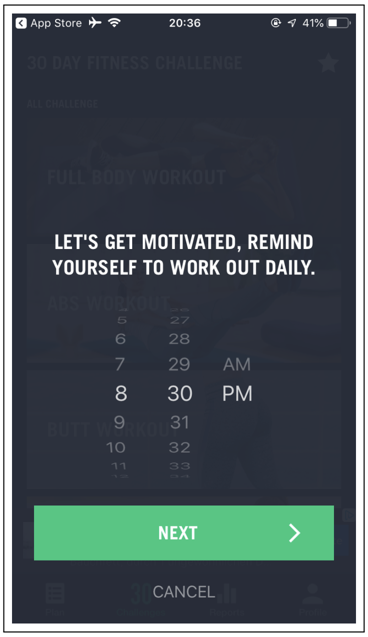

Figure 1.6
Reminders: notification set by the user (`30 Day Fitness Challenge` [Abishkking Ltd., 2019]).

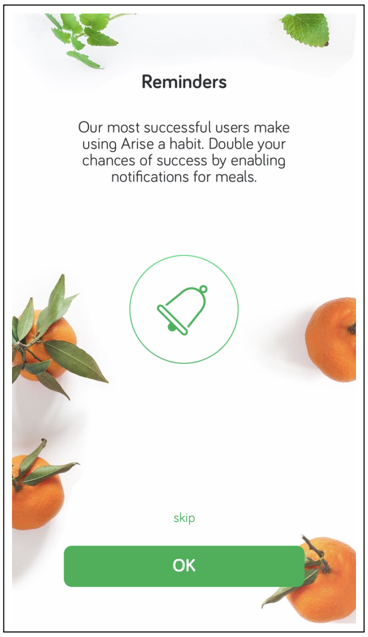

Figure 1.7
Reminders: notification set by the app (`Arise` [ARISE Ltd., 2019]).

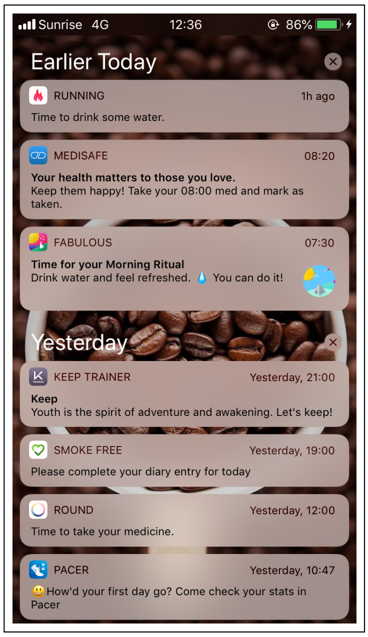

Figure 1.8
Reminders: examples of notifications

messages are identified by the app (see Figure 1.7) or input by the user (see Figure 1.6). Most often, this feature is implemented as a push notification, and we could observe it in most of the apps we reviewed and in all categories of apps: in educational apps (e.g., `iTranslate Lingo` [iTranslate GmbH, 2019], `Memrise` [Memrise Inc., 2019]), fitness apps (e.g., `30 Day Fitness` [Abishkking Ltd., 2019]) and productivity apps (e.g., `Better Habits` [Betterment, 2019], `Today` [The Today App Ltd., 2019]). See Figure 1.8 for an example.

SDT considers that self-organization is a natural effort, toward which individuals lean, and that it occurs under *autonomy* supportive conditions [Niemiec et al., 2010b]. In this context, reminders help the individual stay organized and on track with regard to the target behavior. SDT cautions that, depending on how this feature is implemented, the user might experience the opposite of autonomy, namely *heteronomy*, as when a users acts out of pressures that are experienced as controlling [Ryan and Deci, 2017, p. 86]. Therefore, it is best for reminders to be set by the user (e.g., as for the app '30 Day Fitness'), rather than set by the designers of the app (e.g., `Arise app` [ARISE Ltd., 2019]). When this feature is not perceived as controlling, it can gently move the person to perform the specified activity. For instance, receiving *one* reminder a day to track breakfast for the first three weeks could be considered OK by most users willing to change their eating habits. However, sending multiple reminders every day might have a negative effect on the person, reduce their self-determined interest in using the app, and eventually cause them to stop being interested in the activity [Mehrotra et al., 2016]. Renfree and colleagues [Renfree et al., 2016] also studied the reminders feature on `coach.me` [Lift Worldwide, 2019], the reminders sometimes caused negative reactions because they were deemed annoying, particularly when participants were going through busy or stressful periods.

Goal Setting. This feature provides the user the possibility to input or define the target for the activity they will perform. We identified three patterns: (1) the user proactively sets goals up (e.g, `MyFitnessPal` [Under Armour Inc., 2019]); (2) the app prompts the user directly about what their goals are (e.g., `Feaster` [Feastr GmbH, 2019]); and (3) self-competition: a bid against oneself or against a previously obtained result. For instance, in `Yazio` [Yazio GmbH, 2019] users can challenge themselves on the time elapsed since last eating chocolate. The pattern (2) typically occurs during *onboarding*, and it is pivotal to defining the subsequent interaction (e.g., `fatsecret` [Fatsecret Ltd., 2019]). See Figures 1.9 and 1.10 for examples of this feature. Another way this feature is embedded in behavior change apps is by letting users choose, from a series of predefined activities/exercises, which one they want to perform. These are usually presented as individual blocks labeled with the activity name, or as a list with illustrative icons. Most of the time, these include a description and the requirements to

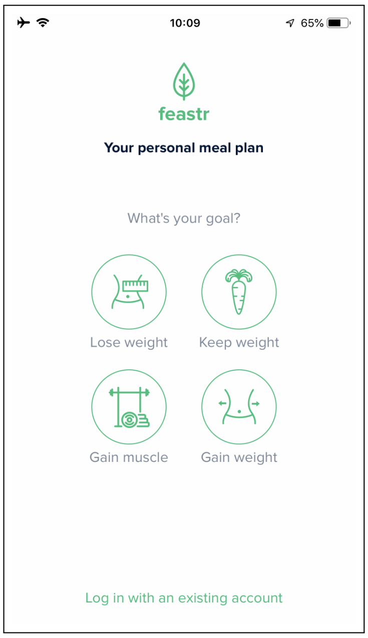

Figure 1.9
Goal Setting: activities suggested by the app (`Feastr` [Feastr GmbH, 2019]).

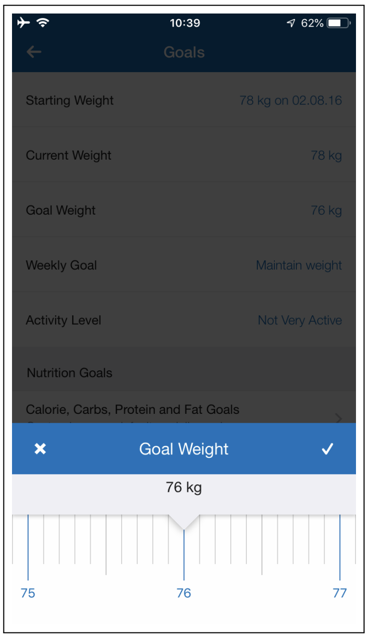

Figure 1.10
Goal Setting: goal defined by user (`MyFitnessPal` [Under Armour Inc., 2019]

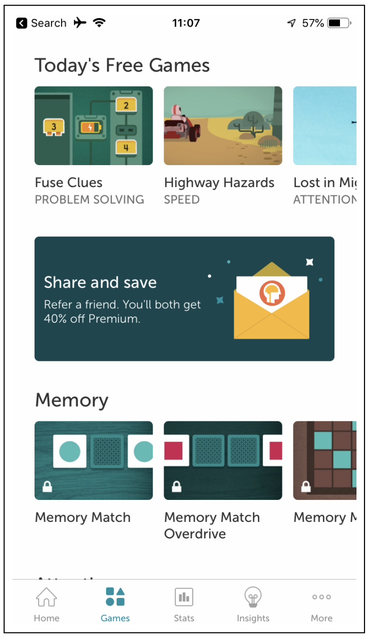

Figure 1.11
Goal Setting: activities suggested by the app (`Lumosity` [Lumos Labs Inc., 2019]).

accomplish it successfully (e.g., `Fabulous` [Fabulous Ltd., 2019], `Luminosity` [Lumos Labs Inc., 2019]). See Figures 1.12 and 1.11 for examples.

The goal-setting feature supports the SDT basic psychological need of *autonomy* because it contributes to an *internal perceived locus of causality* [DeCharms, 1968]. Specifically, an intentional behavior can be either intrinsically motivated (it has an internal perceived locus of causality) or extrinsically motivated (it has an external perceived locus of causality). For example, researchers found that in a learning environment where teachers gave students choices and options, the learning outcomes of the students increased [Deci et al., 1996, Patall et al., 2010]. According to the SDT, a goal imposed by someone else (or by an app) would undermine autonomy, which in turn would reduce the motivation of the subject to perform the activity. Whereas, a self-imposed goal would contribute to the basic need of autonomy and support autonomous motivation to perform the specific activity (cf. [Deci and Ryan, 1980, Ryan et al., 1985, Ryan et al., 1996]).

Motivational Messages. Some apps present the user with pre-set messages that explain why performing the specific activity is good for their health or well-being. These messages are often displayed in the shape of cards (e.g., `Kwit` [Kwit Ltd., 2019],

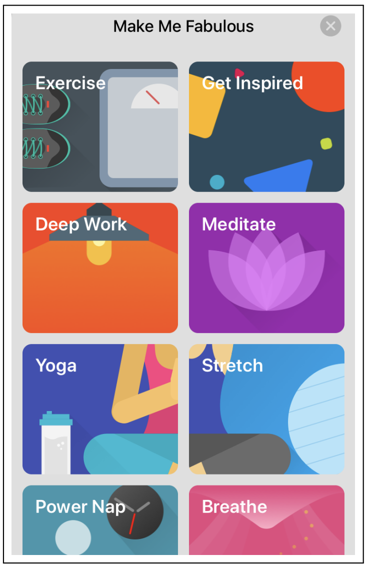

Figure 1.12
Goal Setting: choice of activity (`The Fabulous` [Fabulous Ltd., 2019]).

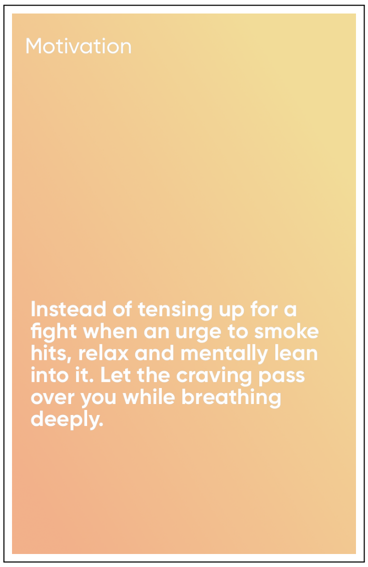

Figure 1.13
Motivational Messages: generic feedback not related to the user activity (`Kwit` [Kwit Ltd., 2019])

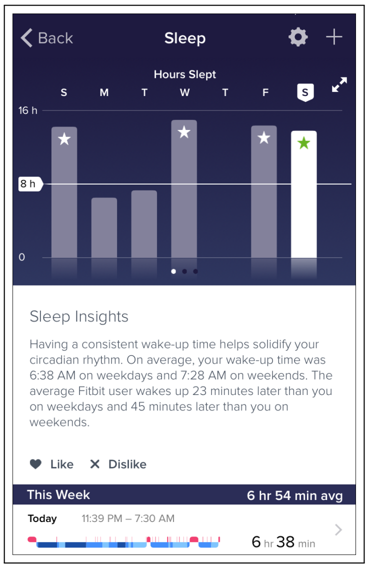

Figure 1.14
Motivational Message: task-specific feedback (`Fitbit` [Fitbit Inc., 2019]). The view displays also Feedback (top and bottom).

`Stop Smoking` [d bel Ltd., 2019]). Other instances for this feature take the form of instructional videos (e.g., `Yoga-Go` [A. L. Amazing Apps Ltd., 2019]). An important quality of these messages, from an SDT standpoint, is that these are not *task inherent*, meaning that they are provided to the users at specified time intervals, regardless of completion of the target behavior or performance. See Figures 1.13 and 1.14 for examples of how these are typically implemented in apps design. Another pattern this feature can take is the *letter to self*: a text field in which users can type a short message about why it is important for them to keep engaging with the target activity. This is then saved in the app and made available whenever needed (e.g., `MyPlate` [Leaf Group Ltd., 2019]) or it gets automatically resurfaced at specific time intervals (e.g., `Brainbuddy` [AppStudio Australia Pty Ltd, 2019]).

From an SDT perspective, these messages can help the user reflect on the reason they want to engage in the specific activity, hence they have the potential to support the basic need of autonomy. Kinnafick et al. studied the effects of supportive text messages on a person doing physical activity and found that regularly receiving this content increased their levels of intrinsic motivation [Kinnafick et al., 2016]. Concerning self-directed messages, encouraging the users to write a message to themselves, is a forcing function that lets them write down the specific reasons they wanted to change their

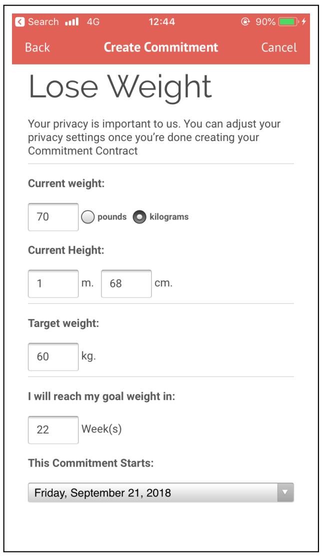

Figure 1.15
Pre-commitment: self-challenge setup page (`Stickk` [StickK Ltd., 2019]).

Figure 1.16
Pre-commitment: self-challenge setup page (`Tiny Habits` [B.J. Fogg, 2019]).

behavior. This contributes to changes in the person's regulatory style, toward more internal and integrated forms of motivation that are associated with autonomous control. Bargh shows how actions that are initiated by the individual have outcomes stronger than those that are more implicit or unconscious [Bargh, 2007].

Pre-commitments. This feature enables users to create *commitment contracts*: a binding agreement the users signs with themselves. Basically, it asks the user to define a goal, a given time frame to accomplish it, and a penalty if the deadline is not met. Typically, the penalty consists in donating a specific amount of money to a charitable organization of choice. Then, to establish whether the challenge was truly accomplished, the challenge is shared with other users (of the same app) who might act as referees. Examples of this feature can be observed on `Stickk` [StickK Ltd., 2019], see Figure 1.15 and 1.16 for visual examples. This feature contributes to the satisfaction of the SDT basic psychological need of *autonomy* because users can bind, or pre-commit, their own behavior [Wertenbroch, 1998]. If the implementation requires other users to act as referees, then the feature contributes as well to the satisfaction of *relatedness* as the users will feel connected and involved with others.

1.5.2 Competence-Supportive Features

"Competence refers to feeling effective in one's interactions with the social environment–that is, experiencing opportunities and support for the exercise, expansion, and expression of an individual's capacities and talents" [Ryan and Deci, 2017, p. 86].

Activity Feedback. This feature provides the user with information about how the task was performed in a given session and it might also present the user with details on the overall progress towards completing a predefined set of activities. This information might also be accompanied by a score that represents the performance (e.g., `Babbel` [Lesson Nine GmbH, 2019]), or a small encouragement message (e.g., `Duolingo` [Duolingo Inc., 2019]). Furthermore, a different instance of this functionality might also present cumulative statistics aggregated over a given time period (e.g., a week, a month); this might help the user evaluate temporal trends and compare current with past performances. For instance, these scores might include the consecutive or non-consecutive days in which a given activity was completed (e.g, `Calm app` [inc., 2019a] shows total number of meditation sessions). See Figures 1.17, 1.18 and 1.19 for examples of this feature.

Both session-specific and cumulative statistics provide *feedback* to the users of behavior change apps. When this feedback is positive, showing growth or improvement trends, this can enhance an individual's sense of competence [Deci and Ryan, 1980]. However, when feedback is negative, this can have the opposite effect, particularly when the information does not provide any actionable advice [Sjöklint et al., 2015]. Unsatisfactory results, such as underachievement, do not lead to behavior change but

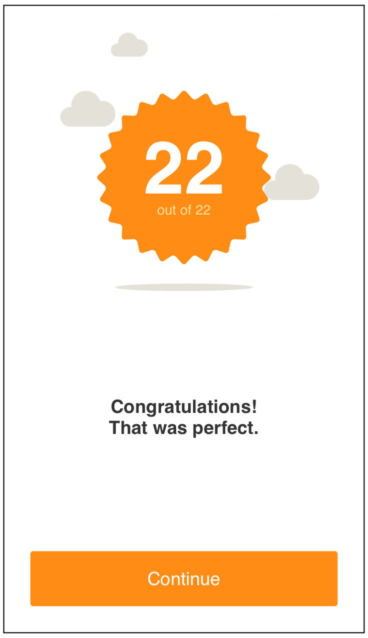

Figure 1.17
Activity Feedback: numerical score after training session (Babbel [Lesson Nine GmbH, 2019]).

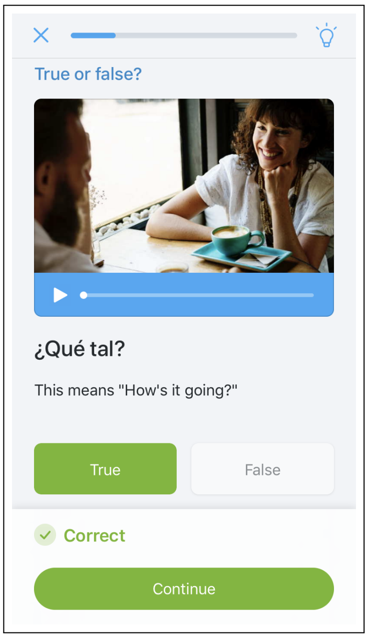

Figure 1.18
Activity Feedback: correct response is visualized after True/False question (Busuu [Busuu Ltd, 2019]).

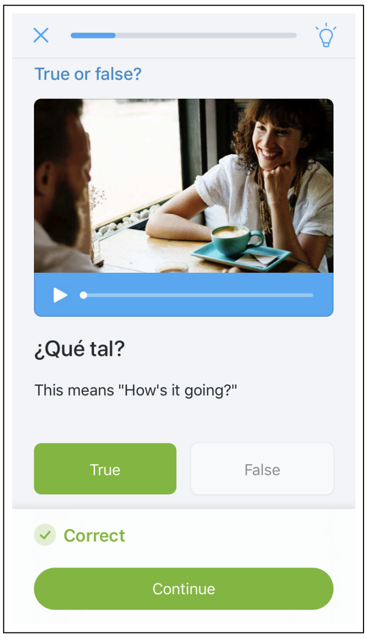

Figure 1.19
Activity Feedback: correct/incorrect answers to quiz (iLingo) [iTranslate GmbH, 2019]

rather the emergence of coping tactics: e.g., disregard, procrastination, and neglect. Research conducted on feedback also revealed that the *timing* of delivering feedback also plays a role in its overall effect on motivation. When feedback in *unexpected*, receiving this information does not make people feel being controlled. When feedback is experienced as an evaluation, pressure or control, it prompts people to perceive the activity as imposed on them (i.e., external perceived locus of causality), hence it undermines intrinsic motivation [Smith and Sarason, 1975]. SDT states also that feedback alone might not be enough to motivate users. Ryan demonstrates that simply providing positive feedback is not enough to motivate people if they do not also experience autonomy [Ryan, 1982]. Another quality of feedback that was found to be connected with its effectiveness in motivating recipients is that feedback must be specific to the performed task (i.e., *task inherent*) and actionable [Hewett and Conway, 2016]. In summary, feedback enables recipients to gain a sense of their effectiveness, hence to enhance their feeling of competence [Suh et al., 2015].

History. This functionality presents the user with a representation of the user activity over a period of time. Whereas Activity Feedback provided statistics, scores, or other information on the person's performance, History is simply a crude chronological

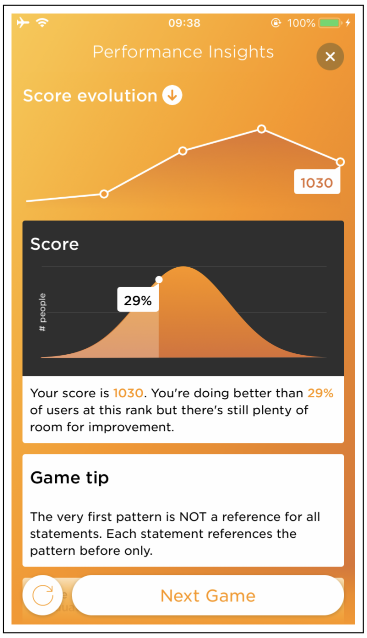

Figure 1.20
History: graph visualization of weekly activity (top of the screen) (`Peak` [Brainbow Ltd., 2019]).

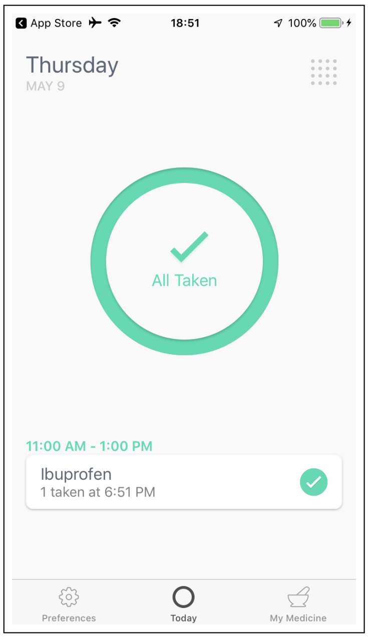

Figure 1.21
Log/Self-Monitoring: round button to report medication intake (`Round` [Circadians Design Inc., 2019]).

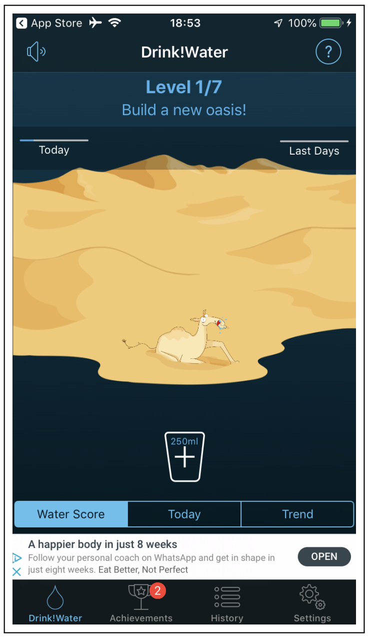

Figure 1.22
Log/Self-Monitoring: page to report water intake (`Drink Water-Health Reminder` [murbit GmbH, 2019]).

representation of what occurred in the various sessions during which the activity was performed (e.g., `Peak` [Brainbow Ltd., 2019] shows a line graph with the score progress for each time an activity was completed). See Figure 1.20 for an example.

History enables users to reflect on the temporal component of the activity of their behavior-change goal. By looking at what occurred on specific days, users are able to relate success or failures to meet specific goals to contextual factors that might have played a role in influencing their activity (e.g., peaks of stress, work deadlines, menstrual cycles). This information helps the users know themselves and how their personal circumstances might influence the behavior their aim to change, hence it supports the basic need of competence. If the temporal representation of the activity demonstrates progress or continuous maintenance of the target activity, this increases the user's sense of effectiveness, thus supporting intrinsic motivation [Grouzet et al., 2004]. Whereas, if the chronological sequence of activity shows the protracted inability of the user to reach the goal of the target activity, this might decrease the user's intrinsic motivation [Carpentier and Mageau, 2013].

Log/Self-Monitoring. This functionality provides the user with the possibility of recording the accomplishment of a goal or the completion of a task related to the specific activity (e.g., drinking a glass of water, taking a medicine, beating the performance of a previous run). Examples of this functionality can be seen in `Drink Water` [murbit GmbH, 2019] and `Round Health` [Circadians Design Inc., 2019]. See Figures 1.21 and 1.22 for examples.

Whenever the users log an activity, this represents a confirmation that they maintained the activity, consequently it enhances their feeling of *competence*. Ryan et al. found self-reporting the achievement of tasks positively associated with an increased intrinsic motivation towards the target activity [Ryan and Deci, 2000a]. At the same time, the simple act of opening the app to input data about a completed session represents an expression of volition that supports the BPN of *autonomy*. By feeding data to the app, users also express their interest in keeping up with the activity and reinforce their willingness to modify their behavior.

Rewards. We identified two forms of rewards: *tangible* and *non-tangible*. Concerning the former, some of the surveyed apps provide points to the users that can be exchanged for vouchers, products, or cash payments. In `GreenApes` [srl, 2019] the user receives "BankoNuts" that are transformed into coupons for obtaining discounts on products. Similarly, `Changers` [Blacksquared GmbH, 2019] assigns "ReCoins" to obtain vouchers for acquiring various goods. See Figures 1.23 and 1.25 for visual examples of this feature. Concerning non-tangible rewards, these might take the form of virtual points that can be exchanged for digital goods, experience points, or badges. In `My Diet Coach` [Levi, 2019], the user gets reward points to dress their avatar. See Figure 1.24 for an example of non-tangible rewards.

According to SDT, rewards can have a detrimental effect on intrinsic motivation, particularly when these are seen as the only reason to engage with the target activity [Ryan and Deci, 2017, p. 128]. Externally administered rewards can be perceived coercive and controlling, hence hinder the basic need of autonomy. Researchers demonstrate that participants who received money for solving puzzles (i.e., task-contingent reward) showed a decrease in their subsequent intrinsic motivation (measured as a free-choice persistence of the target behavior) [Deci, 1971]. In a later study, Deci and Ryan argue that offering an extrinsic reward (e.g., money) for an activity individuals were already interested in performing can prompt them to experience an *external perceived locus of causality* in their behaviour, hence producing the feeling of being controlled [Deci and Ryan, 1980, Deci and Ryan, 1985]. Non-tangible rewards, however, when connected to experience gained while performing the target activity can support the BPN of competence hence yield positive benefits for the intrinsic motivation

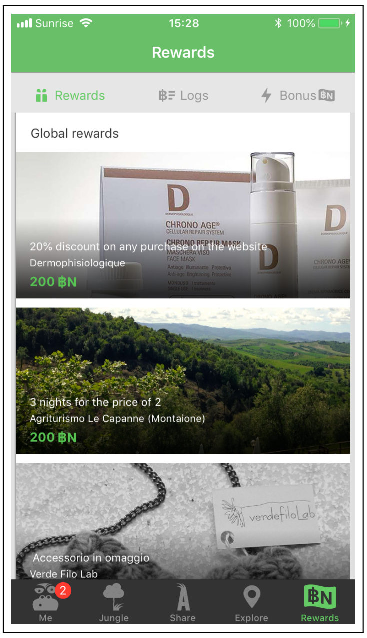

Figure 1.23
Rewards: different physical rewards the user can choose from (`GreenApes` [srl, 2019]).

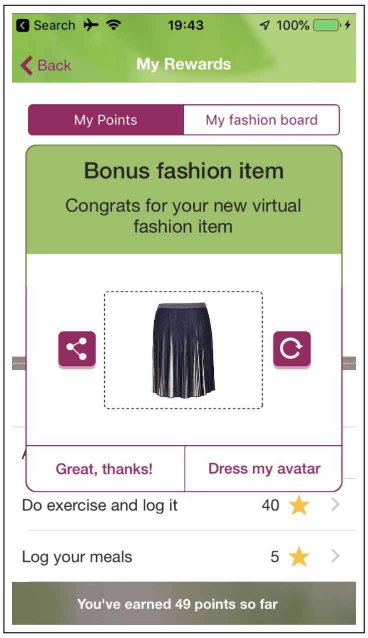

Figure 1.24
Rewards: list of virtual rewards the user can choose from (`MyDietCoach` [Anat Levi, 2019]).

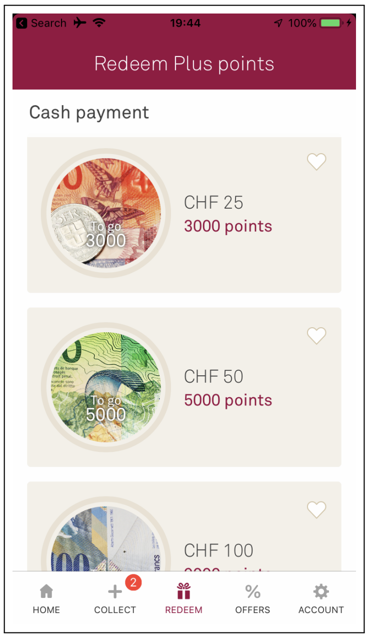

Figure 1.25
Rewards: monetary rewards (`Helsana+` [Helsana Insurance Ltd, 2019]).

of the participants [Ryan and Deci, 2000b]. Rewards such as badges, or unlocked achievements, fosters positive emotions towards the target activity [Deterding, 2012].

1.5.3 Relatedness-Supportive Features

"Relatedness refers to both experiencing others as responsive and sensitive and being able to be responsive and sensitive to them–that is, feeling connected and involved with others and having a sense of belonging" [Ryan and Deci, 2017, p. 86].

Performance Sharing. This feature enables the user to share their achievements with their peers (e.g., through social networks, e-mail, instant messaging or text messages). Typically, the user shares the score of the recently completed task or challenge. When sharing happens on social networks, then a scorecard is published on the news feed of the social network. Other instances of this feature might include the maps of the trajectory the users followed during their activity, the distance walked/ran, or the type of exercise performed (e.g., `Runtastic`, `Strava` [Strava Inc., 2019], `Youper` [Youper Inc., 2019]). See Figures 1.26,1.27, and 1.28 for visual examples of this feature.

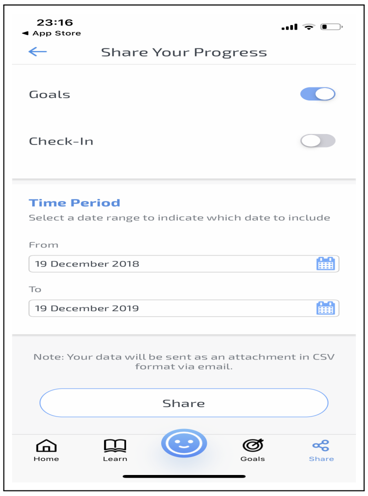

Figure 1.26
Performance Sharing (`MindShift` [Anxiety Canada Association, 2019]).

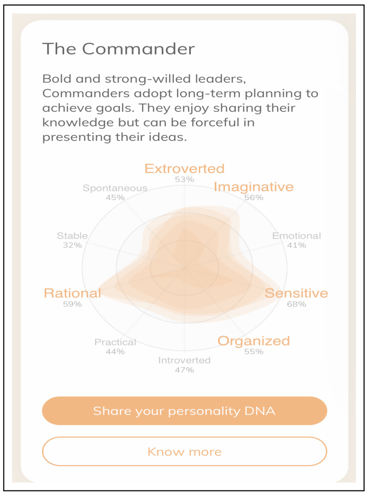

Figure 1.27
Performance Sharing (`Youper` [Youper Inc., 2019]).

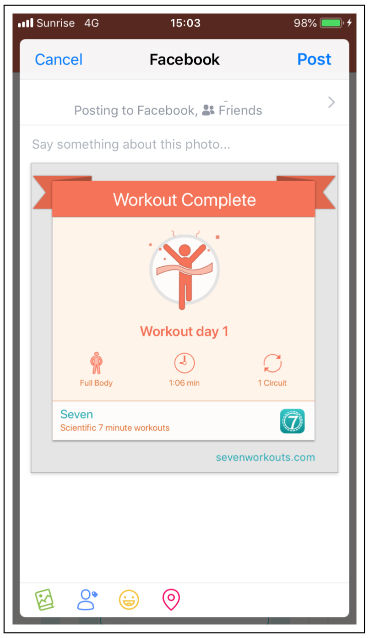

Figure 1.28
Performance Sharing: posting progress on Facebook (`7 Minute Workout` [Bytesize Systems Pty Ltd, 2019]).

By showing their progress to their close contacts, people can receive *acknowledgment* and *support*. Also, through this mechanism people can appreciate whether they matter to others and see the *impact* they have on them [Baumeister and Leary, 1995]. SDT states that by adopting attitudes and acting in manners that are endorsed by peers or significant others, individuals can feel a greater relatedness and sense of belonging that drives self-determined motivation [Ryan and Deci, 2017, p.202]. One of the aspects designers should consider when providing this feature is that although it enables users to receive supportive messages, it also opens the door to possible critics. Receiving negative responses –or even a lack of responses– from peers might lead users to experience opposite effects, specifically feeling that they might not be appreciated and cared for. In turn, this might lead to a detrimental effect on their motivation to perform the activity.

Peer Comparison. This functionality is typically implemented with a list of people who are performing the same activity. The list is ordered using quantitative scores and might include user names, actual user pictures, or avatars (e.g., `Freeletics` [Freeletics GmbH, 2019]). In game-related studies this feature is often referred to as *leader board*. In other instances of this functionality, the comparison might be enabled by other

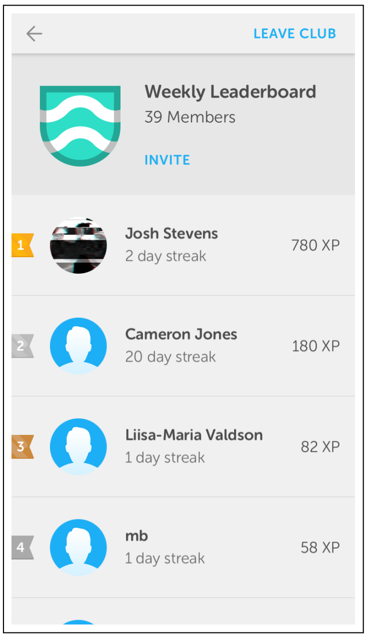

Figure 1.29
Peer Comparison: leader board (`Duolingo` [Duolingo Inc., 2019]), a weekly basis comparison displaying various information

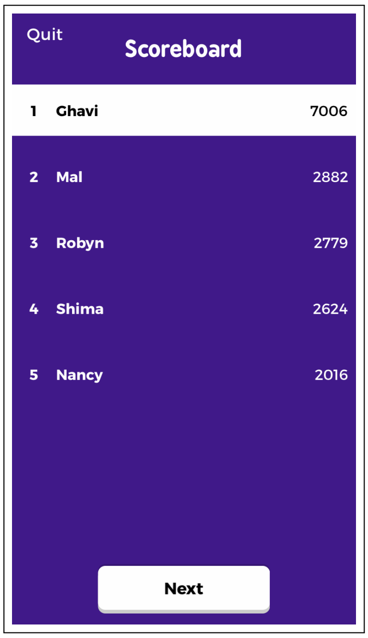

Figure 1.30
Peer Comparison: leader board (`Kahoot!` [Kahoot AS, 2019]). Shows a comparison after completing an activity.

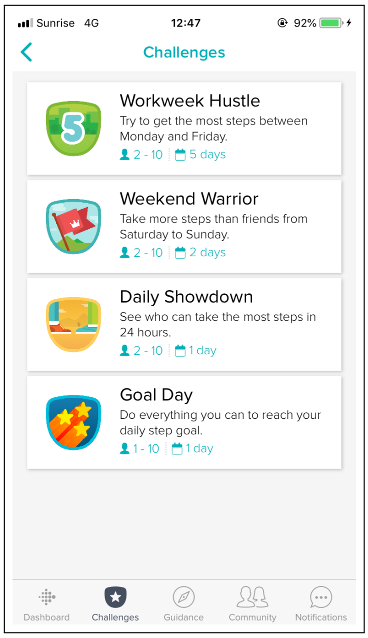

Figure 1.31
Challenge Peer: different types of challenges against other peers (`fitbit` [Fitbit Inc., 2019]).

visualizations (e.g., a Gaussian curve, like for Peak app). See Figure 1.29, 1.30 and 1.20 for an example of this feature.

Through a comparison with their peer, users can assess the impact of their actions on others and feel more effective [Ferguson and Olson, 2014]. Therefore, this feature supports the BPN of *relatedness*. Additionally, 'Peer Comparison' contributes to the SDT basic need of *competence* because, through this feature, individuals can assess their level of efficacy and mastery toward the specific activity [Ryan and Deci, 2017, p.97]. It is important to notice that this feature might also have negative effects on the motivation of the users: losing a direct competition might lead the 'loser' to experience a decrease of intrinsic motivation (i.e., turn towards amotivation) [McAuley et al., 1989]. This is particularly true when the only goal of the competition is seen as winning against the opponent [Vansteenkiste et al., 2004]. The negative effect of losing can be moderated by setting optimal challenges [Deci, 1975], and by providing positive competence feedback [Vansteenkiste et al., 2004].

Challenge Peer. Several behavior change apps we reviewed enabled users to directly challenge other users towards a given goal. These competitions could be limited to one peer or towards a group of people; they can be private (with friends and relatives)

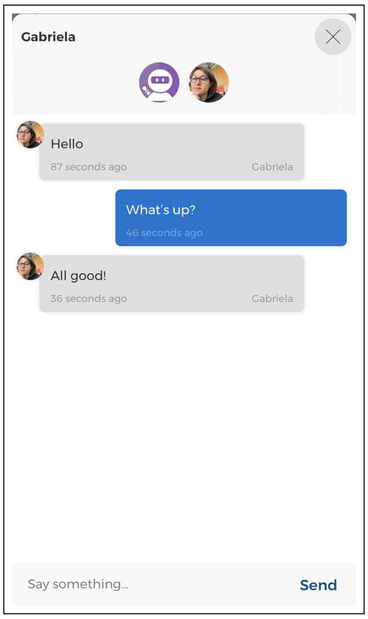

Figure 1.32
Messaging (`Goalify!` [GmbH, 2019]).

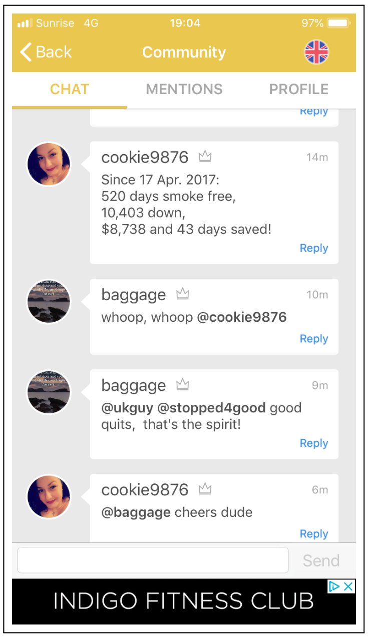

Figure 1.33
Messaging (`QuitNow!` [Fewlaps S.L., 2019]).

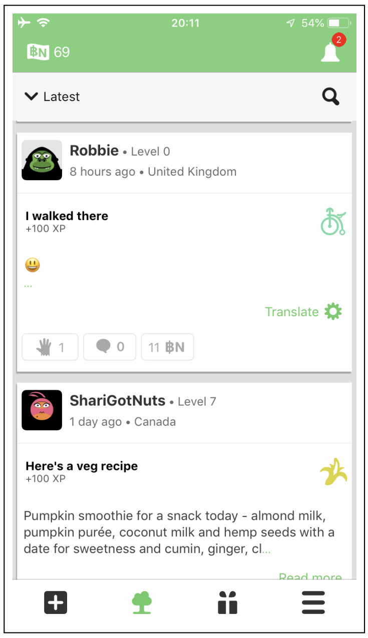

Figure 1.34
Messaging: feed (`Green Apes` [srl, 2019])

or public (with other random users) (e.g., `JouleBug` [Joulebug Inc., 2019]). See Figure 1.31 for a visual example of this feature.

Competitions against other players provide users the ability not only to compare the final outcome of the performance (i.e., the score) with other players, as per the previous category, but also to relate, test strategies, and match executions with those of the opponents. Through the interplay of seeing and being seen, this feature supports the BPN of relatedness by supporting interpersonal connection, recognition, and trust between the players [Ryan and Deci, 2017, p. 87]. Challenging peers might have detrimental effects on intrinsic motivation, especially if the challenge is imposed on the users without letting them choose when and with whom to compete [Standage and Ryan, 2012]. Also, users might feel controlled if they perceive winning as the only objective of performing the target activity.

Messaging. This feature enables the user to exchange text messages with other users who are using the same app. Some apps enable one-to-one communications (e.g., `Goalify` [GmbH, 2019]), whereas others support a group chat (e.g., `QuitNow!` [Fewlaps S.L., 2019]). In other instances of this feature, the communication functionality is afforded through a feed interface where messages are represented as cards that can be scrolled, and to which other users can reply and provide responses. These cards are

BPN	no. apps	% coverage	% aggregated
A	54	26	
C	8	3.8	30.8
R	2	1	
AC	74	35.6	
AR	15	7.1	43.7
CR	2	1	
ACR	53	25.5	25.5
Total	**208**	**100**	**100**

Table 1.4
Frequency of apps implementing behavior change features per BPN identified in the reviewed sample

used to share ideas to help other members of the community improve their skills (e.g., `GreenApes` [srl, 2019]). Typically feeds enable one-click responses (i.e., 'thumbs-up' or down). See Figure 1.33 and 1.34 for a visual example.

Messaging enable users to connect with other users who live similar experiences and face the same challenges. Through this feature, they can exchange experiences, provide and receive support to others, and experience a sense of belonging [Baumeister and Leary, 1995, Ryan, 1993, Deci and Ryan, 2000]. With respect to motivation, prior research finds that when people feel that their relatedness need is satisfied, they tend to be autonomously motivated and they can maintain the specific activity over time [Edmunds et al., 2006]. Researchers also found that the opportunity to interact with others is one of the main driver people have when playing causal games [Ferguson and Olson, 2014].

1.5.4 Coverage of the BPNs in the Sample

The two most popular behavior change features that caters to the BPN of Autonomy are Reminders (71.8% or 149 apps) and Goal Setting (55.9% or 116 apps). If we consider the BPN of Competence, the most popular feature is Activity Feedback (40.9% or 85 apps). Finally, for Relatedness, the most popular feature is Performance Sharing (18.2% or 37 apps). These results are reported in detail in Table 1.5, see also Figure 1.35 for a visual representation of the behavior change features coverage in the Sample. From the total of 208 apps included in the analysis, only 25.5% (or 53) implemented at least one feature that supports all three BPNs. About 44% (or 91) of the reviewed apps provided

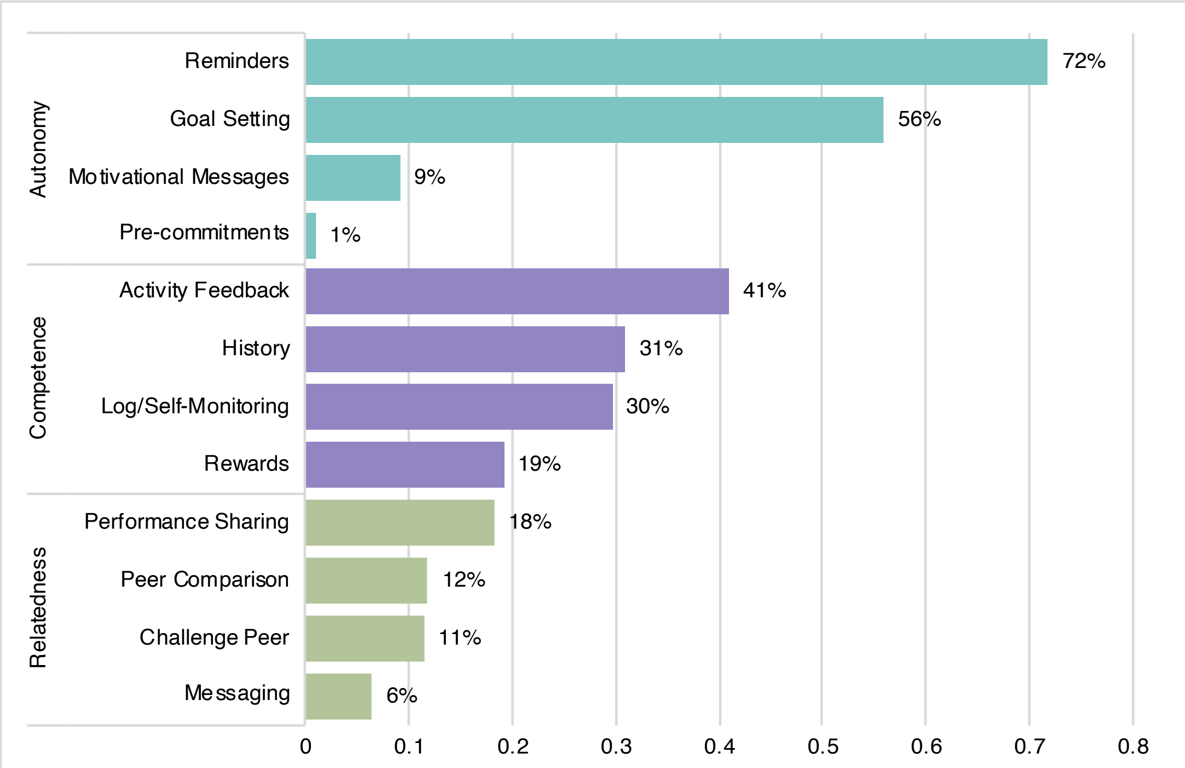

Figure 1.35
Percentage of apps in the sample that afforded a given behavior change feature. The features are listed by decreasing frequency in each BPN (they follow the same order reported in Table 1.5.

support for only two of the basic needs. It is interesting to notice that the most popular combination we observed was 'AC' covering 35.6% (or 74) of reviewed apps, while the least popular combination was 'CR' with only 1% (or 2) of the reviewed apps. Finally, about 31% (or 64) of the reviewed apps provided support for only one of the basic needs with autonomy being the most popular basic need for which features were afforded in behavior change apps (i.e., 26%, or 54, of the sample). Table 1.7 presents a summary of the frequencies of apps implementing the behavior change features described in this section.

1.6 Discussion

In this study, we map the different techniques adopted by current behavior change apps and relate them to the Self-Determination Theory of human motivation. We

Feature	A	C	R	Coverage %	Example of App	Figure no.
Reminders	•			71.8	30 Day Fitness Challenge	1.6
Goal Setting	•			55.9	Feastr	1.9
Motivational Messages	•			9.1	Kwit	1.13
Pre-commitments	•		○	0.9	Stick	1.15
Activity Feedback		•		40.9	Babbel	1.17
History		•		30.9	Peak	1.20
Log/Self-Monitoring	○	•		29.6	Round	1.21
Rewards		•		19.1	GreenApes	1.23
Performance Sharing			•	18.2	MindShift	1.26
Peer Comparison		○	•	11.8	Duolingo	1.29
Challenge Peer			•	11.4	Fitbit	1.31
Messaging			•	6.4	Goalify	1.32

Table 1.5
Behavior change features and % of coverage in the sample, including examples of apps containing the given feature. A: autonomy, C: competence, R: relatedness, •: primary classification, ○: secondary

Feature	A	C	R	coverage %
Reminders	•			71.8
Goal Setting	•			55.9
Motivational Messages	•			9.1
Pre-commitments	•		○	0.9
Activity Feedback		•		40.9
History		•		30.9
Log/Self-Monitoring	○	•		29.6
Rewards		•		19.1
Performance Sharing			•	18.2
Peer Comparison		○	•	11.8
Challenge Peer			•	11.4
Messaging			•	6.4

Table 1.6
Behavior change features and % of coverage in the sample. A: autonomy, C: competence, R: relatedness, •: primary classification, ○: secondary

BPN	no. apps	% coverage	% aggregated
A	54	26	
C	8	3.8	30.8
R	2	1	
AC	74	35.6	
AR	15	7.1	43.7
CR	2	1	
ACR	53	25.5	25.5
Total	**208**	**100**	**100**

Table 1.7
Frequency of apps implementing behavior change features per BPN identified in the reviewed sample

reviewed 208 apps designed purposely to support behavior change and found that the most popular feature is Reminders. This resonates with previous reviews of habit formation apps [Stawarz et al., 2015]. This is likely due to the fact that designers overuse this feature to try to compel individuals to use their app. When mapping design features to SDT, we found that only one fourth of the sample provided users support for the three basic needs. For the apps that do cover all the BPNs, a common design pattern consists of letting users select the target activity (i.e., Goal Setting), then providing statistics after the activity is performed (i.e., Activity Feedback) and enabling users to share their performance through social media (i.e., Performance Sharing). Conversely, we found that 74.5% of behavior change apps do not support all the basic psychological needs: According to the SDT, supporting the three BPNs enables the person to move towards self-determined action for the target activity.

Here, we discuss how the analysis reported in this paper can inform future design and research by pointing to: (i) rarely used design features in current behavior change apps that are underexplored in HCI research; and (ii) feature gaps identified by looking at the app design through the lens of Self-Determination Theory. Finally, we outline limitations and future work.

1.6.1 Research Opportunities Suggested by Rarely Used Design Features

The analysis of the behavior change apps reported in this experiment highlighted design features that are seldom used: pre-commitment and intergroup competition. These might appear less frequently for multiple reasons: because they have been tested and discarded in prior iterations of the design of popular apps, or because they represent unexplored design space yet to be exploited. Both possibilities would warrant HCI experiments that might reveal properties and applicability of these features.

Pre-commitments: This feature was implemented only in $\sim$1% of behavior change apps. Pre-commitment could be a powerful mechanism to counter procrastination that can arise when preferences are inconsistent over time and across contexts. One of the causes for the apparent changes in preferences over time is a change in the saliency of the costs and benefits of the activity in question [Akerlof, 1991]. Although such time-inconsistent preferences can form serious obstacles to following a planned course of action, they can be overcome. In addition to exercising willpower to resist temptation, people can constrain or pre-commit their behavior [Schelling, 1992, Thaler and Shefrin, 1981, Wertenbroch, 1998]. Binding behavior is characterized by the voluntary imposition of constraints (that are costly to overcome) on one's future choices in a strategic attempt to resist future temptations. Ariely et al. experimentally studied pre-commitment [Ariely and Wertenbroch, 2002]. The results of their study show that people are willing to self-impose meaningful (i.e., costly) deadlines to overcome procrastination and that these self-imposed deadlines are effective in improving task performance. This technique was studied in behavior change apps for regulating the use to digital devices. Kim et al. reviewed several behavior change apps to regulate use of mobile devices [Kim et al., 2019]. Similarly, the aforementioned work of Lyngs et al. reviewed apps and browser extensions of which many implement varying levels of friction if users wish to override their own past preferences [Lyngs et al., 2019]. To the best of our knowledge, pre-commitment has not been covered in other domains of behavior change from HCI research. From an SDT perspective, the feature in its most common implementation supports the BPN of autonomy. However, alternative design might involve peers (or family members) as referees on the bids, thus enabling also support to the BPN of relatedness. Social support was studied in the domain of self-regulation (cf. [Ko et al., 2016, Hiniker et al., 2016]). These studies revealed that social support helped users mitigate smartphone distractions. However, we are not aware of studies that focused on pre-commitment *and* social support. A user betting on the achievement of a given task, might feel more compelled to bring it to completion if a friend or another user of the same system will be checking on her/him (as opposed to an algorithm).

Intergroup Competition. Many of the behavior change applications that we reviewed give users the ability to compete against other users. However, we could not identify apps that enable users to cooperate towards a given goal. Furthermore, we could not identify apps that enable users to compete in groups (i.e., inter-group competition). Both competition and cooperation can affect intrinsic motivation in a number of ways. Research has demonstrated a positive effect of competition on intrinsic motivation [Epstein and Harackiewicz, 1992, Reeve and Deci, 1996, Tauer and Harackiewicz, 1999]. There are two main mechanisms in which competition affects intrinsic motivation: (1) through the competitive context established at the outset of an activity, which can affect how individuals approach a task, and (2) through performance feedback [Sansone and Harackiewicz, 1996]. Other research has revealed that if individuals focus on winning rather than the activity itself, their intrinsic motivation can decrease [Deci and Ryan, 1985, Harackiewicz et al., 1998]. Cooperation also has the potential to affect intrinsic motivation in a number of ways, because individuals can experience the benefits of being part of a team that works toward a common goal; this engenders a sense of relatedness among their teammates. Cooperation also has the potential to provide positive feedback if a team completes the goal. This can promote perceived competence and, in turn, intrinsic motivation [Deci and Ryan, 1991, Ryan and Deci, 2000b, Ryan and Deci, 2000a, Vallerand and Losier, 1999]. However, cooperation can have negative effects on motivation if they perceive the group goal as externally controlling (loss of autonomy), or if they fail to meet their goal. A safer approach –with regard to affecting intrinsic motivation– could be letting the users compete in groups (i.e., inter-group competition). Tauer et al. found that inter-group competition leads individuals to experience levels of intrinsic motivation higher than pure cooperation and pure competition [Tauer and Harackiewicz, 2004]. Therefore, it would be relevant to empirically compare cooperation and inter-group competition with individual competition in their ability to support behavior-change interventions.

1.6.2 Feature Design and Research Gaps Suggested by the SDT

By classifying and looking at the behavior change features from the SDT perspective, we also identify three areas that are currently under-explored in research focusing on these apps: (a) Design that support the individual in reaching higher level of intrinsic motivation; (b) Design that provide support for all the three basic needs as identified by the SDT; and (c) Tailoring of the interventions that resonate with the constructs of

SDT. We argue that conducting more research in these areas can lead to new powerful designs for behaviour change apps.

(a) Nurturing or Thwarting Intrinsic Motivation. Through this analysis, we observed *Reminders* and *Activity Feedback* features in almost every app we analyzed. We have also observed examples of *Motivational Messages* whose content was disconnected from the performance of the recipient of the messages. The connection provided by the taxonomy between design and theory, allows us to derive implications for the *content* and for the *deployment* of these messages. SDT research has demonstrated that providing feedback that shows progress increases intrinsic motivation [Vallerand and Reid, 1984]. However, if the feedback does not show consistent improvements on the target activity, it might discourage the user [Burgers et al., 2015]. An SDT-informed design for activity feedback would require information provided to the user to be: i. *personal* (i.e., specific to the participant); ii. *contextual* (i.e., providing task-inherent information that can help the user connect their performance of the activity with its outcomes); and iii. *goal-oriented* (i.e., providing the next challenge to push their work further by being phrased in a way that is specific to the level of the user) [Cherubini et al., 2020]. Concerning the deployment of these messages, we note that these are often sent through a channel that is already overloaded by other communications and might lead the user to experience *notification fatigue* [Pielot and Rello, 2017]. Also, if the delivery of these messages becomes repetitive and predictable, the user might experience them as *controlling* and this might be detrimental to intrinsic motivation [Ryan, 1982, Kast and Connor, 1988]. Instead, it would be more beneficial to deliver these messages opportunistically when the user performs a spontaneous activity, perhaps at a time or place where this did not occur in the past [Cherubini et al., 2020].

Another point of discussion concerns the use of *rewards*. SDT researchers conducted many experimental studies on rewards, punishments, and other extrinsic events [Deci, 1971, Deci, 1972, Deci and Ryan, 1980, Deci and Ryan, 1985]. SDT specifies that these external events might support or thwart a person's feeling of *autonomy* and *competence* and this, in turn, influences intrinsic motivation. SDT research showed that if rewards are seen as the only goal of performing the activity, these can yield detrimental effects on intrinsic motivation [Deci, 1971, Deci and Ryan, 1980, Deci and Ryan, 1985]. External rewards can be perceived as controlling or coercive, consequently harming the basic need of *autonomy*. An SDT-informed design for rewards should provide tokens that celebrate the users' renewed competences, rather than anchoring users on the extrinsic value of the prize. In this context, we can think about non-monetary rewards (e.g., badges, experience points) that can be given to the user when specific goals or sustained performance are achieved. These incentives might be perceived by the recipients as recognizing their knowledge, rather than placing a value on their behavior.

In summary, to ensure that users will have self-determined motivation for the specific activity, it is not sufficient to implement any of the behavior change features described in the taxonomy. Erroneous designs of these features might lead users to feel controlled, inapt, or not at the level of their peers, thus hurting the users' intrinsic motivation.

(b) Providing Support to the Three Basic Needs. SDT posits that it is fundamental to a person's growth, well-being, and integrity to fulfill the three basic needs [Ryan and Deci, 2017, p.98]. In the analysis of the pool of apps selected for this work, we found that only one fourth of the apps (i.e., 25.5%) include features that trigger all three of these psychological sensibilities in some form. However, we lack controlled experiments that could shed light on the effect of implementing multiple features that cater to the BPNs on the motivation of users for the specified activity. We suggest three open questions: (i) It is still unclear whether providing support for only one, or two of the basic needs can yield positive effects on a user's motivation. (ii) The majority of behavior change apps implement multiple behavior change features that provide support to the same basic need (i.e., 74.5%). However, we do not know whether implementing multiple features that support the same BPN would actually increase the overall positive effect, or be detrimental towards supporting self-determined action towards the target activity. (iii) In the analysis, we found that 43.8% of the apps provide support to two BPNs. It is not clear whether a particular combination of supports for the three basic needs would be better suited to help users with varying levels of intrinsic motivation (measured at the onset of the intervention). Longitudinal and large-scale studies that include a post-experiment observation are necessary to understand the long-term effects of the interventions (cf. [Patel et al., 2016, Cherubini et al., 2020]). These experimental designs might also account for individual differences, and record effects on ceiling performance and lapsed use.

(c) Optimal Challenge. When surveying the behavior change apps, we realized that there are very few applications that tailor the intervention to the specific characteristics of the user. Recent research focusing on serious games revealed that users respond differently to behavior change strategies [Orji et al., 2014, Orji et al., 2017b, Sundar and

Marathe, 2010]. For instance, for a given user who walks an average of 8K steps a day, walking or running 10K steps a day is a challenging but realistic goal. However, the same goal, for a person that walks or runs 4.5K steps a day, might be completely unrealistic (we referred to this concept as *optimal challenge* [Deci, 1975]). The analysis reported in this paper reveals that app designers often opt for a one-solution-fits-all approach; during the few weeks of testing, we could not identify any tailoring or personalization mechanisms. However, SDT cautions that although some users might be motivated for a specific activity by challenges they consider interesting, others might simply react the opposite way if they perceive the challenge as too difficult. More research in this area could demonstrate the effect of providing personalized challenges to users of behavior change apps.

1.6.3 Feature Design Variations

The app features presented in this Taxonomy can be designed in multiple forms; we present some examples of these designs in this research effort. Given that creativity plays a critical role in the app feature design, it is possible that some designs might support not only one but two BPNs. For instance, a goal-setting feature allowing participants to set their goals based on past performance might support not only *autonomy* but also *competence*. Support to *autonomy* by allowing the individual to choose their goal and support to *competence* by offering the choice to increase the goal based on past performance. Furthermore, the classification presented in this project is based on a specific set of apps corresponding to the specific present time. It might occur that as technology evolves, new app features designs will arise supporting multiple BPNs forcing an upgrade in the present Taxonomy.

1.6.4 Limitations and Future Work

Our work has some limitations. As the Apple App Store and App Annie do not provide information about the number of users, we focused on functionality, leaving considerations on the number of installs or the content of user reviews to future work. This is similar to the approach taken by [Lyngs et al., 2019, Stawarz et al., 2014, Stawarz et al., 2015]. The study of behavior change apps reported in this paper was limited to the iOS App Store and to free apps. Behavior change features that could have been provided after in-app purchases might have not been analyzed. In our sample this might have been the case for 30 apps that we excluded from the initial sample. Future work should extend the analysis also to these paid features. The analysis of the feature was also limited in time. The apps were tested for several hours across multiple weeks, which provided us with a longitudinal exposure to the features. We created accounts and tested it with real interactions (e.g., drinking water when the app told us so). The features that tailored interventions during this time frame were analyzed. Our analysis does not take into account tailoring strategies on long term interventions. Also, the methodology employed in this research is limited to expert evaluation of the features. Employing another approach such as crowdsourcing would have provided us with a more significant number of apps to be classified. However, training crowdsourcers to understand SDT constructs fully requires a considerable amount of time out of this project research scope. Additionally, in this study we did not look at hardware counterparts for behavior change apps (i.e., wearables). Looking at our sample, only two apps in our sample have a hardware counterpart but they can be used without it and recent research shows that wearable trackers have high attrition rate [Lazar et al., 2015]. Future work should also look at other app stores. As highlighted by Lyngs et al. [Lyngs et al., 2019], iOS apps tend to have fewer features than their Android counterparts (especially for pre-commitment and tracking), because iOS provides fewer permissions to developers. Therefore, reviewing only apps for iOS might give a limited picture of what features have been explored in behaviour change apps. Furthermore, future work should cover user reviews (cf. [Cowan et al., 2012]).

Finally, SDT as a theoretical framework is not exempt from criticisms. As we have reviewed in Sec. 1.3.3, SDT focuses on self-determined behaviors, which are chosen consciously. However, recent research also highlighted the importance of non-conscious mechanisms to form habits and modify behaviors [Verplanken and Wood, 2006, Pinder et al., 2018] but also theories that focus on rational deliberative processes, as SDT, are typically insufficient to explain the intention-behavior gap in the presence of strong habits [Sheeran et al., 2017]. Other streams of research focused on the BPNs and suggested that other basic needs also play an important role on human behavior [Sheldon et al., 2001, Martela and Ryan, 2016]. Finally, other scholars looked at rewards and reached different conclusions from those suggested by SDT, however the topic is still debated [Deci et al., 1999, Cherubini et al., 2020].

1.7 Conclusion

The challenge of designing effective behaviour-change interventions is important to address. In this paper, we contribute to this effort by providing a functionality analysis, according to the Self-Determination Theory, of current apps for behaviour change. This survey reveals gaps for future studies that can further develop our understanding of the domain and intervention design. We hope that this research informs a future where technology will be used to reinforce and enable the autonomy of individuals, rather than necessitating dependencies.

1.8 Acknowledgements

We would like to thank Amon Rapp and Florian Brühlmann for their valuable comments that contributed to the development of this work. We also thank Tolulope Omuku and Khaled Lala (master students at UNIL) for their help analysing the data. Finally, we would like to thank Holly Cogliati-Bauereis, James Tyler and Mercedes Alpízar for their precious comments. Finally, we would like to thank the anonymous reviewers for their precious comments.

1.9 Conclusion to the Chapter

In this chapter, we learned that SDT could support the design of behavior change app features. Also, only a quarter of the apps we reviewed support the BPNs suggesting that behavior change apps can benefit from exploring multiple BPNs and implementing features that are rarely used, such as coopetition and pre-commitments. We acknowledge that having a taxonomy of app features is insufficient to inform the design of behavior change apps. Therefore, it is necessary to test different implementations of these features to determine if the individuals perceive these as supporting the BPNs. Thus, in the subsequent chapter, we create a physical activity app that implements one particular combination of features described in this taxonomy. We allowed individuals to interact with the app, and then we collected their app feature perceptions to understand if they perceived their BPNs were supported by the features.

First Chapter References

[A. L. Amazing Apps Ltd., 2019] A. L. Amazing Apps Ltd. (2019). Yoga-Go: Weight loss workouts. Last accessed August 22, 2019.

[Abishkking Ltd., 2019] Abishkking Ltd. (2019). 30 Day Fitness Challenge Pro: 7 minute home workout exercise. Last accessed August 22, 2019.

[Abroms et al., 2011] Abroms, L. C., Padmanabhan, N., Thaweethai, L., and Phillips, T. (2011). iPhone Apps for Smoking Cessation: A Content Analysis. *American Journal of Preventive Medicine*, 40(3):279–285.

[Ajzen, 1985] Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In *Action control*, pages 11–39. Springer, Salmon Tower Building New York City, USA.

[Akerlof, 1991] Akerlof, G. (1991). Procrastination and Obedience. *American Economic Review*, 81(2):1–19.

[Alharthi et al., 2018] Alharthi, S. A., Alsaedi, O., Toups, Z. O., Tanenbaum, J., and Hammer, J. (2018). Playing to Wait: A Taxonomy of Idle Games. In *Proceedings of CHI'18*, pages 1–15, New York, USA. ACM Press.

[Anat Levi, 2019] Anat Levi (2019). My Diet Coach - Weight Loss. Last accessed August 22, 2019.

[Anxiety Canada Association, 2019] Anxiety Canada Association (2019). MindShift™ CBT. Last accessed August 22, 2019.

[App Annie Inc., 2019] App Annie Inc. (2019). AppAnnie: The Mobile Performance Standard. Last accessed August 22, 2019.

[AppStudio Australia Pty Ltd, 2019] AppStudio Australia Pty Ltd (2019). Brainbuddy: Addicted to Porn? Reboot your Brain. Last accessed August 22, 2019.

[Ariely and Wertenbroch, 2002] Ariely, D. and Wertenbroch, K. (2002). Procrastination, deadlines, and performance: Self control by precommitment. *Psychological Science*, 13(3):219–224.

[ARISE Ltd., 2019] ARISE Ltd. (2019). Arise: Weight Loss Made Simple. Last accessed August 22, 2019.

[Asana Rebel GmbH, 2019] Asana Rebel GmbH (2019). Asana Rebel - Design Your Perfect Day. Last accessed August 22, 2019.

[Bandura, 1986] Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall, Inc., Englewood Cliffs, NJ, US.

[Bargh, 2007] Bargh, J. A. (2007). Social psychological approaches to consciousness. In Thompson, E., Moscovitch, M., and Zelazo, P. D., editors, *The Cambridge Handbook of Consciousness*, Cambridge Handbooks in Psychology, pages 555–569. Cambridge University Press, Cambridge.

[Baumeister and Leary, 1995] Baumeister, R. F. and Leary, M. R. (1995). The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychological bulletin*, 117(3):497–529.

[Betterment, 2019] Betterment, W. (2019). Better Habits: The smartest mobile app for forming habits that last. Last accessed August 22, 2019.

[Birk et al., 2016] Birk, M. V., Atkins, C., Bowey, J. T., and Mandryk, R. L. (2016). Fostering Intrinsic Motivation through Avatar Identification in Digital Games. In *Proceedings of CHI'16*, pages 2982–2995, New York, USA. ACM Press.

[B.J. Fogg, 2019] B.J. Fogg (2019). Tiny Habits. Last accessed August 22, 2019.

[Blacksquared GmbH, 2019] Blacksquared GmbH (2019). Changers: Healthcare platform to become healthier, happier and greener. Last accessed August 22, 2019.

[Brainbow Ltd., 2019] Brainbow Ltd. (2019). Peak: challenge your brain. Last accessed August 22, 2019.

[Braun and Clarke, 2006] Braun, V. and Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101.

[Brynjarsdottir et al., 2012] Brynjarsdottir, H., Håkansson, M., Pierce, J., Baumer, E., DiSalvo, C., and Sengers, P. (2012). Sustainably unpersuaded: how persuasion narrows our vision of sustainability. In *Proceedings of CHI'12*, page 947, New York, USA. ACM Press.

[Burgers et al., 2015] Burgers, C., Eden, A., van Engelenburg, M. D., and Buningh, S. (2015). How feedback boosts motivation and play in a brain-training game. *Computers in Human Behavior*, 48:94–103.

[Busuu Ltd, 2019] Busuu Ltd (2019). Busuu: Learn Languages Online. Last accessed August 22, 2019.

[Bytesize Systems Pty Ltd, 2019] Bytesize Systems Pty Ltd (2019). 7 Minute Workout: Daily bodyweight exercises. Last accessed August 22, 2019.

[Caraban et al., 2019] Caraban, A., Karapanos, E., Gonçalves, D., and Campos, P. (2019). 23 Ways to Nudge: A review of technology-mediated nudging in human-computer interaction. In *Conference on Human Factors in Computing Systems - Proceedings*, pages 1–15. Association for Computing Machinery.

[Carpentier and Mageau, 2013] Carpentier, J. and Mageau, G. A. (2013). When change-oriented feedback enhances motivation, well-being and performance: A look at autonomy-supportive feedback in sport. *Psychology of Sport and Exercise*, 14(3):423–435.

[Chefkoch GmbH, 2019] Chefkoch GmbH (2019). Chefkoch: SmartList. Last accessed August 22, 2019.

[Cherubini et al., 2020] Cherubini, M., Villalobos-Zuñiga, G., Boldi, M.-O., and Bonazzi, R. (2020). The Unexpected Downside of Paying or Sending Messages to People to Make Them Walk. *ACM Transactions on Computer-Human Interaction*, 27(2):1–44.

[Chiriac et al., 2011] Chiriac, N., Hölttä-Otto, K., Lysy, D., and Suh, E. S. (2011). Three approaches to complex system decomposition. In *Proceedings of the 13th International DSM Conference*, pages 3–17, Cambridge, MA, USA. MIT Press.

[Circadians Design Inc., 2019] Circadians Design Inc. (2019). Round Health: The best way to remember and record your medications and vitamins. Last accessed August 22, 2019.

[Clue by Biowink GmbH, 2019] Clue by Biowink GmbH (2019). Clue app: birth control pill reminders. Last accessed August 22, 2019.

[Consolvo et al., 2008] Consolvo, S., Libby, R., Smith, I., Landay, J. A., McDonald, D. W., Toscos, T., Chen, M. Y., Froehlich, J., Harrison, B., Klasnja, P., LaMarca, A., and LeGrand, L. (2008). Activity sensing in the wild: a field trial of ubifit garden. In *Proceeding of CHI'08*, pages 1797 – 1806, New York, USA. ACM Press.

[Coughlin et al., 2015] Coughlin, S. S., Whitehead, M., Sheats, J. Q., Mastromonico, J., Hardy, D., and Smith, S. A. (2015). Smartphone Applications for Promoting Healthy Diet and Nutrition: A Literature Review. *Jacobs journal of food and nutrition*, 2(3):021.

[Cowan et al., 2012] Cowan, L. T., Van Wagenen, S. A., Brown, B. A., Hedin, R. J., Seino-Stephan, Y., Hall, P. C., and West, J. H. (2012). Apps of Steel: Are Exercise Apps Providing Consumers With Realistic Expectations?: A Content Analysis of Exercise Apps for Presence of Behavior Change Theory. *Health Education & Behavior*, 40(2):133–139.

[d bel Ltd., 2019] d bel Ltd. (2019). Quit Smoking: We are your motivation! Last accessed August 22, 2019.

[Davidson, 1998] Davidson, R. (1998). The transtheoretical model: A critical overview. In *Treating addictive behaviors, 2nd ed.*, Applied clinical psychology., pages 25–38. Plenum Press, New York, NY, US.

[Davis et al., 2015] Davis, R., Campbell, R., Hildon, Z., Hobbs, L., and Michie, S. (2015). Theories of behaviour and behaviour change across the social and behavioural sciences: a scoping review. *Health Psychology Review*, 9(3):323–344.

[de Oliveira et al., 2010] de Oliveira, R., Cherubini, M., and Oliver, N. (2010). MoviPill: improving medication compliance for elders using a mobile persuasive social game. In *Proceedings of Ubicomp'10*, page 251, New York, USA. ACM Press.

[DeCharms, 1968] DeCharms, R. (1968). *Personal causation: the internal affective determinants of behavior*. Academic Press, Cambridge, Massachusetts, US.

[Deci and Ryan, 1985] Deci, E. and Ryan, R. M. (1985). *Intrinsic Motivation and Self-Determination in Human Behavior*. Springer US, New York, USA.

[Deci, 1971] Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of personality and Social Psychology*, 18(1):105.

[Deci, 1972] Deci, E. L. (1972). The effects of contingent and noncontingent rewards and controls on intrinsic motivation. *Organizational behavior and human performance*, 8(2):217–229.

[Deci, 1975] Deci, E. L. (1975). *Intrinsic Motivation*. Bibliographie pp.295. Plenum Publishing Company Limited, New York, USA.

[Deci and Ryan, 1980] Deci, E. L. and Ryan, R. M. (1980). The empirical exploration of intrinsic motivational processes1. In *Advances in experimental social psychology*, volume 13, pages 39–80. Elsevier, Atlanta, USA.

[Deci and Ryan, 1991] Deci, E. L. and Ryan, R. M. (1991). A motivational approach to self: Integration in personality. In *Nebraska Symposium on Motivation, 1990: Perspectives on motivation.*, Current theory and research in motivation, Vol. 38., pages 237–288. University of Nebraska Press, Lincoln, NE, US.

[Deci and Ryan, 2000] Deci, E. L. and Ryan, R. M. (2000). The “what” and “why” of goal pursuits: Human needs and the self-determination of behavior. *Psychological inquiry*, 11(4):227–268.

[Deci and Ryan, 2008] Deci, E. L. and Ryan, R. M. (2008). Self-determination theory: A macrotheory of human motivation, development, and health. *Canadian Psychology/Psychologie canadienne*, 49(3):182–185.

[Deci et al., 1999] Deci, E. L., Ryan, R. M., and Koestner, R. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125(6):627–668.

[Deci et al., 1996] Deci, E. L., Ryan, R. M., and Williams, G. C. (1996). Need satisfaction and the self-regulation of learning. *Learning and individual differences*, 8(3):165–183.

[Deterding, 2012] Deterding, S. (2012). Gamification: designing for motivation. *Interactions*, 19(4):14.

[Deterding, 2016] Deterding, S. (2016). Contextual Autonomy Support in Video Game Play: A Grounded Theory. In *Proceedings of CHI’16*, pages 3931–3943, New York, USA. ACM Press.

[DiClemente, 2003] DiClemente, C. C. (2003). *Addiction and change: How addictions develop and addicted people recover.* Guilford substance abuse series. The Guilford Press, New York, USA.

[Duolingo Inc., 2019] Duolingo Inc. (2019). Duolingo: Learn a language for free. Last accessed August 22, 2019.

[Edmunds et al., 2006] Edmunds, J., Ntoumanis, N., and Duda, J. L. (2006). A test of self-determination theory in the exercise domain. *Journal of Applied Social Psychology*, 36(9):2240–2265.

[Edwards et al., 2016] Edwards, E. A., Lumsden, J., Rivas, C., Steed, L., Edwards, L. A., Thiyagarajan, A., Sohanpal, R., Caton, H., Griffiths, C. J., Munafò, M. R., Taylor, S., and Walton, R. T. (2016). Gamification for health promotion: systematic review of behaviour change techniques in smartphone apps. *BMJ Open*, 6(10):6–0.

[Epstein and Harackiewicz, 1992] Epstein, J. A. and Harackiewicz, J. M. (1992). Winning is Not Enough: The Effects of Competition and Achievement Orientation on Intrinsic Interest. *Personality and Social Psychology Bulletin*, 18(2):128–138.

[Erskine et al., 2010] Erskine, J. A., Georgiou, G. J., and Kvavilashvili, L. (2010). I suppress, therefore I smoke: Effects of thought suppression on smoking behavior. *Psychological Science*, 21(9):1225–1230.

[Fabulous Ltd., 2019] Fabulous Ltd. (2019). The Fabulous: Embark on a journey to reset your habits. Last accessed August 22, 2019.

[Fatsecret Ltd., 2019] Fatsecret Ltd. (2019). fatsecret: calorie counter app. Last accessed August 22, 2019.

[Feastr GmbH, 2019] Feastr GmbH (2019). Feastr: Ealthy eating has never been so easy. Last accessed August 22, 2019.

[Ferguson and Olson, 2014] Ferguson, C. J. and Olson, C. K. (2014). Video Game Violence Use Among "Vulnerable" Populations: The Impact of Violent Games on Delinquency and Bullying Among Children with Clinically Elevated Depression or Attention Deficit Symptoms. *Journal of Youth and Adolescence*, 43(1):127–136.

[Ferrara, 2013] Ferrara, J. (2013). Games for persuasion: Argumentation, procedurality, and the lie of gamification. *Games and Culture*, 8(4):289–304.

[Fewlaps S.L., 2019] Fewlaps S.L. (2019). QuitNow!: Quitting smoking made easy. Last accessed August 22, 2019.

[Fitbit Inc., 2019] Fitbit Inc. (2019). FitBit App: the fitness app for everyone. Last accessed August 22, 2019.

[Fox and Duggan, 2012] Fox, S. and Duggan, M. (2012). Mobile Health 2012. Technical report, Pew Research Center.

[Freeletics GmbH, 2019] Freeletics GmbH (2019). Freeletics: Train with the best digital coach. Last accessed August 22, 2019.

[Friederichs et al., 2015] Friederichs, S. A., Oenema, A., Bolman, C., and Lechner, L. (2015). Long term effects of self-determination theory and motivational interviewing in a web-based physical activity intervention: randomized controlled trial. *International Journal of Behavioral Nutrition and Physical Activity*, 12(1):101.

[Fritz et al., 2014] Fritz, T., Huang, E. M., Murphy, G. C., and Zimmermann, T. (2014). Persuasive technology in the real world: a study of long-term use of activity sensing devices for fitness. In *Proceedings of CHI'14*, pages 487–496, New York, USA. ACM Press.

[Geuens et al., 2016] Geuens, J., Swinnen, T. W., Westhovens, R., De Vlam, K., Geurts, L., and Abeele, V. V. (2016). A review of persuasive principles in mobile apps for chronic arthritis patients: opportunities for improvement. *JMIR mHealth and uHealth*, 4(4):1–15.

[Gimpel et al., 2015] Gimpel, H., Regal, C., and Schmidt, M. (2015). myStress: Unobtrusive Smartphone-Based Stress Detection. In *ECIS 2015 Research-in-Progress Papers*, pages 1–12, Atlanta, USA. Association for Information Systems.

[Glanz et al., 2008] Glanz, K., Rimer, B. K., and Viswanath, K. (2008). What is Theory? In *Health behavior and health education: Theory, research, and practice*, page 552. Jossey-Bass, San Francisco, CA, US.

[GmbH, 2019] GmbH, O. (2019). Goalify helps you to stay on top of your goals and habits. Last accessed August 22, 2019.

[Graham et al., 2006] Graham, C., Benda, P., Howard, S., Balmford, J., Bishop, N., and Borland, R. (2006). "heh - keeps me off the smokes...": Probing technology support for personal change. In *Proceedings of OZCHI'06*, pages 221–228, New York, NY, USA. ACM.

[Grant, 2012] Grant, A. M. (2012). An integrated model of goal-focused coaching: An evidence-based framework for teaching and practice. *International Coaching Psychology Review*, 7(2):146–165.

[Grist et al., 2017] Grist, R., Porter, J., and Stallard, P. (2017). Mental Health Mobile Apps for Preadolescents and Adolescents: A Systematic Review. *Journal of medical Internet research*, 19(5):e176.

[Grouzet et al., 2004] Grouzet, F. M. E., Vallerand, R. J., Thill, E. E., and Provencher, P. J. (2004). From Environmental Factors to Outcomes: A Test of an Integrated Motivational Sequence. *Motivation and Emotion*, 28(4):331–346.

[Hagger and Chatzisarantis, 2015] Hagger, M. and Chatzisarantis, N. (2015). Self-Determination Theory. In Conner, M. and Norman, P., editors, *Predicting and changing health behaviour: Research and practice with social cognition models*, pages 107–141. Open University Press, Maidenhead, UK.

[Hagger and Chatzisarantis, 2014] Hagger, M. S. and Chatzisarantis, N. L. D. (2014). An Integrated Behavior Change Model for Physical Activity. *Exercise and Sport Sciences Reviews*, 42(2):62–69.

[Harackiewicz et al., 1998] Harackiewicz, J. M., Barron, K. E., and Elliot, A. J. (1998). Rethinking achievement goals: When are they adaptive for college students and why? *Educational Psychologist*, 33(1):1–21.

[Harrison et al., 2015] Harrison, D., Marshall, P., Bianchi-Berthouze, N., and Bird, J. (2015). Activity tracking: Barriers, workarounds and customisation. In *UbiComp 2015 - Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 617–621. Association for Computing Machinery, Inc.

[Haskell et al., 2007] Haskell, W. L., Lee, I. M., Pate, R. R., Powell, K. E., Blair, S. N., Franklin, B. A., MacEra, C. A., Heath, G. W., Thompson, P. D., and Bauman, A. (2007). Physical activity and public health: Updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. *Medicine and Science in Sports and Exercise*, 39(8):1423–1434.

[Heffner et al., 2015] Heffner, J. L., Vilardaga, R., Mercer, L. D., Kientz, J. A., and Bricker, J. B. (2015). Feature-level analysis of a novel smartphone application for smoking cessation. *The American Journal of Drug and Alcohol Abuse*, 41(1):68–73.

[Hekler et al., 2013] 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 CHI'13*, page 3307, New York, New York, USA. ACM Press.

[Helsana Insurance Ltd, 2019] Helsana Insurance Ltd (2019). Helsana+: We reward you for a healthy lifestyle. Last accessed August 22, 2019.

[Hewett and Conway, 2016] Hewett, R. and Conway, N. (2016). The undermining effect revisited: The salience of everyday verbal rewards and self-determined motivation. *Journal of Organizational Behavior*, 37(3):436–455.

[Hiniker et al., 2016] Hiniker, A., Hong, S., Kohno, T., and Kientz, J. A. (2016). MyTime: Designing and evaluating an intervention for smartphone non-use. In *Conference on Human Factors in Computing Systems - Proceedings*, pages 4746–4757, New York, NY, USA. Association for Computing Machinery.

[Hsu et al., 2014] Hsu, A., Yang, J., Yilmaz, Y. H., Haque, M. S., Can, C., and Blandford, A. E. (2014). Persuasive technology for overcoming food cravings and improving snack choices. In *Proceedings of CHI'14*, pages 3403–3412, New York, USA. ACM.

[IAC Search & Media Technologies Limited, 2019] IAC Search & Media Technologies Limited (2019). SnapCalc: Math Problem Solver. Last accessed August 22, 2019.

[inc., 2019a] inc., C. (2019a). Calm app: Meditation techniques for sleep and stress reduction. Last accessed August 22, 2019.

[inc., 2019b] inc., S. (2019b). Success Factors: Human Resource Management. Last accessed August 22, 2019.

[iTranslate GmbH, 2019] iTranslate GmbH (2019). iTranslate Lingo: The easiest way to learn a language. Last accessed August 22, 2019.

[iTranslate GmbH, 2019] iTranslate GmbH (2019). iTraslate Lingo: The easiest way to learn a language. Last accessed August 22, 2019.

[Jebelle and Burrows, 2019] Jebelle, H. and Burrows, T. (2019). A weight loss app may be a risky way to address obesity in children.

[Jeffery et al., 2000] Jeffery, R. W., Epstein, L. H., Wilson, G. T., Drewnowski, A., Stunkard, A. J., and Wing, R. R. (2000). Long-term maintenance of weight loss: Current status. *Health Psychology*, 19(1, Suppl):5–16.

[Jonathan and Keith, 2009] Jonathan, E. and Keith, F. (2009). *In Two Minds: Dual Processes and Beyond - Open Research Online*. Oxford University Press, London, UK.

[Joulebug Inc., 2019] Joulebug Inc. (2019). JouleBug: make your everyday habits more sustainable. Last accessed August 22, 2019.

[Kahoot AS, 2019] Kahoot AS (2019). Kahoot! mobile app: Play, create, host quizzes, even on the go. Last accessed August 22, 2019.

[Kappen and Nacke, 2013] Kappen, D. L. and Nacke, L. E. (2013). The kaleidoscope of effective gamification: Deconstructing gamification in business applications. In *Proceedings of the First International Conference on Gameful Design, Research, and Applications*, Gamification '13, pages 119–122, New York, USA. ACM.

[Kast and Connor, 1988] Kast, A. and Connor, K. (1988). Sex and Age Differences in Response to Informational and Controlling Feedback. *Personality and Social Psychology Bulletin*, 14(3):514–523.

[Kennedy and Gregoire, 2009] Kennedy, K. and Gregoire, T. K. (2009). Theories of Motivation in Addiction Treatment: Testing the Relationship of the Transtheoretical Model of Change and Self-Determination Theory. *Journal of Social Work Practice in the Addictions*, 9(2):163–183.

[Khaled et al., 2009] Khaled, R., Barr, P., Biddle, R., Fischer, R., and Noble, J. (2009). Game design strategies for collectivist persuasion. In *Proceedings of the 2009 ACM SIGGRAPH Symposium on Video Games*, Sandbox '09, pages 31–38, New York, USA. ACM.

[Kim et al., 2019] Kim, J., Jung, H., Ko, M., and Lee, U. (2019). GoalKeeper. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(1):1–29.

[Kinnafick et al., 2016] Kinnafick, F. E., Thøgersen-Ntoumani, C., and Duda, J. (2016). The effect of need supportive text messages on motivation and physical activity behaviour. *Journal of Behavioral Medicine*, 39(4):574–586.

[Ko et al., 2016] Ko, M., Choi, S., Yatani, K., and Lee, U. (2016). Lock n' LoL: Group-based limiting assistance app to mitigate smartphone distractions in group activities. In *Conference on Human Factors in Computing Systems - Proceedings*, pages 998–1010, New York, NY, USA. Association for Computing Machinery.

[Konrad et al., 2015] Konrad, A., Bellotti, V., Crenshaw, N., Tucker, S., Nelson, L., Du, H., Pirolli, P., and Whittaker, S. (2015). Finding the Adaptive Sweet Spot: Balancing Compliance and Achievement in Automated Stress Reduction. In *Proceedings of CHI'15*, pages 3829–3838, New York, New York, USA. ACM Press.

[Kwit Ltd., 2019] Kwit Ltd. (2019). Kwit: Be sure to never smoke again. Last accessed August 22, 2019.

[Lally et al., 2010] Lally, P., Van Jaarsveld, C. H. M., W. W. POTTS, H., and Wardle, J. (2010). How are habits formed: Modeling habit formation in the real world.

European Journal of Social Psychology Eur., 40(June 2009):625–634.

[Lazar et al., 2015] Lazar, A., Koehler, C., Tanenbaum, J., and Nguyen, D. H. (2015). Why we use and abandon smart devices. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 635–646. ACM.

[Leaf Group Ltd., 2019] Leaf Group Ltd. (2019). MyPlate Calorie Counter: Calories, Macros and Fitness. Last accessed August 22, 2019.

[Leavell, 2017] Leavell, J. P. (2017). Controlling and Informational Planned Behavior: Self-Determination Theory and the Theory of Planned Behavior. *Atlantic Marketing Journal*, 5(3):81–91.

[Lee et al., 2018] Lee, M., Lee, H., Kim, Y., Kim, J., Cho, M., Jang, J., and Jang, H. (2018). Mobile App-Based Health Promotion Programs: A Systematic Review of the Literature. *International journal of environmental research and public health*, 15(12):2838.

[Legault et al., 2007] Legault, L., Green-Demers, I., Grant, P., and Chung, J. (2007). On the Self-Regulation of Implicit and Explicit Prejudice: A Self -Determination Theory Perspective. *Personality and Social Psychology Bulletin*, 33(5):732–749.

[Lesson Nine GmbH, 2019] Lesson Nine GmbH (2019). Babbel: Speak a new language with confidence. Last accessed August 22, 2019.

[Levi, 2019] Levi, A. (2019). My Diet Coach: Weight Loss. Last accessed August 22, 2019.

[Lift Worldwide, 2019] Lift Worldwide, I. (2019). A Home for Ambitious People: Build your career. Get in shape. Learn a new skill. Last accessed August 22, 2019.

[Lister et al., 2014] 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):e9.

[Locke and Latham, 2002] 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):705–717.

[Lumos Labs Inc., 2019] Lumos Labs Inc. (2019). Luminosity: Discover what your mind can do. Last accessed August 22, 2019.

[Lyngs et al., 2019] Lyngs, U., Lukoff, K., Slovak, P., Binns, R., Slack, A., Inzlicht, M., Van Kleek, M., and Shadbolt, N. (2019). Self-Control in Cyberspace: Applying Dual

Systems Theory to a Review of Digital Self-Control Tools. In *Proceedings of CHI'19*, pages 1–18, New York, USA. ACM Press.

[Martela and Ryan, 2016] Martela, F. and Ryan, R. M. (2016). The Benefits of Benevolence: Basic Psychological Needs, Beneficence, and the Enhancement of Well-Being. *Journal of Personality*, 84(6):750–764.

[McAuley et al., 1989] McAuley, E., Duncan, T., and Tammen, V. V. (1989). Psychometric Properties of the Intrinsic Motivation Inventory in a Competitive Sport Setting: A Confirmatory Factor Analysis. *Research Quarterly for Exercise and Sport*, 60(1):48–58.

[McLeroy et al., 1988] McLeroy, K. R., Bibeau, D., Steckler, A., and Glanz, K. (1988). An Ecological Perspective on Health Promotion Programs. *Health Education Quarterly*, 15(4):351–377.

[Mehrotra et al., 2016] Mehrotra, A., Pejovic, V., Vermeulen, J., Hendley, R., and Musolesi, M. (2016). My Phone and Me: Understanding People's Receptivity to Mobile Notification. In *Proceedings of CHI'16*, pages 1021–1032, New York, USA. ACM Press.

[Memrise Inc., 2019] Memrise Inc. (2019). Memrise: The fastest way to learn a language. Last accessed August 22, 2019.

[Michie et al., 2013] Michie, S., Richardson, M., Johnston, M., Abraham, C., Francis, J., Hardeman, W., Eccles, M. P., Cane, J., and Wood, C. E. (2013). The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. *Annals of behavioral medicine : a publication of the Society of Behavioral Medicine*, 46(1):81–95.

[Midden and Ham, 2018] Midden, C. and Ham, J. (2018). Environmental Psychology. In Steg, L. and Groot, J. I., editors, *Environmental Psychology*, Wiley Online Books, chapter Persuasive, pages 283–294. Wiley, Hoboken, NJ, US.

[Minutes, 2020] Minutes, . (2020). La suisse reste le pays de l'iphone. Last accessed February 26, 2020.

[murbit GmbH, 2019] murbit GmbH (2019). Drink Water: Health Reminder. Last accessed August 22, 2019.

[Nexercise Inc., 2019] Nexercise Inc. (2019). Sworkit: Fitness Made Simple. Last accessed August 22, 2019.

[Niemiec et al., 2010a] Niemiec, C. P., Brown, K. W., Kashdan, T. B., Cozzolino, P. J., Breen, W. E., Levesque-Bristol, C., and Ryan, R. M. (2010a). Being present in the face of existential threat: The role of trait mindfulness in reducing defensive responses to mortality salience. *Journal of Personality and Social Psychology*, 99(2):344–365.

[Niemiec et al., 2010b] Niemiec, C. P., Ryan, R. M., and Deci, E. L. (2010b). Self-determination theory and the relation of autonomy to self-regulatory processes and personality development. In Hoyle, R. H., editor, *Handbook of Personality and Self-Regulation*, Wiley Online Books, pages 169–191. Wiley, Hoboken, NJ, US.

[Oinas-Kukkonen and Harjumaa, 2009] Oinas-Kukkonen, H. and Harjumaa, M. (2009). Persuasive Systems Design: Key Issues, Process Model, and System Features. *Communications of the Association for Information Systems*, 24(28):485–500.

[Okumus et al., 2016] Okumus, B., Bilgihan, A., and Ozturk, A. B. (2016). Factors Affecting the Acceptance of Smartphone Diet Applications. *Journal of Hospitality Marketing & Management*, 25(6):726–747.

[Orji et al., 2017] Orji, R., Nacke, L. E., and Di Marco, C. (2017). Towards personality-driven persuasive health games and gamified systems. In *Proceedings of CHI'17*, pages 1015–1027, New York, USA. ACM.

[Orji et al., 2014] Orji, R., Vassileva, J., and Mandryk, R. L. (2014). Modeling the efficacy of persuasive strategies for different gamer types in serious games for health. *User Modeling and User-Adapted Interaction*, 24(5):453–498.

[Patall et al., 2010] Patall, E. A., Cooper, H., and Wynn, S. R. (2010). The effectiveness and relative importance of choice in the classroom. *Journal of Educational Psychology*, 102(4):896–915.

[Patel et al., 2016] Patel, M. S., Asch, D. A., Rosin, R., Small, D. S., Bellamy, S. L., Heuer, J., Sproat, S., Hyson, C., Haff, N., Lee, S. M., Wesby, L., Hoffer, K., Shuttleworth, D., Taylor, D. H., Hilbert, V., Zhu, J., Yang, L., Wang, X., and Volpp, K. G. (2016). Framing Financial Incentives to Increase Physical Activity Among Overweight and Obese Adults. *Annals of Internal Medicine*, 164(6):385.

[Pielot and Rello, 2017] Pielot, M. and Rello, L. (2017). Productive, anxious, lonely: 24 hours without push notifications. In *Proceedings of MobileHCI'17*, pages 1–11, New York, USA. ACM.

[Pinder et al., 2018] Pinder, C., Vermeulen, J., Cowan, B. R., and Beale, R. (2018).

Digital behaviour change interventions to break and form habits. *ACM Transactions on Computer-Human Interaction*, 25(3).

[Pinder et al., 2017] Pinder, C., Vermeulen, J., Cowan, B. R., Beale, R., and Hendley, R. J. (2017). Exploring the feasibility of subliminal priming on smartphones. In *Proceedings of MobileHCI'17*, pages 1–15, New York, New York, USA. ACM Press.

[Prochaska and Di Clemente, 1983] Prochaska, J. O. and Di Clemente, 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):390–395.

[Purpura et al., 2011] Purpura, S., Schwanda, V., Williams, K., Stubler, W., and Sengers, P. (2011). Fit4life: The design of a persuasive technology promoting healthy behavior and ideal weight. In *Proceedings of CHI'11*, pages 423–432, New York, USA. ACM.

[Radel et al., 2017] Radel, R., Pelletier, L., Pjevac, D., and Cheval, B. (2017). The links between self-determined motivations and behavioral automaticity in a variety of real-life behaviors. *Motivation and Emotion*, 41(4):443–454.

[Rapp et al., 2019] Rapp, A., Tirassa, M., and Tirabeni, L. (2019). Rethinking Technologies for Behavior Change: A View from the Inside of Human Change. *ACM Transactions on Computer-Human Interaction*, 26(4):1–30.

[Reeve and Deci, 1996] Reeve, J. and Deci, E. L. (1996). Elements of the Competitive Situation that Affect Intrinsic Motivation. *Personality and Social Psychology Bulletin*, 22(1):24–33.

[Rejeski and Fanning, 2019] Rejeski, W. J. and Fanning, J. (2019). Models and theories of health behavior and clinical interventions in aging: a contemporary, integrative approach. *Clinical Interventions in Aging*, 14:1007–1019.

[Renfree et al., 2016] 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 *Conference on Human Factors in Computing Systems - Proceedings*, volume 07-12-May-2016, pages 2932–2939. Association for Computing Machinery.

[Rosenstock, 1974] Rosenstock, I. M. (1974). The Health Belief Model and Preventive Health Behavior. *Health Education Monographs*, 2(4):354–386.

[Runtastic Inc., 2019] Runtastic Inc. (2019). Runtasty - Healthy Recipes. Last accessed August 22, 2019.

[Ryan and Deci, 2000a] Ryan, R. and Deci, E. (2000a). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *The American psychologist*, 55(1):68–78.

[Ryan, 1982] Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of personality and social psychology*, 43(3):450.

[Ryan, 1993] Ryan, R. M. (1993). Agency and organization: Intrinsic motivation, autonomy, and the self in psychological development. In *Nebraska Symposium on Motivation, 1992: Developmental perspectives on motivation.*, Current theory and research in motivation, Vol. 40., pages 1–56. University of Nebraska Press, Lincoln, NE, US.

[Ryan et al., 1985] Ryan, R. M., Connell, J. P., and Deci, E. L. (1985). A motivational analysis of self-determination and self-regulation in education. *Research on motivation in education: The classroom milieu*, 2:13–51.

[Ryan and Deci, 2000b] Ryan, R. M. and Deci, E. L. (2000b). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology*, 25(1):54–67.

[Ryan and Deci, 2000c] Ryan, R. M. and Deci, E. L. (2000c). When rewards compete with nature. In *Intrinsic and Extrinsic Motivation*, pages 13–54. Elsevier, London, UK.

[Ryan and Deci, 2017] Ryan, R. M. and Deci, E. L. (2017). *Self-Determination Theory Basic Psychological Needs in Motivation, Development and Wellness*. The Guilford Press, New York, USA.

[Ryan et al., 2006] Ryan, R. M., Rigby, C. S., and Przybylski, A. (2006). The Motivational Pull of Video Games: A Self-Determination Theory Approach. *Motivation and Emotion*, 30(4):344–360.

[Ryan et al., 1996] Ryan, R. M., Sheldon, K. M., Kasser, T., and Deci, E. L. (1996). All goals are not created equal: An organismic perspective on the nature of goals and their regulation. In Gollwitzer, P. M. and Bargh, J. A., editors, *The psychology of action: Linking cognition and motivation to behavior.*, pages 7–26. Guilford Press, New York, US.

[Sansone and Harackiewicz, 1996] Sansone, C. and Harackiewicz, J. M. (1996). "I don't feel like it": The function of interest in self-regulation. In *Striving and feeling:*

Interactions among goals, affect, and self-regulation., pages 203–228. Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, US.

[Schelling, 1992] Schelling, T. C. (1992). Self-command: A new discipline. In *Choice over time.*, pages 167–176. Russell Sage Foundation, New York, NY, US.

[Schultz and Ryan, 2015] Schultz, P. P. and Ryan, R. M. (2015). The "Why," "What," and "How" of Healthy Self-Regulation: Mindfulness and Well-Being from a Self-Determination Theory Perspective. In Ostafin, B. D., Robinson, M. D., and Meier, B. P., editors, *Handbook of Mindfulness and Self-Regulation*, pages 81–94. Springer New York, New York, NY.

[Sheeran et al., 2017] Sheeran, P., Godin, G., Conner, M., and Germain, M. (2017). Paradoxical Effects of Experience: Past Behavior Both Strengthens and Weakens the Intention-Behavior Relationship. *Journal of the Association for Consumer Research*, 2(3):309–318.

[Sheldon et al., 2001] Sheldon, K. M., Elliot, A. J., Kim, Y., and Kasser, T. (2001). What is satisfying about satisfying events? Testing 10 candidate psychological needs. *Journal of Personality and Social Psychology*, 80(2):325–339.

[Simple Habits Inc., 2019] Simple Habits Inc. (2019). Simple Habits: Stress relief for busy people. Last accessed August 22, 2019.

[Sjöklint et al., 2015] Sjöklint, M., Constantiou, I. D., and Trier, M. (2015). The complexities of self-Tracking - An inquiry into user reactions and goal attainment. In *23rd European Conference on Information Systems, ECIS 2015*, volume 2015-May, pages 1–15. Association for Information Systems.

[Skarecki, 2015] Skarecki, B. (2015). Running Apps Have Hurt My Progress, Not Helped It.

[Smith and Sarason, 1975] Smith, R. E. and Sarason, I. G. (1975). Social anxiety and the evaluation of negative interpersonal feedback. *Journal of consulting and Clinical Psychology*, 43(3):429.

[srl, 2019] srl, G. (2019). GreenApes: The social network rewarding sustainable actions and ideas. Last accessed August 22, 2019.

[Standage and Ryan, 2012] Standage, M. and Ryan, R. M. (2012). Self-determination theory and exercise motivation: facilitating self-regulatory processes to support and maintain health and well-being. In Roberts, G. C. and Treasure, D. C., editors, *Advances in motivation in sport and exercise*, pages 233–270. Human Kinetics, Champaign, US.

[Stanovich, 2010] Stanovich, K. (2010). *Rationality and the Reflective Mind*. Oxford University Press, New York, NY, US.

[Statcounter, 2020] Statcounter (2020). Mobile Operating System Market Share Switzerland. Last accessed February 26, 2020.

[Stawarz et al., 2015] Stawarz, K., Cox, A. L., and Blandford, A. (2015). Beyond Self-Tracking and Reminders: Designing Smartphone Apps That Support Habit Formation. In *Proceedings of CHI'15*, pages 2653–2662, New York, USA. ACM Press.

[Stawarz et al., 2014] Stawarz, K., Cox, A. L., Blandford, A., Stawarz, K., Cox, A. L., and Blandford, A. (2014). Don't forget your pill!: designing effective medication reminder apps that support users' daily routines. In *Proceedings of CHI'14*, pages 2269–2278, New York, New York, USA. ACM Press.

[Stawarz et al., 2018] Stawarz, K., Preist, C., Tallon, D., Wiles, N., and Coyle, D. (2018). User Experience of Cognitive Behavioral Therapy Apps for Depression: An Analysis of App Functionality and User Reviews. *J Med Internet Res*, 20(6):e10120.

[Stibe and Cugelman, 2016] Stibe, A. and Cugelman, B. (2016). Persuasive backfiring: When behavior change interventions trigger unintended negative outcomes. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 9638, pages 65–77. Springer Verlag.

[StickK Ltd., 2019] StickK Ltd. (2019). Stickk: Stick to your commitments. Last accessed August 22, 2019.

[Strack and Deutsch, 2004] Strack, F. and Deutsch, R. (2004). Reflective and Impulsive Determinants of Social Behavior. *Personality and Social Psychology Review*, 8(3):220–247.

[Strava Inc., 2019] Strava Inc. (2019). Strava: app for runners and cyclists. Last accessed August 22, 2019.

[Suh et al., 2015] Suh, A., Wagner, C., and Liu, L. (2015). The Effects of Game Dynamics on User Engagement in Gamified Systems. In *48th Hawaii International Conference on System Sciences HICSS*, volume 2015, pages 672–681, Kauai, HI, USA.

IEEE Computer Society Press.

[Sundar and Marathe, 2010] Sundar, S. S. and Marathe, S. S. (2010). Personalization versus customization: The importance of agency, privacy, and power usage. *Human Communication Research*, 36(3):298–322.

[Tauer and Harackiewicz, 1999] Tauer, J. M. and Harackiewicz, J. M. (1999). Winning Isn't Everything: Competition, Achievement Orientation, and Intrinsic Motivation. *Journal of Experimental Social Psychology*, 35(3):209–238.

[Tauer and Harackiewicz, 2004] Tauer, J. M. and Harackiewicz, J. M. (2004). The effects of cooperation and competition on intrinsic motivation and performance. *Journal of personality and social psychology*, 86(6):849–61.

[Thaler and Shefrin, 1981] Thaler, R. H. and Shefrin, H. M. (1981). An Economic Theory of Self-Control. *Journal of Political Economy*, 89(2):392–406.

[The Today App Ltd., 2019] The Today App Ltd. (2019). Today: Creating Daily Habits for a Better World. Last accessed August 22, 2019.

[Under Armour Inc., 2019] Under Armour Inc. (2019). MyFitnessPal: Fitness starts with what you eat. Last accessed August 22, 2019.

[Vallerand and Losier, 1999] Vallerand, R. J. and Losier, G. F. (1999). An integrative analysis of intrinsic and extrinsic motivation in sport. *Journal of Applied Sport Psychology*, 11(1):142–169.

[Vallerand and Reid, 1984] Vallerand, R. J. and Reid, G. (1984). On the Causal Effects of Perceived Competence on Intrinsic Motivation: A Test of Cognitive Evaluation Theory. *Journal of Sport Psychology*, 6(1):94–102.

[van Hooft, 2018] van Hooft, E. (2018). Self-Regulatory Perspectives in the Theory of Planned Job Search Behavior: Deliberate and Automatic Self-Regulation Strategies to Facilitate Job Seeking. In Klehe, U.-C. and Hooft, E., editors, *The Oxford Handbook of Job Loss and Job Search*, page 28. Oxford University Press, London, UK.

[Vansteenkiste et al., 2004] Vansteenkiste, M., Simons, J., Lens, W., Sheldon, K. M., and Deci, E. L. (2004). Motivating learning, performance, and persistence: The synergistic effects of intrinsic goal contents and autonomy-supportive contexts. *Journal of personality and social psychology*, 87(2):246.

[Verplanken and Wood, 2006] Verplanken, B. and Wood, W. (2006). Interventions to

Break and Create Consumer Habits. *Journal of Public Policy & Marketing*, 25(1):90–103.

[Wehmeyer et al., 2017] Wehmeyer, M. L., Shogren, K. A., Little, T. D., and Lopez, S. (2017). *Development of self-determination through the life-course*. Springer, Dordrecht, The Netherlands.

[Wertenbroch, 1998] Wertenbroch, K. (1998). Consumption Self-Control by Rationing Purchase Quantities of Virtue and Vice. *Marketing Science*, 17(4):317–337.

[Yazio GmbH, 2019] Yazio GmbH (2019). Yazio: Welcome to a Healthier Life. Last accessed August 22, 2019.

[Youper Inc., 2019] Youper Inc. (2019). Youper: Emotional Health Assistant. Last accessed August 22, 2019.

Chapter 2

Informed Choices, Progress Monitoring and Comparison with Peers

2.1 Introduction to the Chapter

In this chapter, we introduce the second research stream. Grounding on the results from chapter 2, we followed a design through theory approach to create app features that supported the Basic Psychological Needs. We used these features to create a steps-counter app which we tested and captured participants' impressions of the app features.

2.2 Introduction

Many people around the world use apps on their smartphones (e.g., MyFitnessPal [Under Armour Inc., 2020], Endonomdo [Endomondo, 2020]) to monitor their physical exercises and improve their level of physical activity. Recent representative surveys show that over 100k smartphone health apps available worldwide, and 500M users use mobile health applications to keep track of their everyday activities [Edwards et al., 2016, Fox and Duggan, 2012]. In this paper, we refer to this group of applications as *behavior change apps*, as these have been explicitly designed *to foster and assist behavior change and sustainment* [Hekler et al., 2013, p. 3308]. Behavior change apps incorporate various features to modify users' behavior, many of which are used simultaneously (e.g., goal-setting, performance sharing, reminders).

Recently, researchers started to question the usefulness and efficacy of some of these apps [Ferrara, 2013,Skarecki, 2015,Jebelle and Burrows, 2019,Covolo et al., 2017], and call for further exploration in the area, for example, looking at the optimal number and combination of app features [Schoeppe et al., 2016].

Given that commercial apps, in most cases, incorporate multiple features, they are often unsuited for controlled experiments as it remains challenging to identify the effects of the individual features on one's behavior. This identification would be necessary to design more effective and efficient behavior change apps.

Furthermore, prior research has not extensively explored users' perceptions of behavior-change apps on a minimal set of behavior change features based on established behavioral change theories. Users' perceptions can be seen as good predictors of engagement, motivation and well-being [Ryan and Deci, 2017, p.20,p.213] [Peters et al., 2018]. Therefore, capturing and understanding the users' perception can offer researchers valuable insights into how technological features might influence users' behavior. This marks a salient motivation of our work.

In this study, we design, develop, and deploy AGON, a step-counter app for smartphones. This app's design is grounded on the Self-Determination Theory (SDT), a human motivation theory that focuses on the types and sources of motivation that impact behavior. The SDT has been successfully applied across multiple life domains (e.g., see [Deci and Ryan, 2008] for a review). The theory posits that human beings have three basic psychological needs (BPNs): autonomy, competence, and relatedness (we will define them in the next section). Furthermore, the theory states that the satisfaction of the BPNs is a requirement for optimal development, integrity, and well-being [Ryan and Deci, 2017, p.242]. Recently a taxonomy [Villalobos-Zúñiga and Cherubini, 2020] mapped app features to these BPNs, providing a tool for designers to evaluate how app features may motivate individuals towards their goals.

As a first attempt to design SDT-informed app features, we build on this taxonomy and create AGON. This app incorporates three distinctive features: (1) goal setting to support *autonomy*; (2) history to support *competence*; and (3) peers comparison (in the form of step-counts) to support *relatedness*. We hypothesized that in a longitudinal deployment, participants would perceive support to the BPNs and feel motivated towards the target activity (i.e., walking or running). We then formulated the following research question:

RQ: What are the individuals' perceptions about the hypothesized app features that aim to support autonomy, competence, and relatedness when it comes to improving their physical activity?

We conducted a field study of the app that spanned 4 weeks to address this research question. We recruited 49 participants for the study. We asked our participants to fill a diary documenting positive and negative experiences. Subsequently, we invited 15 of them to a follow-up contextual interview to better understand their experience with the app and their reflections. Our results indicated that: participants experienced feelings of being autonomous when setting a step goal; also the use of the app provoked participants reflections, self-evaluations, and contemplation of their physical activity routines; finally, the design of the relatedness feature elicit feelings of companionship, comparison, and competition, but not feelings of being connected to each other, as we initially expected.

We contribute a novel design of a fitness app derived directly from the app features' taxonomy based on the SDT. This app was intentionally designed to support *autonomy*, *competence*, and *relatedness* towards the goal of improving the individuals' physical activity levels. This study also contributes qualitative empirical evidence of the users' perceptions of how the app supported their BPNs. We support app designers with suggestions on how to improve the design of features that aim to support the BPNs.

2.3 Background and Related work

Our work lies at the intersection of two principal research areas: (i) a contemporary theory of motivation (i.e., Self-Determination Theory) and (ii) empirical research in HCI on designing behavior change app features to support physical activity.

2.3.1 Self-Determination Theory

The SDT posits that people have different *levels* or *amounts* of motivation to perform a specific activity. It also states that people have different types of *orientation of motivation* i.e. the underlying attitudes, goals and values that give rise to action [Deci and Ryan, 1985]. These types are classified as *intrinsic* and *extrinsic* motivation. Intrinsic motivation refers to doing an activity because the person finds it inherently interesting or enjoyable (e.g., reading a book). Extrinsic motivation refers to doing an activity because it leads to a separate outcome (e.g., preparing for an exam). Therefore, different types of motivation differ in the sources that initiate them, in magnitude, in affects, and in the experiences of the individual and their behavioral consequences [Ryan and Deci, 2017, p. 14].

Moreover, intrinsically motivated behaviors are autonomous and experienced as being volitional. In contrast, extrinsically motivated behaviors can *vary widely in the degree to which they are controlled versus autonomous* [Ryan and Connell, 1989]. For example, a student may be extrinsically motivated to study for an exam to avoid the punishment of parents but also could be motivated because they observe a valued outcome (i.e., getting a degree).

More specifically, the SDT affirms that the extrinsic motivation can be *internalized* and that the degree of internalization demonstrates the degree to which the behavioral regulation is relatively autonomous versus controlled [Ryan and Deci, 2017, p. 14]. Consequently, the SDT introduces a *control–autonomy continuum* to explain this internalization process. It spans from *amotivation* (or absence of intention to act) to *external regulation* (to obtain a reward) to *introjected regulation* (to avoid guilt) to *identification* (accepted external regulation) to *integration* (self-determined action).

Additionally, the SDT explains that these –previously mentioned– extrinsic motivation types can urge a person to behave a certain way in the short-term but will fail to maintain the behavior over more extended periods [Deci and Ryan, 1985]. As a result, behavior-change interventions designed for extrinsic motivation types may not sustain the new behavior after the intervention ends.

Mainly, the theory describes three Basic Psychological Needs (or BPN), that when satisfied by the contextual conditions, leads to a self-determined action.

In this work, we adopt the following definitions for each BPNs:

1. *Autonomy* "refers to feeling willingness and volition with respect to one's behavior. The need for autonomy describes the need of individuals to experience self-endorsement and ownership of their actions." [Ryan and Deci, 2017, p. 86]

2. *Competence* "refers to feeling effective in one's interactions with the social environment–that is, experiencing opportunities and support for the exercise, expansion, and expression of an individual's capacities and talents." [Ryan and Deci, 2017, p. 86]

3. *Relatedness* "refers to both experiencing others as responsive and sensitive and being able to be responsive and sensitive to them–that is, feeling connected and involved with others and having a sense of belonging." [Ryan and Deci, 2017, p. 86]

In our work, we set to understand how the BPNs constructs could inform the design of a mobile app to support physical activity and, subsequently, elicit the users' perceptions of them.

2.3.2 Behavior Change App Features

Goal-Setting

Researchers have created various implementations of goal-setting features in smartphone apps. For example, Consolvo et al. [Consolvo et al., 2006] devised and field-tested an interactive prototype for mobile devices where the step goal was determined based on 1-week data of previously recorded users' step activity. Subsequently, they [Consolvo et al., 2009] developed yet another mobile phone prototype, which offers participants to specify their physical activity goals by themselves. Later, Munson et al. [Munson and Consolvo, 2012] experimented with a physical activity mobile app where users selected predefined categories in which their weekly physical activity goals fit best. More recently, Gouveia et al. [Gouveia et al., 2015], developed HABITO, a step-tracking app, which defines two goal-setting mechanisms: one in which users established the daily distance they want to walk and a second one in which the app offered a default walking distance. Hartzler et al. [Hartzler et al., 2016] developed NUTRIWALKING, an app, which offered personalized daily exercise goals. Their goal-setting feature suggested options with exercise duration, based on participants' self-reported baseline level of physical activity.

Finally, other previous research efforts identified the components of appropriate goal-setting strategies to support physical activity. They suggest tailoring the goal difficulty to the user's ability level and re-evaluating the goals based on achievements to increase the qualities of the goal-setting functionality [Baretta et al., 2019]. In sum, this research strand focused on a technological mechanism to self-set goals by selecting from a list of options or inputting the objective directly. In light of this prior research, our work explores how providing participants with information about their physical activity (e.g., previous week's daily average step-counts) and suggesting goals based on their performance can help them make an informed decision concerning their goals.

Progress Monitoring

Researchers followed different approaches to communicate app users their activity progress. For example, in their UbiFit system, Consolvo et al. [Consolvo et al., 2009] used a garden metaphor that blooms throughout the week as users conduct their physical activities. Harries et. al [Harries et al., 2016] presented a step-tracking mobile phone app where participants could see their total daily steps (in a numerical form and in the form of line graphs to overview weekly step progress) after a running workout. Their participants also had the option of viewing step data for the previous day, past week, and their history. Munson et al. [Munson and Consolvo, 2012] explored various progress visualizations: included bar and line charts with completion percentages of the user's goals. Sankaran et al. [Sankaran et al., 2016] developed a specific app for cardiac tele-rehabiliatation: the participants monitored their progress on a horizontal progress bar with an animated person running towards the goal. Oyibo et al. [Oyibo et al., 2019] in their BEN'FIT system adopted a horizontal bar to show the users' weekly physical activity levels. In sum, this research strand focused on the use of visual elements or graphs to communicate activity progress to its users. We aim to extend this research by investigating how a text-based list of steps counts with a temporal component can help participants relate their activity performance with their day-to-day activities.

Peers Comparison

Several prior research efforts developed interactive prototypes to encourage physical activity, exploiting various social support strategies and techniques. One of them is Houston [Consolvo et al., 2006], a mobile application that shares step-counts with friends in the form of achieved activity levels and progress towards the goal. Its field deployment suggested that the participants felt social pressure to achieve a given objective since they did not want to be the last in a leader-board list or wanted to perform better than a friend from the list. Colusso et al. [Colusso et al., 2016] studied the concept of closeness to comparison in the context of a video game, where participants compared their scores (using bar graphs) to the one they compete with.

Hartzler et al. [Hartzler et al., 2016] in their NutriWalking app incorporated an exercise feature with teams of 10 members, a digital "coach", and free-form interaction through team posts and private peer-to-peer messaging. Altmeyer et al. [Altmeyer et al., 2018] presented a gamified system consisting of a physical activity tracker, a mobile application, and a publicly-accessible (web) application. They demonstrated that social sharing of personal step counts increased the overall number of steps for an individual, arguing that the public disclosure increased the participants' level of responsibility. In sum, this strand of the research focused on reflective strategies to mindfully motivate people to exercise more, and confirmed that sharing physical activity-related details with peers not only contributes to the overall user experience and enjoyment of workouts (e.g., [Munson and Consolvo, 2012, Wozniak et al., 2017]) but can also be a powerful motivator for health activities at large (e.g., [Toscos et al., 2006, Epstein et al., 2015]).

Collectively, prior research endeavors were not necessarily grounded to a motivational theory to create behavior change mobile applications and evaluate users' motivation and attitude towards physical activity (e.g., [Middelweerd et al., 2014]). Whereas our work specifically adopts the SDT to develop a mobile application (to facilitate an increase in physical activity) and explore the users' perceptions on supporting features for the BPNs: autonomy, competence, and relatedness. This marks the novelty of our contribution.

2.4 Research Prototype

2.4.1 Design Process and Rationale

Our goal was to design a steps-tracking app with a minimal set of functionalities that support the BPNs as posited by the SDT. We started the design process by studying the examples of apps and the characteristics of the features presented in the taxonomy of behavior change apps features based on SDT [Villalobos-Zúñiga and Cherubini, 2020]. Then, we filtered by high coverage of taxonomy features and picked the top 2 of each BPNs (we did not limit our choice to one feature to avoid any bias or prejudice towards any feature). Then, we discarded from this selection: *Reminders* as it uses an overcrowded communication channel; and *Performance Sharing* as its implementation may disclose the participants' identity. Finally, the selected feature-set

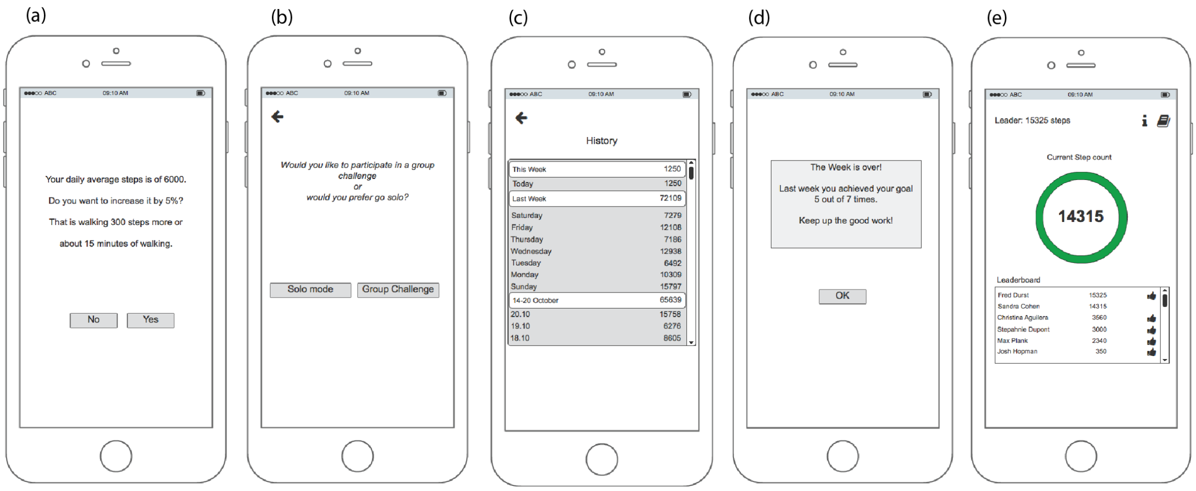

Figure 2.1

(a) Goal setting feature with options to accept or deny the proposition. (b) Goal setting feature to choose weekly group steps competition. (c) History feature: Each row shows the pair date-steps counts. The list is ordered from most recent to oldest. (d) Weekly pop-up message with activity feedback. (e) Peers Comparison showing a list of participants with their total step-counts; at the end of each row a thumbs-up button to provide feedback to other participants.

was: *Goal Setting*, *Activity Feedback*, *History*, and *Peers Comparison*. Next, we used these categories as the foundation for our design process, which we describe in the following paragraphs.

Initial Approach

We considered the *Goal Setting* taxonomy category as the foundation of AGON's *autonomy* support feature. This category describes apps that prompt the user directly with their goals [Villalobos-Zúñiga and Cherubini, 2020]. In our approach, AGON calculates the user's daily average steps using past logs. Then, it displays this information to the user and suggests a percentage increase of this daily average. The user can accept or deny this proposition (see Figure 2.1a). We considered this design might support *autonomy* because it provides information to the user, suggests a concrete goal, and provides an option to accept or refuse this new goal.

Our second feature attempt to support *autonomy* allowed the user to choose between participating in a weekly group steps competition or working towards their weekly step goal individually (see Figure 2.1b). We considered this design might support *autonomy*, because it allows the user to express their will concerning the method that will lead them to their step-goal.

We considered the *History* and *Feedback* taxonomy categories as the foundations of AGON's *competence* support feature. The former presents the user with a representation of their activity over a time period, and the latter provides the user with information about how the task was performed in a given session [Villalobos-Zúñiga and Cherubini, 2020]. Initially, we thought of having a *history* of step-counts including daily-step totals and corresponding dates (see Figure 2.1c). We considered this design might support *competence* because it allows users to reflect on their step-counts and feel effective about their walking activity levels. We thought of using *feedback* as a second way to support *competence* by displaying a pop-up message with weekly activity performance results (see Figure 2.1d). Similarly to the *history* of step-counts, this design allows users to feel effective while doing the walking activity. We must note that the individual's sense of effectiveness will depend on the actual activity performance (e.g., little steps lead to bad performance and consequently bad effectiveness, and vice versa).

We considered the *Peers Comparison* taxonomy category as the foundation of AGON's *relatedness* support feature. This category presents an ordered list of scores and people who perform the same activity [Villalobos-Zúñiga and Cherubini, 2020]. Therefore, we thought of designing a *peers comparison list* including a list of users names with weekly steps-totals (see Figure 2.1e). As a second way to support relatedness, we thought about how users could encourage other competitors by tapping on a thumbs-up icon next to each competitor's name. This design allows users to feel connected and important to each other by knowing that they are taking part in the same activity and by having the chance of supporting themselves through the thumbs-up action.

Refined Approach

In this phase, we performed an expert evaluation of the initial app design, which consisted of reviewing the design and interaction of each app feature from the lens of the Self-Determination Theory. For this activity, we involved five researchers from our institution, all familiar with the SDT, to study a detailed design document and UI maps. Notably, we asked the experts to evaluate whether the app's features provided support to the BPNs posited by the SDT. Further, we asked experts to perform a Heuristic Evaluation of the design [Molich and Nielsen, 1990], to iron out usability issues. The feedback we collected in this phase allowed us to update the design of AGON in the following ways: first, for the *autonomy* feature, we kept the design where

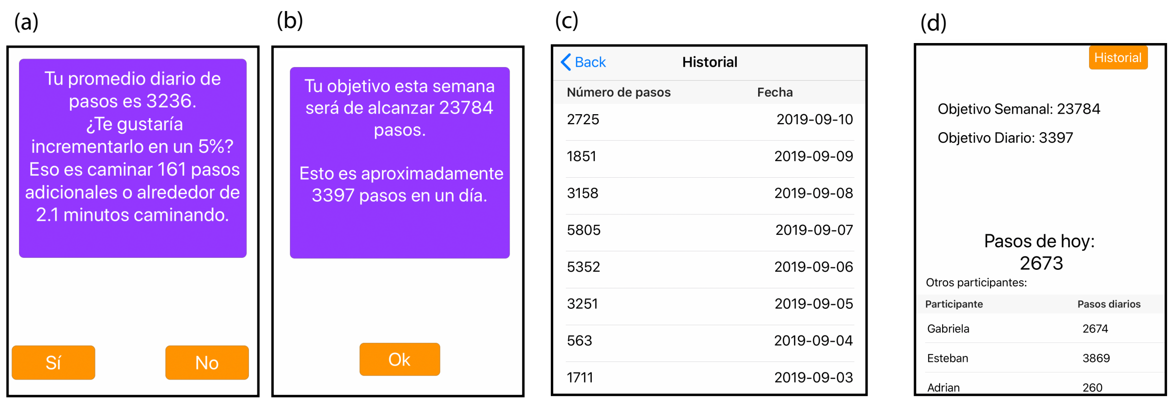

Figure 2.2
(a) Goal-setting feature (first dialogue): *"Your daily step goal is 3236 steps. Would you like to increase it by 5%?. That is walking 161 additional steps or about 2.1 minutes walking."* (b) Goal setting feature (second dialogue): *Your weekly goal will be to reach 23784. This is approximately 3397 steps per day.* (c) History feature: Each row shows the pair steps counts-date. The list is ordered from most recent to oldest. (d) Main Dashboard Screen: Displays the history button on top-right; below the weekly and daily step goals; below the daily step-counter; at the bottom the peers comparison feature.

the app showed the daily steps average, with the option to increase this average or not. We discarded the design that allowed users to choose a group or individual competition, because leaving this option would make the relatedness feature (peers comparison) optional, which was not our intent. Instead, we thought of removing this choice and leaving the group competition (represented by the peers comparison) as a fix app feature.

Next, for the *competence* feature, we kept the history design, which provided more information to satisfy the sense of effectiveness while walking or running. We preferred this design on top of the pop-up message because we could guarantee a longer exposure to the history that remains accessible in the app all the time versus the pop-up message that appears on the screen a limited time (i.e., couple of seconds once a week). Finally, for the *relatedness* feature, we kept the peers comparison (as we mentioned earlier) but removed the thumbs-up because we considered it a double support for the *relatedness* need.

Final Approach

Once we completed the refined version of the designs and the UI-map, we developed AGON. This app implements three features based on each of the BPNs: (1) goal-setting to support *autonomy*; (2) steps-history to support *competence*; and (3) peers comparison to incorporate *relatedness*. AGON differs from other commercial apps in the following ways. First, unlike many commercial apps and services, it is developed upon an established behavior change theory in mind. Second, AGON was purposefully designed with a minimal set of features to research the distinct effects these group of features has on the users' behavior. In the following paragraphs, we detailed the rationale and the relation to the SDT that lead to our final design.

We designed the *goal-setting* feature to support the basic psychological need of *autonomy*. We hypothesized that a *goal-setting-autonomy-support* feature could be represented by a dialogue-box displaying personal information to the users. This information helps users in the goal-setting-decision-making process by informing them of their previous week step-count average and proposing to walk a higher number of steps the following week (see Figure 2.2a). In the same dialogue-box users can see the equivalent number of steps for this increase and the approximate amount of time it would take to walk those steps (e.g., *"Your daily step goal is 2936. Would you like to increase it by 5%? That is to walk 146 additional steps or about 1.9 minutes walking."*)[1]. Below the dialogue box and on the left side, we placed a *yes* button (to accept the step increase) and on the right side a *no* button (to reject the step increase and continue targeting the same average number of steps from their previous week). After the user taps on any of the two buttons, the text in the dialogue box updates to present a sentence indicating the weekly objective and the approximate daily steps (see Figure 2.2b). In the next lines, we relate this feature design to the SDT literature.

According to the SDT, by offering users information (i.e., their average number of steps counts) about a decision they need to make (i.e., defining weekly step-goal), they can take an informed choice and thus feel autonomous when specifying their weekly step-goal [Ryan and Deci, 2017, p. 461]. Also, the action of goal-setting contributes to an *internal perceived locus of causality* (or I-PLOC) [DeCharms, 1968]. In concrete, an intentional behavior can be either internally motivated (i.e., intrinsic motivation) or externally motivated (i.e., extrinsic motivation). For instance, researchers found that students' learning outcomes increased when teachers provided them options and choices in a learning environment [Deci et al., 1996, Patall et al., 2010]. The SDT states that when a goal is imposed by someone else (or by an app), it will undermine autonomy and reduce the person's motivation to perform the activity. On the other

[1] The recommendation for the weekly physical increase is no more than 10% [University of Colorado Hospital, 2003]

side, a self-imposed goal would contribute to the basic need of autonomy and supports autonomous motivation to perform the target activity (cf. [Deci and Ryan, 1980, Ryan et al., 1985, Ryan et al., 1996]).

We designed the *steps-history* feature to support the basic psychological need of *competence*. The *steps-history* feature is a simplified version of related functionalities adapted from FITBIT [Fitbit Inc., 2020] and STEPS APP [StepsApp GmbH, 2020]; it can be accessed from the AGON's main screen by tapping on top-right button. We designed this feature to allow users to see the progress in steps they have made through time. Therefore, AGON displays two columns one indicates the steps counts, and the other shows the date when those steps were taken (see Figure 2.2c). This feature differs from an *activity feedback* feature, which provides statistics, scores, or other information on the user's performance. We offered users to examine their collected data for one year back. Showing past steps allows the users to reflect on the temporal component of the walking activity. Looking at the events from specific days, users can relate success or failures to meet the steps-goal to environmental factors that might have influenced their activity (e.g., peaks of stress, work deadlines, menstrual cycles, family problems). By contrasting this information, users know how their circumstances might influence the walking behavior, therefore supporting the basic need of *competence*. According to the SDT, receiving information that is useful and allowing individuals to learn and improve contributes to the support of their *competence* BPNs [Rigby and Ryan, 2011].

We designed the *peers comparison* feature to support the basic psychological need of *relatedness*. This feature was in parts motivated by the leader boards from STRAVA [Strava Inc., 2019] and FREELETICS [Freeletics GmbH, 2019], both renowned fitness apps. The *peers comparison* can be observed from the AGON's main screen, right below the daily step counter label (see Figure 2.2d). We designed this feature by creating a two-column list. The left column displays participants' names, and the right column participants' daily steps sum. In this way, users can compare their performance with other participants who participate in the same activity over time. However, in contrast to STRAVA and FREELETICS, we adopted a privacy by design approach [Cavoukian, 2013] and partially anonymized the participants' data by displaying only their first name and step counts. Besides, the *step-counts feature for peers* does not include either icons or profile images of the participants. The actual step-counts data was gathered from the APPLE'S HEALTHKIT app database for which participants granted access after installing the AGON app. Through a peer comparison, users can evaluate the impact of their actions (i.e., walking daily) on others and feel more effective and related to other participants [Ferguson and Olson, 2014], supporting the *relatedness* BPN. We now discuss our research methodology.

2.5 Methodology

We present a pilot study where we employed a qualitative methodology for two purposes: First, this method allowed us to check our app features design before deploying the following quantitative study (see chapter 3). Second, it served to understand the participants' needs, behaviors and to evaluate the situated use of technology [Blandford et al., 2016]. In this pilot study, we tested the research app design by capturing participants' perceptions. These perceptions allowed us to understand if the design of the proposed features display support to the BPNs. Our methodology was inspired by research from Peng et al. [Peng et al., 2016]. These authors studied smartphone owners' perceptions of mobile health apps by conducting interviews and identifying themes. Given our study's object, namely behavior change apps for fitness, we opted for a field study spanning 4 weeks as we deemed users would require weeks to adjust to the new routines and reflect on their experience.

2.5.1 Participants and Recruitment

The participants sample included 49 students, age ranged from 18 to 30 years ($M = 22.2$, $SD = 3.1$). Of these participants, 30 were women (61.2%). Participants were compensated with the equivalent of 25 USD for their participation, regardless if they completed the study. We captured demographic data (i.e., name, gender, age) through a screener that also served to check whether respondents qualified for the study. We recruited participants through flyers placed on the university campus and social media sites. The flyers contained a basic explanation of the study and an email address to write to sign up. If they replied to the study call, the first author, acting as a recruiter, contacted them to screen them over the phone. If they qualified, the recruiter provided instructions on how to download and install the app and the following steps of the research.

We excluded participants who: (1) were younger than 18 years old (as we did not have resources to collect approval from the legal guardians or parents of minors); (2) did not possess an IPHONE 5 or newer model (given that our technological intervention was developed exclusively for IPHONE); (3) expected to not have Internet access for more than 7 consecutive days during the study; (4) could foresee instances in which they would be unable to undertake physical activity for more than 7 consecutive days; participants needed to be able to use the app for at least a week to guarantee they were exposed to the app features. Seventy-seven individuals replied to the study call. The first author contacted them to schedule a phone screen. Of 77, 8 did not answer the invitation email. The screening of the remaining 69 individuals for participation in the study found that 12 did not qualify for the study for one of these reasons: (1) they did not have an iPhone 5 or newer (7 respondents); (2) did not reach the end of the screener (5 respondents). Therefore, 57 participants began the study, and during the 1-month deployment we registered 8 dropouts. The final number of participants of the study was N=49.[2]. We informed participants that their involvement was voluntary, that they could withdraw from the study at any time, and that anonymity was guaranteed. The Ethics in Research Committee of our university approved the research protocol.

2.5.2 Study Procedure

We conducted a 1-month field study in a mid-size city in Central America. This period allowed participants –considering AGON minimal set of features– to be exposed, understand, and react to the three app features (*goal-setting*, *history*, and *peers comparison*). Our study consisted of three phases.

Phase 1: On boarding and Setup

Participants signed the consent form online and received setup instructions onto their emails (i.e., download the research app from APPLE APP STORE, grant access to their HEALTHKIT data, carry their phones, and use the app for four weeks). This phase lasted 4 days to give participants enough time to follow the setup instructions.

Phase 2: Interacting with AGON app

In this phase, we collected participants' step-counts. Every Monday, the app suggested a weekly and daily step goal, based on the steps average from the previous week.

[2]See the flowchart describing the recruitment process in the supplementary material: $https://osf.io/rb43c/?view_only = a7c5a118b235410abed4495b5aa91ab6$

If the app did not obtain steps from the previous week, stemming from the prior research [Marshall et al., 2009], it suggested 5K steps or approximately 30 minutes of walking each day to start with. Participants had to complete a diary through a different online platform by the end of the day. It aimed to record users' perceptions about AGON app.

Phase 3: Interview

At the end of the four weeks, we invited 15 participants to an interview session based on the frequency of use of AGON as reported in their diary entries. We did this because we aimed to collect reflections and reactions on the app from active and inactive participants. More specifically, we recruited 9 "power users" (i.e., who used the app every day) and 6 participants who were not particularly active with the app, as expressed in their diary entries. Collecting insight perceptions from active participants allowed us to understand better the app feature characteristics that made participants feel more engaged with the app. Similarly, by interviewing participants who interacted little with AGON, we captured their perceptions concerning the disengagement they had towards the app. Engagement is relevant in application-mediated studies where participants must be exposed to the mobile system's characteristics and functionalities to capture meaningful insights. For the semi-structured interviews, we designed an interview's script unpacking our *RQ*, aiming to elicit the participants' perceptions of the app's supporting features regarding their physical activity. The first two authors conducted the interviews.

The interview protocol [3] was as follows: We explained the purpose of the interview (2 minutes). We followed the interview script to capture their perceptions about the AGON app features. We inquired about their motivation to use the app and suggestions to improve it (25-40 minutes). While one author conducted the interview, the other served as a note-taker. We used a voice recorder to capture the interviews. Our dataset consists of 6 hours and 46 minutes of interview records. Finally, to aid the data analysis, we transcribed the recordings verbatim. The interviews were conducted in Spanish, the native language of the participants, and the two researchers who performed the interviews.

[3] See the interview protocol on OSF: https://osf.io/rb43c/

2.5.3 Data Analysis

Our data analysis drew from two sources of insights: the participants' diary entries which covered all 49 participants, and the semi-structured interviews conducted with a sub-sample of 15 participants. We conducted data analysis in Spanish. At the end of the analysis, the quotes taken directly from the data were translated from Spanish to English (by the first author, a native Spanish speaker) to keep the meaning as close as possible to the original.

At the inception, two researchers engaged in an affinity diagram process to analyze the interviews [Beyer and Holtzblatt, 2016, Hartson and Pyla, 2012]. This method is used to organize large amounts of unstructured qualitative data, such as participants' interviews, and has been extensively applied in HCI studies [Woodward et al., 2018, Mayer et al., 2018, Harboe and Huang, 2015, Lucero, 2015, Fedosov, 2020]. As the first step, two researchers created post-it notes independently from each other. These notes were distilled from the audio recordings of the 15 interviews and diary entries. The notes included comments, ideas, or quotes that caught the researchers' attention. Then, each researcher placed their notes on a separate wall to read the notes of each other.

Next, we spend approximately 8 hours conceptualizing categories and subcategories in an iterative, interpretative and synthesized analysis process. We cluster the post-it notes by their semantic affinity (common patterns) and aligned them with the research question. We repeated this process until we reached a consensus on the categories. To ensure the result's methodological accuracy, we used a triangulation strategy [Streubert and Carpenter, 2011], where two researchers with different points of view analyzed the data (inter-subjective). As a result, in about 4 hours, we compiled a document with the categories (i.e., themes) and representative quotes from our participants. We also held meetings with researchers outside of the project to challenge our assumptions and corroborate the themes.

Next, we present the main themes or categories we identified from our data corpus analysis and support them with quotes from the participants. The identified themes are not orthogonal; they describe interesting characterizations of our participants' experiences with AGON app. In the remainder of the paper, we use pseudonyms to describe study participants.

2.6 Findings

Our study reveals that the goal-setting feature supported the need for *autonomy* because participants perceived the decision to set the step goal their own. This feature also contributed to motivating participants to achieve the goal. The steps-history supported the need for *competence* because participants reflected on their step progress through time. The step-counts for peers feature partially supported the need for *relatedness* because participants reported feelings of companionship, curiosity, and at times competitiveness against self and the others, but not feelings of being connected to others.

2.6.1 Autonomy Support Feature

Drawing on the perceptions and the use of the *autonomy* support feature, we identified two empirical categories from our data corpus. Namely, (1) the perception of setting a goal increment; and (2) the emotions experienced through maintaining an individual steps goal.

Perception of step goal

Every Monday, AGON app proposed a step goal to participants. Gigi, 26 emphasized the value of having a choice on the number of steps the app suggested to increase: *"I felt control when accepting or denying [this recommendation]. I was the one deciding if I wanted [it] or not. However, in a sense, [the app] was imposing me the number, but because I could at least say if I wanted to go for that number or not, I had somewhat decision power."* Giss, 30 offered a similar perception: *"I like to have control of my daily and long term step goals."* Similarly, Lily, 20 expressed her feeling of choice when deciding on her step goal: *"I feel I was deciding [on the number of steps], and it was my own amount. I never felt pushed to walk."*. Also, Faby, 20 explained how he decided not to increase the daily goal: *"This number is too much and I can't. When I realized I couldn't make it I told the app I can't increase 5%"*

Besides, participants perceived the step goal as a form of personalized goal they were striving to meet by the end of the week. For example, Gigi discussed: *"The app adjusted the step goal. If I walked less steps during the week, it didn't tell me: "This was last* *week's goal, now you need to do more", instead it adjusted the goal to myself."*. Furthermore, Mar, 22 commented on a realistic, achievable step goal AGON app set for her: *"I like [that] the app suggests a more realistic step goal based on the previous week steps average."*. Mary, 19 added to that endorsing privacy-aware architecture of the app: *"I appreciated [that] the app used the data that was stored in my phone's database and used it to calculate a reasonable increase in my step counts."*

Emotional spectrum of step goal perceptions

This category relates to the feelings evoked by our participants when attempting to maintain a daily step goal. Feelings varied between positive affect, interest in the activity, and feelings of challenge. For example, Addy, 20 expressed her feeling of contentment produced by having a step goal: *"I feel very well, because I have a daily step goal, it pushes me to try to reach it."* Anny, 21, when speaking about the step goal feature elaborated on her personal experience with AGON app: *"I felt interested, because [the app] shows you the weekly step goal and at the same time the number of steps I should do every day to achieve it."* Setting a daily step goal pushed our participants to take actions to reach it, for example, Gigi, 26 explained: *"The weekly step goal motivated me to do more than what I was [usually] doing.*

Some other participants felt they were participating in a challenge, which motivated them to initiate physical activities. For example, while reflecting on using the app Cara, 22 said: *"Even though I did not walk much, I feel it is an internal competition like a challenge."* The feeling of effort was also commonly expressed from our participants, Gigi elaborated: *"I try to use the stairs more instead of the elevator and to park my car further away so that I can register more steps. It would be good if we receive an alert when the day is about to end and we have not reached our step goal."*

Collectively, these reflections help illustrate how AGON app provoked feelings of being in control of the step goal. Also, customized manageable step-goal increments helped participants to observe the steps-objective as something reachable. Additionally, it encouraged participants to engage in more physical activities autonomously.

2.6.2 Competence Support Feature

The perception and the use of the *competence* support feature turned participants to focus on their self as well as made them reflect on their own performance and improvement while engaging in physical activity. From our data corpus, we identified the following three categories: (1) reflections on self-evaluation; (2) the feelings of empowerment; and (3) the sensations produced by acknowledging past performance.

Self-Evaluation

By self-evaluating their activity, participants could monitor their progress and reflect on their own physical activity Nicky, 19 explained: *"The purpose of the history is to compare how the step counts were at the beginning and at the end of the week."* Kary, 22 speaking about her self-reflection process, restated its value to motivate her to make an extra effort in her workout routine: *"With the history I could see if I have made some progress, if I make some effort and see the improvement, then it motivates me to exercise more."* Kary also discussed the emphasis on self-improvement as an integral part of sport and activity tracking apps, and opened up a discussion about the lack of its support. This was corroborated in the following statement: *"This history [feature] is a personal retrospective, that allows me to see the progress or regression in my exercise pattern."*. Sammy, 18 reflected on how the app made him re-think the importance of doing physical activity. He explained: *"Today I feel well using the app, it is very monotonous but good and useful. Besides, I truly felt it had pushed me to walk more. These last days, doing physical activity became something that took my mind over the course of a month. This made my life change and became a little more active. I often chose walking instead of taking the bus, and I think that's what it is about, about little changes."*.

Empowerment

A feeling of empowerment was felt by some of the participants, which encouraged them to have greater confidence in themselves. Addy, 20 when contemplating about the activity progress mentioned: *"I feel I am a strong person that can do it and I can do more"*. Also, some felt AGON app gave them the courage to do more physical activity, Dean,19 exclaimed: *"It's not about keeping me in the comfort zone, it's attaining the goal and being able to say – "I'm going for more""*.

Hedonic Aspects

On a few occasions, participants expressed some feelings of sadness when looking at the history of steps. Al, 20 clarified: *"I felt bad when I saw I had less than 3000 steps, I felt powerless knowing that I could not increase my step counts."* Similarly, Gigi, 26 explained: *"When I saw my history [feature], I got disappointed because there were so many days with 400 steps, so little! I said to myself: "How is it possible that I walked so little?" So it was sad, like depressing."* Some of our participants experienced joyful feelings while using the history feature, Kary explained: *"In the history tab, I started with 2000 steps, next week I reached 4000 and that motivated me, it is like I can do more each day. When I moved forward [in step counts] I felt joy!"*. These statements illustrate the potential of the step tracking apps or services alike to incorporate hedonic aspects of both positive and negative experiences after physical activity beyond simply visualizing dry statistical summaries.

Collectively, these reflections and reactions helped illustrate how being aware of the physical activity level pushes individuals to be more rigorous in evaluating their performance and therefore experimenting with various emotions related to their activity.

2.6.3 Relatedness Support Feature

Our data corpus yielded insights into our participants' perception of the *relatedness* support feature of the app. It revealed subtle connections that relate to their overall motivation to exercise and provoked the comparison of their own results against those of other participants. We distilled three empirical categories from our data corpus: (1) companionship; (2) comparison with others doing physical activity; as well as (3) the feelings of curiosity concerning others' activity.

Companionship

Participants felt they were not the only ones performing the walking activity because they could see the step count increments from others. For example, Kary, 22, when reflecting on the feeling generated by the peers comparison feature mentioned: *"I am not alone, when I see the list I think: "these other people are doing physical activity somewhere else [around the city]."*. Consequently, the ability to see the step counts from other participants provoked personal motivations to walk more. This was described by Anny, 22: *"It is super interesting to see the steps counts from other participants, in addition to mine. I think it is a good motivation."*

Comparison and Competition

Even though AGON app did not explicitly offer a competition functionality, participants perceived a feeling of competing against each other to score a higher number of step counts. For example, Nicky, 19 mentioned: *"I should go out and walk more, because others are walking more."* Similarly, Kary, 22 echoed this through indicating the peer pressure the app triggered: *"I found myself thinking, this person has more steps than I do, I will go to the park so that I will have more steps than he does."*. These quotes might also be indicative that the participants were acting out of guilt. However, we could not find support for this in the remainder of the data we collected.

Curiosity

Some participants revealed the urge to be constantly aware of the level of physical activity from other participants and why they have walked that particular number of steps. For example, Teb, 28 reported: *"I noticed that I walk more than other people using the app."* In turn, AGON app provoked speculations about possible activities and routines of other people: Kate, 22 explained: *"I found myself thinking, she has walked so little, what might she be doing?"*.

In sum, these insights demonstrate the importance of the social features of the app. Even incidental or unintended interactions with other people through sharing of peer statistics (e.g., in the form of leader boards) allowed our participants to interact with others by comparing, competing, and keeping their physical activity status present.

2.7 Discussion

The key to any behavior change is the development of intrinsic–or self-determined–motivation towards the target activity. Self-Determination Theory [Ryan and Deci, 2017] is a well-established and empirically validated approach to evaluate behavior-change interventions. More than four decades of empirical research has demonstrated that the basic psychological needs posited by the SDT are predictive and reliable mediators to motivation [Chatzisarantis et al., 2003, Hagger, 2009, Vansteenkiste and Ryan, 2013, Ng et al., 2012, Ryan and Deci, 2017]. SDT has already been studied and applied in the field of HCI to enable behavior change. However, the road to translate this theory into concrete design guidelines is still long, and scholars are asked to "make decisions about *which* functionalities to support and *how* to implement such functionalities." [Hekler et al., 2013] Building on recent work [Villalobos-Zúñiga and Cherubini, 2020], which mapped specific app functionalities to the SDT, we contribute a concrete design of a pedometer app, whose design originates from the theory. It provides three features specifically tailored to support the basic psychological needs of its users. As we will detail in the following subsections, our findings also contribute empirical evidence that AGON users experienced feelings of self-control, empowerment, and comparison with other participants. These findings are encouraging and will have to be further validated with quantitative research, as discussed in Sec. 2.7.4. Of course, the implementation we tested in the current study is not the only possible way to provide support for autonomy, competence, and relatedness. The discussion with our participants revealed additional avenues of design and research that we will discuss in the subsections below. Finally, we contribute recommendations for other researchers who might want to study behavior-change technology in the wild (see Sec. 2.7.5).

2.7.1 Informed and Personalized Choices Supports Autonomy

While reviewing the perceptions derived from using the goal-setting feature of our app, we noticed that participants elicited feelings of *owning* the decision about setting their weekly step-goal. Even though the step-goal increase suggestion came from AGON, participants noticed they decide to pursue the goal or not depending on how much effort they foresee it implies. These feelings are aligned to the SDT *autonomy* definition, which states the need of an individual to experience ownership of their actions [Ryan and Deci, 2017, p. 86]. Also, this resonates with the characteristics of the *goal-setting* feature described in chapter 1, indicating that this feature contributes to an *internal perceived locus of causality*

Other perceptions captured from using the AGON's goal-setting feature show that when participants had information about their level of physical activity (e.g., average daily step-counts) to make a decision, they displayed effort and determination to achieve the objective. According to what the SDT postulates, feelings of effort and determination are expected to arise when individuals make good choices after thoughtfully considering the relevant options and information [Ryan and Deci, 2017, p. 462]. Further, this observation resonates with previous studies on physical activity in sport psychology, where researchers established that when goals are set autonomously, they positively predict effort, and consequently, goal attainment [Smith et al., 2007, Smith et al., 2011]. Our study extends this research by incorporating tailored and specific goals (i.e., "walk 7550 steps this week") and not generic open goals (i.e., "improve your upper body strength") like Smith et al. [Smith et al., 2007, Smith et al., 2011] do; differentiating in the ability to set goal metrics.

Another exciting aspect that stood from our findings is the positive reception of a *personalized* weekly step goal. Participants felt more inclined to accept a weekly step goal increase since this increment was tailored to their previous week's performance. The suggested increase of 5% was perceived as achievable –optimal– by our participants. The concept of optimal challenge was already discussed within the SDT [Deci, 1975], and it has been empirically tested in immersive games [Qin et al., 2010, Rigby and Ryan, 2011]. Similarly, we see the value of applying this concept to behavioral change scenarios, such as those featured in modern fitness apps. This approach explains why participants reported they accepted the weekly goal increases.

In sum, these observations extend prior research in the following ways: (1) we provide an initial attempt towards translating the SDT *autonomy* construct to AGON's goal-setting feature; (2) previous research allowed individuals to self-set their goals without activity-related information to make a knowledgeable decision (e.g., [Consolvo et al., 2009, Munson and Consolvo, 2012, Gouveia et al., 2015]), or through a self-reported baseline (e.g, [Hartzler et al., 2016]), we extend this design space by improving the individual's ability to set goals by making autonomous and informed decisions; (3) the positive response of our participants to the suggested weekly step goal increase is significant because it outlines the potential of personalized goal increments and how this can engage participants to achieve the step goals they have committed to.

Therefore, these findings suggest that app designers might want to provide users with information (e.g., how much effort is needed to achieve a goal, personal activity performance) to contribute to more autonomous decisions and foster goal attainment. Furthermore, designers should suggest adaptable goals to each individual's ability level [Baretta et al., 2019].

2.7.2 Performance Monitoring Supports Competence

While reviewing the perceptions derived from using our app's history feature, we noticed that participants elicited feelings of progress and empowerment concerning the walking activity. These feelings distill from comparing their performance at the beginning of the week with the end of the week. This resonates with chapter 1, where we described the sense of effectiveness produced by implementing a functionality with *history* taxonomy category characteristics. We also noticed that in some occasions these feelings were positive (e.g., feeling joy due to making progress toward the goal), while in other cases expressing negative feelings (e.g., the disappointment caused by not progressing towards the goal). The former feelings are aligned to the SDT *competence* definition, which states the need of feeling effective in one's interactions [Ryan and Deci, 2017, p. 86]. However, concerning the latter feelings, we believe *Agon*'s design could be improved. In moments where despite not making progress towards the exercise goal, the app could provide information that may encourage alternative ways to fulfill the objectives (e.g., display a message with the text: "Keep trying. There are 3 more days to go!").

Another highlight in our findings is the eudaimonic effect [Mekler and Hornbæk, 2016] of striving towards one's personal best evoked by the app's use. By observing their daily-step records and history, participants aimed for a constant need for fulfillment and self-improvement.

Other perceptions captured from using the AGON's history feature show an increase of participants' reflection on their physical activity and, in turn, this reflection contributed to motivating them to reach their step-goal. This self-reflection produced by the history feature relates to previous research, demonstrating that having a display that supports activity-awareness may lead to positive outcomes such as keeping the participants' physical activity levels and engagement throughout research interventions [Consolvo et al., 2008]. AGON app extends this display approach by using a two-column list with past weeks and steps, allowing time-progress comparison; this is limited in the garden metaphor of Consolvo et al. [Consolvo et al., 2009]. Participant's reflections on their own –effective– performance and the consequent positive effect on their sense of competence were expected as the SDT postulates [Ryan and Deci, 2017, p. 154]. This finding is relevant because it supports the importance of providing personal activity-related information to individuals to increase their sense of competence.

In sum, these observations extend prior research in the following ways: (1) we provide an initial attempt towards translating the SDT *competence* construct to AGON's history feature; (2) previous research provides users with progress bars (e.g., [Sankaran et al., 2016]) showing a single progress indicator without the option to compare to previous scores, some others use line or bar graphs (e.g., [Harries et al., 2016, Munson and Consolvo, 2012, Oyibo et al., 2019]) allowing comparison between days of the same week, some others allow score comparison with just the competition leader (e.g., [Gouveia et al., 2015]). We contribute to this design space by allowing users of AGON to visualize and reflect on a broad time frame (one year). In our design, we enabled this exploration by providing a scrollable two-column text, displaying the list with past days and their related steps. (3) our participant's positive response to the history feature is significant because it outlines the power and intention that rises from participants when feeling capable of achieving the step goals they have committed to.

Therefore, these findings suggest that fitness app designers, who are interested in supporting users' attainment to their fitness goals, should: i. provide elements that increase the awareness of the activity (e.g., steps-history), and ii. cater to positive feelings (e.g., encouraging messages when bad performance happen).

2.7.3 Comparison with Peers Partially Supports Relatedness

While reviewing the users' perceptions derived from using the peers comparison feature of our app, we observed two types of competition behavior among our participants. Some participants compare their step count with that of other participants through the peers comparison feature (see Figure 2.2d). The SDT scholars define this behavior as a direct competition, which occurs when players struggle against each other to maximize their success. This finding resonates with the description of the *peers comparison* feature introduced in chapter 1 which states that by having this comparison, users can evaluate the impact of their actions on others and therefore feel more effective. In contrast, other participants perceived the daily step goal as a competition with themselves. The SDT considers this behavior an indirect competition [Ryan and Deci, 2017, p.488], which occurs when people compete against themselves when performing better than what they have done previously [Ryan and Deci, 2017]. Developing mechanisms that foster indirect competition can be significant to fitness app designers interested in skill-building and performance, both reflected in the individual's adoption of mastery goals ([Ryan and Deci, 2017], op. cit.).

Through a comparison with their peer, users can assess the impact of their actions on others and feel more effective [Ferguson and Olson, 2014]. Therefore, this feature supports the BPN of relatedness. Additionally, 'Peer Comparison' contributes to the SDT basic need of competence because, through this feature, individuals can assess their level of efficacy and mastery toward the specific activity [Ryan and Deci, 2017, p.97]

Our relatedness insights showed a partial support of the peers comparison feature on the corresponding BPNs. More specifically, participants felt companionship, meaning they were not the only ones in the study doing physical activity, which in turn increased their motivation to be more physically active. Because of this perception of having a companion, they compared and competed with others (as previously mentioned). However, participants did not perceive themselves *connected* to other participants. This insight leads us to reflect on why the peers comparison was not effective in developing a feeling of connectedness with others. We believe that the relatedness-support feature did not produce an anticipated effect due to two factors. First, participants did not know whom the other people on the list were, producing a lack of empathy and connection with others. Second, we did not place participants in groups comprising people with same physical activity levels. An alternative group assignment –which grouped participants who had similar performance– might have increased the participants' self-efficacy evaluation. This insight opens up an opportunity to explore other designs for relatedness-support features. For example, placing individuals in groups or circles (which might develop a sense of team) and developing a sense of competition against other groups, instead of competing at the individual level. Therefore, these findings suggest that fitness app designers who are interested in supporting meaningful connections between users should: i. provide elements that increase the perception that more people are doing the same activity (e.g., a list of users pseudonyms with the option to display user information like name, age, hobbies.), and ii. foster the feeling of between-user connection (e.g., create competitions between familiar with each other and with similar physical activity levels).

2.7.4 Future Research and Design Directions

The subjective accounts collected in this research demonstrate that AGON provided support to the autonomy, competence, and relatedness of our study participants. *Would this support be sufficient to motivate the participants to increase the physical activity* *in their routine?* Unfortunately, the current research is unable to answer this question comprehensively. In this study, we learned that physical activity levels might drastically change from one week to the next due to seasonal effects (e.g., holidays), weather conditions, or other schedule constraints. We also discovered a large variability of behaviors associated with physical activity levels in the sample population: some people might regularly train, while others might walk or run only sporadically. It is necessary to conduct longitudinal studies spanning multiple months and involving hundreds of participants to demonstrate the effects of behavior-change technology. This experimental design might allow researchers to average out participants and seasonal effects and demonstrate the long-term impact of behavior-change technology.

Furthermore, in this study, we tested the three features supporting the BPNs concurrently. However, it would be valuable to understand whether these features need to be available simultaneously to yield benefits to the app users. *Would offering only one (or two) features provide incremental support to users?* A full factorial design, assigning different combinations of features to other user groups, would need testing to answer this question. Such a study can produce quantified observations to relate each feature –or combination of features– to physical activity levels.

From a design perspective, the interviews we conducted suggested exciting avenues to explore further. Several study participants mentioned having a hard time agreeing to the increases in physical activity recommended by AGON on specific days because of personal or work commitments. In the future, designers might develop *autonomy* features that could allow users to adhere to personalized goal increments and customize the day of the week these goals are feasible. Also, our study participants mentioned that the *competence* feature supported their self-reflection. However, none of the participants said they had used this information to follow week-by-week progress, perhaps, because of an additional cognitive load required to infer this information. Furthermore, designers might highlight non-obvious trends in the data series to aid data exploration for end-users. Finally, as discussed in Sec. 2.7.3 designers might want to develop *relatedness* features that support optimal [Deci, 1975] and coopetitive [Wolf et al., 2020] challenges. We believe these design ideas could also inspire the designers of supporting technologies in other behavior-change domains (e.g., acquiring new skills, subscribing to a conscientious consumption lifestyle, lowering one's ecological footprint).

2.7.5 Barriers and Recommendations when Running Research on Behavior-Change Apps

The current study made us reflect on two critical aspects of researching technology that aims at supporting behavior change: (a) providing enough time to study participants to *adjust to the intervention*; and (b) *considering the ecosystem* upon which the deployment of interventions occur. Concerning the first aspect, our participants installed the app on different days of the first week of the study, interacting with the app at varying frequencies, and all had periods of unavailability throughout the month-long study. The analysis of the diaries revealed that in most cases participants developed: a concrete understanding of AGON only towards the end of the study. In hindsight, we might have designed a more prolonged deployment to allow participants to experience more opportunities of support provided by the app. The support provided by behavior change apps becomes meaningful only when this is provided close to the target activity the intervention is aiming to change (e.g., walking or running). Given that the target activities often occur sporadically (e.g., once or twice a week), the observation window of user studies must necessarily exceed several expected occurrences of the target activity. Therefore, we recommend researchers define the length of the user study on based on the typical frequency of the target activity in the study population. For instance, with the daily performance of target activities, a study spanning three to five weeks could be sufficient to provide exposure to the intervention. A lower frequency of occurrence of the target activity must reflect in a longer observation window.

Concerning the second aspect, we noticed that our study participants sometimes had difficulty accepting suggested increases in physical activity. Not because they did not want to, but because of schedule constraints or improper conditions (e.g., no lovely places for a walk during the lunch break). As discussed in the previous subsection, this made us consider that an improved version of the Goal Setting feature could have provided more flexibility to allow our participants to choose *when* was best to perform the extra activity. Most behavior change interventions are oblivious to the constraints existing in the users' lives and cannot personalize the recommendations to different life circumstances, unavailability, and logistic constraints. This lack of consideration for the ecosystem in intervention deployment reduces its impact. From now on, we recommend designers explore strategies to capture and model this ecosystem and allow users of behavior change apps to schedule activities around constraints.

2.7.6 Limitations

We want to acknowledge a few limitations of our study. Since, we deliberately opt-in for a qualitative research approach our findings may not and are not intended to be generalized to other domains. Our approach allowed us to develop a rich and descriptive account of participant perceptions of the AGON app. A future longitudinal study could capture objective and more detailed accounts of an intervention's effectiveness based on the proposed design of AGON.

Furthermore, our study was limited to one particular embodiment of the BPN supportive features. Studying (and comparing) alternative designs might reveal specificities of the features that we could not capture in this study.

Finally, given that we deployed our recruitment fliers on a university campus, we recruited participants in their 20s and 30s. We purposefully aimed for young adults since they are often more active and more involved in sharing personal experiences online [Acquisti and Gross, 2006], and generally have emerged as rapid adopters of digital technology [Dee Dickerson and Gentry, 1983]. However, this strategy undoubtedly limits the generalizability of our findings to older users. Future research should recruit a more heterogeneous sample from various cultural, geographic (e.g., suburban, rural), and demographic contexts.

2.8 Conclusion

We explored the perceptions and reflections of 49 individuals from the 4-week field deployment of a Self-Determination Theory-based mobile app. In this paper, we made two primary contributions. First, we present a novel design of a fitness app distilled from the taxonomy of app features based on the SDT. Second, we presented insights on how the perceptions of the app features supported *autonomy*, *competence*, and *relatedness* needs. We hope that our study will inform and inspire future research in personal and persuasive computing that looks at behavior change practices and interventions within the context of physical activities and beyond.

2.9 Acknowledgments

We want to thank Mercedes Alpízar Soto for her timely valuable contributions during this work's app design discussions and writing process. Finally, we would like to thank the anonymous reviewers for their precious comments that strengthen the quality of this research.

2.10 Conclusion to the Chapter

In this chapter, we learned that the design of the physical activity app features we implemented produced perceptions on participants that indicated that the features correctly mapped the BPNs. These perceptions correspond to our particular feature implementation; therefore, testing other feature implementations might provide a broader understanding of supporting the BPNs. We must remark that having perceptions that mapped the BPNs is crucial to provoke behavior change, but it is not sufficient. There is a difference between perceiving to act in a particular manner and behaving in that way. Therefore, in the following chapter, we collect objective behavioral data to test whether the perceived BPNs app features provoke a behavior change.

Second Chapter References

[Acquisti and Gross, 2006] Acquisti, A. and Gross, R. (2006). Imagined communities: Awareness, information sharing, and privacy on the facebook. In *International workshop on privacy enhancing technologies*, pages 36–58, Berlin, Heidelberg. Springer, Springer Berlin Heidelberg.

[Altmeyer et al., 2018] Altmeyer, M., Lessel, P., Sander, T., and Krüger, A. (2018). Extending a gamified mobile app with a public display to encourage walking. In *Proceedings of the 22Nd International Academic Mindtrek Conference*, Mindtrek '18, pages 20–29, New York, NY, USA. ACM.

[Baretta et al., 2019] Baretta, D., Bondaronek, P., Direito, A., and Steca, P. (2019). Implementation of the goal-setting components in popular physical activity apps: Review and content analysis. *Digital Health*, 5:1–10.

[Beyer and Holtzblatt, 2016] Beyer, H. and Holtzblatt, K. (2016). *Contextual Design*. Morgan Kaufmann, 340 Pine Street, 6th Floor San Francisco, CA 94104, USA.

[Blandford et al., 2016] Blandford, A., Furniss, D., and Makri, S. (2016). Qualitative hci research: Going behind the scenes. *Synthesis lectures on human-centered informatics*, 9(1):1–115.

[Cavoukian, 2013] Cavoukian, A. (2013). Privacy by design: leadership, methods, and results. In *European Data Protection: Coming of Age*, pages 175–202. Springer, Dordrecht, The Netherlands.

[Chatzisarantis et al., 2003] Chatzisarantis, N. L. D., Hagger, M. S., Biddle, S. J. H., Smith, B., and Wang, J. C. K. (2003). A Meta-Analysis of Perceived Locus of Causality in Exercise, Sport, and Physical Education Contexts. *Journal of Sport and Exercise Psychology*, 25(3):284–306.

[Colusso et al., 2016] Colusso, L., Hsieh, G., and Munson, S. A. (2016). Designing closeness to increase gamers' performance. In *Conference on Human Factors in*

Computing Systems - Proceedings, pages 3020–3024, New York, NY, USA. ACM.

[Consolvo et al., 2006] Consolvo, S., Everitt, K., Smith, I., and Landay, J. (2006). Design requirements for technologies that encourage physical activity. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, pages 457–466, New York, NY, USA. ACM.

[Consolvo et al., 2008] Consolvo, S., Libby, R., Smith, I., Landay, J. A., McDonald, D. W., Toscos, T., Chen, M. Y., Froehlich, J., Harrison, B., Klasnja, P., LaMarca, A., and LeGrand, L. (2008). Activity sensing in the wild: a field trial of ubifit garden. In *Proceeding of CHI'08*, pages 1797 – 1806, New York, USA. ACM Press.

[Consolvo et al., 2009] Consolvo, S., McDonald, D. W., and Landay, J. A. (2009). Theory-driven design strategies for technologies that support behavior change in everyday life. In *Conference on Human Factors in Computing Systems - Proceedings*, number May, pages 405–414, New York, NY, USA. Association for Computing Machinery.

[Covolo et al., 2017] Covolo, L., Ceretti, E., Moneda, M., Castaldi, S., and Gelatti, U. (2017). Does evidence support the use of mobile phone apps as a driver for promoting healthy lifestyles from a public health perspective? A systematic review of Randomized Control Trials. *Patient Education and Counseling*, 100(12):2231–2243.

[DeCharms, 1968] DeCharms, R. (1968). *Personal causation: the internal affective determinants of behavior*. Academic Press, Cambridge, Massachusetts, US.

[Deci and Ryan, 1985] Deci, E. and Ryan, R. M. (1985). *Intrinsic Motivation and Self-Determination in Human Behavior*. Springer US, New York, USA.

[Deci, 1975] Deci, E. L. (1975). *Intrinsic Motivation*. Bibliographie pp.295. Plenum Publishing Company Limited, New York, USA.

[Deci and Ryan, 1980] Deci, E. L. and Ryan, R. M. (1980). The empirical exploration of intrinsic motivational processes1. In *Advances in experimental social psychology*, volume 13, pages 39–80. Elsevier, Atlanta, USA.

[Deci and Ryan, 2008] Deci, E. L. and Ryan, R. M. (2008). Self-determination theory: A macrotheory of human motivation, development, and health. *Canadian Psychology/Psychologie canadienne*, 49(3):182–185.

[Deci et al., 1996] Deci, E. L., Ryan, R. M., and Williams, G. C. (1996). Need satisfaction and the self-regulation of learning. *Learning and individual differences*, 8(3):165–183.

[Dee Dickerson and Gentry, 1983] Dee Dickerson, M. and Gentry, J. W. (1983). Characteristics of adopters and non-adopters of home computers. *Journal of Consumer research*, 10(2):225–235.

[Edwards et al., 2016] Edwards, E. A., Lumsden, J., Rivas, C., Steed, L., Edwards, L. A., Thiyagarajan, A., Sohanpal, R., Caton, H., Griffiths, C. J., Munafò, M. R., Taylor, S., and Walton, R. T. (2016). Gamification for health promotion: systematic review of behaviour change techniques in smartphone apps. *BMJ Open*, 6(10):6–0.

[Endomondo, 2020] Endomondo (2020). The personal trainer in your pocket.

[Epstein et al., 2015] Epstein, D. A., Jacobson, B. H., Bales, E., McDonald, D. W., and Munson, S. A. (2015). From "nobody cares" to "way to go!": A design framework for social sharing in personal informatics. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '15, page 1622–1636, New York, NY, USA. Association for Computing Machinery.

[Fedosov, 2020] Fedosov, A. (2020). *Supporting the Design of Technology-Mediated Sharing Practices*. Carl Grossmann Verlag, Berlin.

[Ferguson and Olson, 2014] Ferguson, C. J. and Olson, C. K. (2014). Video Game Violence Use Among "Vulnerable" Populations: The Impact of Violent Games on Delinquency and Bullying Among Children with Clinically Elevated Depression or Attention Deficit Symptoms. *Journal of Youth and Adolescence*, 43(1):127–136.

[Ferrara, 2013] Ferrara, J. (2013). Games for persuasion: Argumentation, procedurality, and the lie of gamification. *Games and Culture*, 8(4):289–304.

[Fitbit Inc., 2020] Fitbit Inc. (2020). FitBit App: the fitness app for everyone.

[Fox and Duggan, 2012] Fox, S. and Duggan, M. (2012). Mobile Health 2012. Technical report, Pew Research Center.

[Freeletics GmbH, 2019] Freeletics GmbH (2019). Freeletics: Train with the best digital coach. Last accessed August 22, 2019.

[Gouveia et al., 2015] Gouveia, R., Karapanos, E., and Hassenzahl, M. (2015). How do we engage with activity trackers? a longitudinal study of habito. In *UbiComp 2015 - Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 1305–1316, New York, NY, USA. ACM.

[Guyton and Hall, 2015] Guyton, A. and Hall, J. E. (2015). *Textbook of Medical Physiology*. Saunders, New York, NY, US, 13th ed. edition.

[Hagger, 2009] Hagger, M. S. (2009). Theoretical integration in health psychology: Unifying ideas andcomplementary explanations. *British Journal of Health Psychology*, 14:189–194.

[Harboe and Huang, 2015] Harboe, G. and Huang, E. M. (2015). Real-world affinity diagramming practices: Bridging the paper-digital gap. *Conference on Human Factors in Computing Systems - Proceedings*, 2015-April:95–104.

[Harries et al., 2016] Harries, T., Eslambolchilar, P., Rettie, R., Stride, C., Walton, S., and Van Woerden, H. C. (2016). Effectiveness of a smartphone app in increasing physical activity amongst male adults: A randomised controlled trial. *BMC Public Health*, 16(1):1–10.

[Hartson and Pyla, 2012] Hartson, R. and Pyla, P. (2012). *The UX Book: Process and Guidelines for Ensuring a Quality User Experience*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1st edition.

[Hartzler et al., 2016] Hartzler, A. L., Venkatakrishnan, A., Mohan, S., Silva, M., Lozano, P., Ralston, J. D., Ludman, E., Rosenberg, D., Newton, K. M., Nelson, L., and Pirolli, P. (2016). Acceptability of a team-based mobile health (mHealth) application for lifestyle self-management in individuals with chronic illnesses. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2016-Octob(November):3277–3281.

[Hekler et al., 2013] 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 CHI'13*, page 3307, New York, New York, USA. ACM Press.

[Jebelle and Burrows, 2019] Jebelle, H. and Burrows, T. (2019). A weight loss app may be a risky way to address obesity in children.

[Lucero, 2015] Lucero, A. (2015). Using Affinity Diagrams to Evaluate Interactive Prototypes. In *Proceedings of the 15th IFIP TC 13 International Conference on Human-Computer Interaction (INTERACT 2015), Part II*, volume 9297 of *Lecture Notes in Computer Science*, pages 231–248, Gewerbestrasse 11, 6330 Cham, Switzerland. Springer, Cham.

[Marshall et al., 2009] Marshall, S. J., Levy, S. S., Tudor-Locke, C. E., Kolkhorst, F. W., Wooten, K. M., Ji, M., Macera, C. A., and Ainsworth, B. E. (2009). Translating Physical Activity Recommendations into a Pedometer-Based Step Goal. 3000 Steps in 30 Minutes. *American Journal of Preventive Medicine*, 36(5):410–415.

[Mayer et al., 2018] Mayer, S., Lischke, L., Wozniak, P. W., and Henze, N. (2018). Evaluating the Disruptiveness of Mobile Interactions. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, pages 1–14, New York, NY, USA. Association for Computing Machinery.

[Mekler and Hornbæk, 2016] Mekler, E. D. and Hornbæk, K. (2016). Momentary pleasure or lasting meaning? distinguishing eudaimonic and hedonic user experiences. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, page 4509–4520, New York, NY, USA. Association for Computing Machinery.

[Middelweerd et al., 2014] Middelweerd, A., Mollee, J., van der Wal, C., Brug, J., and te Velde, S. (2014). Apps to promote physical activity among adults: a review and content analysis. *International Journal of Behavioral Nutrition and Physical Activity*, 11(97):189–194.

[Molich and Nielsen, 1990] Molich, R. and Nielsen, J. (1990). Improving a Human-Computer Dialogue. *Communications of the ACM*, 33(3):338–348.

[Munson and Consolvo, 2012] Munson, S. a. and Consolvo, S. (2012). and Sharing to Motivate Physical Activity. *6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)*, pages 25–32.

[Ng et al., 2012] Ng, J. Y. Y., Ntoumanis, N., Thøgersen-Ntoumani, C., Deci, E. L., Ryan, R. M., Duda, J. L., and Williams, G. C. (2012). Self-Determination Theory Applied to Health Contexts: A Meta-Analysis. *Perspectives on Psychological Science*, 7(4):325–340.

[Oyibo et al., 2019] Oyibo, K., Adaji, I., Olagunju, A. H., Deters, R., Olabenjo, B., and Vassileva, J. (2019). Ben'fit: Design, implementation and evaluation of a culture-tailored fitness app. In *ACM UMAP 2019 Adjunct - Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization*, pages 161–166, New York, NY, USA. ACM.

[Patall et al., 2010] Patall, E. A., Cooper, H., and Wynn, S. R. (2010). The effectiveness and relative importance of choice in the classroom. *Journal of Educational Psychology*, 102(4):896–915.

[Peng et al., 2016] Peng, W., Kanthawala, S., Yuan, S., and Hussain, S. A. (2016). A qualitative study of user perceptions of mobile health apps. *BMC Public Health*, 16(1):1–11.

[Peters et al., 2018] Peters, D., Calvo, R. A., and Ryan, R. M. (2018). Designing for motivation, engagement and wellbeing in digital experience. *Frontiers in Psychology*, 9(MAY):1–15.

[Qin et al., 2010] Qin, H., Rau, P.-L. P., and Salvendy, G. (2010). Effects of different scenarios of game difficulty on player immersion. *Interacting with Computers*, 22(3):230–239.

[Rigby and Ryan, 2011] Rigby, S. and Ryan, R. M. (2011). *Glued to games: How video games draw us in and hold us spellbound.* New directions in media. Praeger/ABC-CLIO, Santa Barbara, CA, US.

[Ryan and Connell, 1989] Ryan, R. M. and Connell, J. P. (1989). Perceived Locus of Causality and Internalizatoin: Examining Reasons for Acting in Two Domains. *Journal of Personality and Social Psychology*, 57(5):749–761.

[Ryan et al., 1985] Ryan, R. M., Connell, J. P., and Deci, E. L. (1985). A motivational analysis of self-determination and self-regulation in education. *Research on motivation in education: The classroom milieu*, 2:13–51.

[Ryan and Deci, 2017] Ryan, R. M. and Deci, E. L. (2017). *Self-Determination Theory Basic Psychological Needs in Motivation, Development and Wellness*. The Guilford Press, New York, USA.

[Ryan et al., 1996] Ryan, R. M., Sheldon, K. M., Kasser, T., and Deci, E. L. (1996). All goals are not created equal: An organismic perspective on the nature of goals and their regulation. In Gollwitzer, P. M. and Bargh, J. A., editors, *The psychology of action: Linking cognition and motivation to behavior.*, pages 7–26. Guilford Press, New York, US.

[Sankaran et al., 2016] Sankaran, S., Frederix, I., Haesen, M., Dendale, P., Luyten, K., and Coninx, K. (2016). A grounded approach for applying behavior change techniques in mobile cardiac tele-rehabilitation. *ACM International Conference Proceeding Series*, 29-June-20:1–8.

[Schoeppe et al., 2016] Schoeppe, S., Alley, S., Van Lippevelde, W., Bray, N. A., Williams, S. L., Duncan, M. J., and Vandelanotte, C. (2016). Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: A systematic review. *International Journal of Behavioral Nutrition and Physical Activity*, 13(1):3–26.

[Skarecki, 2015] Skarecki, B. (2015). Running Apps Have Hurt My Progress, Not Helped It.

[Smith et al., 2007] Smith, A., Ntoumanis, N., and Duda, J. (2007). Goal striving, goal attainment, and well-being: Adapting and testing the self-concordance model in sport. *Journal of Sport and Exercise Psychology*, 29(6):763–782.

[Smith et al., 2011] Smith, A. L., Ntoumanis, N., Duda, J. L., and Vansteenkiste, M. (2011). Goal striving, coping, and well-being: A prospective investigation of the self-concordance model in sport. *Journal of Sport and Exercise Psychology*, 33(1):124–145.

[StepsApp GmbH, 2020] StepsApp GmbH (2020). Step Counter and Walking Tracker.

[Strava Inc., 2019] Strava Inc. (2019). Strava: app for runners and cyclists. Last accessed August 22, 2019.

[Streubert and Carpenter, 2011] Streubert, H. and Carpenter, D. (2011). *Qualitative research in nursing, in Advancing the Humanistic Imperative*. Lippincott Williams & Wilkins, Philadelphia, PA, USA, 5th edition edition.

[Toscos et al., 2006] Toscos, T., Faber, A., An, S., and Gandhi, M. P. (2006). Chick Clique: Persuasive Technology to Motivate Teenage Girls to Exercise Tammy. In *Extended Abstracts of the 2006 CHI Conference on Human FActors in Computing Systems - CHI'06*, pages 1873–1878, New York, NY, USA. Association for Computing Machinery.

[Tsigilis, 2005] Tsigilis, N. (2005). The influence of intrinsic motivation on an endurance field test. *The Journal of sports medicine and physical fitness*, 45(2):213–6.

[Under Armour Inc., 2020] Under Armour Inc. (2020). MyFitnessPal: Fitness starts with what you eat.

[University of Colorado Hospital, 2003] University of Colorado Hospital (2003). Training for Cardiovascular Fitness. Technical report, University Sports Medicine.

[Vansteenkiste and Ryan, 2013] Vansteenkiste, M. and Ryan, R. M. (2013). On psychological growth and vulnerability: Basic psychological need satisfaction and need frustration as a unifying principle. *Journal of Psychotherapy Integration*, 23(3):263–280.

[Villalobos-Zúñiga and Cherubini, 2020] Villalobos-Zúñiga, G. and Cherubini, M. (2020). Apps that motivate: a taxonomy of app features based on self-determination theory. *International Journal of Human-Computer Studies*, 140(102449):102449.

[Wolf et al., 2020] Wolf, T., Jahn, S., Hammerschmidt, M., and Weiger, W. H. (2020). Competition versus cooperation: How technology-facilitated social interdependence initiates the self-improvement chain. *International Journal of Research in Marketing*.

[Woodward et al., 2018] Woodward, J., McFadden, Z., Shiver, N., Ben-Hayon, A., Yip, J. C., and Anthony, L. (2018). Using co-design to examine how children conceptualize intelligent interfaces. *Conference on Human Factors in Computing Systems - Proceedings*, 2018-April:1–14.

[Wozniak et al., 2017] Wozniak, P. W., Fedosov, A., Mencarini, E., and Knaving, K. (2017). Soil, rock, and snow: On designing for information sharing in outdoor sports. In *Proceedings of the 2017 Conference on Designing Interactive Systems*, DIS '17, pages 611–623, New York, NY, USA. ACM.

Chapter 3

Let's compete, we'll be rewarded!

"The Devil is in the Details."
— Ludwig Mies van der Rohe, 1886–1969

3.1 Introduction to the Chapter

In this chapter, we present an ongoing study where we evaluate during 5-months a physical activity app that combines multiple features supporting the Basic Psychological Needs. We assessed participants' intrinsic motivation and daily average number of steps. Then we present our preliminary results indicating a positive trend caused by the app supporting *competence* and *relatedness*.

3.2 Introduction

From the previous chapter, we understand the importance of learning about *human motivation* and designing theory-informed apps to support individuals in improving their overall health and well-being.

Grasping users' perceptions about SDT app features was a first step towards finding an effective behavior change app. This chapter extends these findings by conducting a quantitative study where we collect objective participants' behavior.

We phrase the following research question:
RQ3: What are the BPNs feature combinations that produce the most positive effects on the individuals' behavior when it comes to improving their physical activity?

We designed an app with three features, each supporting one of the Basic Psychological Needs, and tested different combinations of these features to answer this question. We captured participants' steps counts and their app interaction logs.

Our results contribute to having a more comprehensive understanding of the effects of behavior change features.

3.3 Related Work

As much of the literature from the previous chapter informs this present one. I decided to consider this section as an extension of the related work from the previous chapter by including additional literature blocks with more recent literature which I structured in the following sections: SDT Informed Physical Activity Studies, Behavior Change App Features, and User Engagement.

3.3.1 SDT Informed Physical Activity Studies

A recent review covering 170 papers from 2003 to 2019 shows few (i.e., 45) behavior change studies in which the technological artifact is informed by theory [Aldenaini et al., 2020]. Moreover, from these theory-informed studies, only 7 are based on the Self-Determination Theory. We summarize them in the following lines.

Haque et al. [Haque et al., 2017], developed a prototype informed by the Basic Psychological Needs aiming to improve physical activity behavior at the workplace. They conducted a usability test for 1-week with 26 participants and found out their app works well for physical activity promotion. In a more recent study, Haque et al. [Haque et al., 2020], conducted a 4-week study with 220 participants using the same app as in the previously mentioned study. The goal was to increase the physical activity levels of people in an office setting across 4 countries. Their results show that their app would help increase the physical activity level compared to a paper diary used by the control group. Their results also display the app supported the need for autonomy and competence. They also mentioned that compliance with the app usage was deficient, with 27 participants completing the study. In another SDT-informed study, Spruijt-Metz et al. [Spruijt-Metz et al., 2008], studied a mechanism to reduce sedentary behavior in minority girls (n=136) using a classroom animation intervention during 5 to 7 consecutive school days. They found a trend for mediation effects of intrinsic motivation, but without reaching significance. In another study, Lacroix J. et al. [Lacroix et al., 2009], deployed an intervention in which participants used an activity monitor on the body for 10 days. They aimed to explore the relationship of behavioral regulation, types of motives, and self-efficacy with actual levels of daily-life physical activity. They found out that active people experiment higher levels of self-determine behavioral regulation. Also, they experience stronger motives to be active and bear higher levels of self-efficacy for daily-life physical activity than inactive people. Continuing with the SDT-supporting studies, Cercos et al. [Cercos and Mueller, 2013], employed a shared semi-public display of physical activity data. This display was placed in a shared work space where people could see it. To capture the user's steps, researchers employed a Fitbit device. They aimed at understanding how to promote physical activity between a group of co-located people using wearable self-monitoring devices and semi-public displays. Their preliminary findings indicate that the display motivated players to use a self-monitoring device every day and facilitated new discussions between players without producing privacy issues. They also reported emotional connections with non-collocated participants.

In a different research effort, Stragier J. et al. [Stragier et al., 2015], studied the physical activity tracking shared on Twitter. Their results show that is intrinsic motivation instead of extrinsic motivation that determines a person's willingness to share physical activity via social networking sites. Other researchers [De Cocker et al., 2016], developed a theory-driven, Web-based, interactive, computer-tailored intervention aimed at reducing and interrupting sitting at work. They received personalized feedback and tips on how to reduce or interrupt workplace sitting. Their results, enforces the importance of computer tailoring for sedentary behavior and its promising utilization in public health promotion.

Finally, Altmeyer M. et al. [Altmeyer et al., 2018], created a system using fitness trackers to collect step counts and displayed them on gamified mobile app and a gamified public display. They evaluated the system during 4 weeks with 12 participants and discovered that the public display significantly increase the number of steps from the users and their motivation to walk. They attributed these findings to the significant increase in social relatedness.

In sum, all these past research has as common ground the use of the SDT to design interventions where the technological component includes trackers, displays, or mobile apps. The piece of research the most similar to the study described in this chapter is presented by Haque et al., where they designed a mobile app prototype to improve physical activity and test it during a short time with a small number of participants. We extend this research by describing how each app feature was conceived to support the Basic Psychological Needs, then implementing a mobile app and conducting a 4-month factorial in-the-wild study that teases out the effect of each app feature.

Next, to understand the association between the app features and the desired behavior, it is essential to study how past research has established a link between behavior change features and behavioral outcomes. Therefore, in the following section, we introduce behavior change app features relevant to the design scope of our research.

3.3.2 Behavior Change App Features

Goal-Setting

In a recent study Chevance et al. [Chevance et al., 2020], explored the effects of goal-setting and achievement for walking by employing a mobile app and a pedometer. In their study, participants received through an app an experimental manipulated daily step goal based on their baseline number of steps. Their results show a positive and significant relation between the goal performance difficulty and the physical activity level; and a negative and significant association with goal achievement.

Progress Monitoring

Concerning the progress monitoring features, Lloyd et al. [Lloyd et al., 2017], designed a smartphone app for people with mild cognitive impairment, aiming to promote healthy lifestyle choices. As a mechanism to measure progress, they implemented a reward system where the user is awarded stars for meeting targets on any day and trophies if the user continually wins stars. Users could accumulate gold, silver, and bronze trophies according to their achievement level, which are displayed on their screens. Their results show their approach has some potential and might have positive implications for supporting positive health-related behavior change on individuals.

Peers Comparison

In relation to the peers comparison features, in a recent study Edney et al. [Edney et al., 2020], explored a social networking and gamified app to increase physical activity. In their app, participants were allowed to interact with their Facebook friends who also use the app. Participants can post messages and photographs on a Facebok-style newsfeed, send and receive virtual gifts, have daily and weekly competitions aiming to get the highest step count. Their results showed that 8 weeks into the intervention, participants had significantly increased their total weekly moderate-to-vigorous physical activity.

In another study, Altmeyer et al. [Altmeyer et al., 2018] tested a gamified mobile app with a public display to encourage walking. Their results show that the public display significantly increased step counts and participant’s motivation to walk, which they attributed to the significant increase in social relatedness. Also, they found that the public display made participants aware that their steps data was visible for outsiders and that they could be confronted with their performance, adding extra motivation to increment their step counts.

In sum, all this research show the positive behavioral effects these specific app feature designs have on individuals. They show the relevance of setting challenging goals to support higher physical activity levels; also the importance of introducing levels when delivering rewards to participants, and finally, the support of social aspects when performing physical activities.

Moreover, one of the fundamental aspects of technological interventions is that the individuals should be exposed to the treatments, meaning that they should interact with the technological system. Otherwise, it is hard to relate the behavioral outcomes with the technological intervention. Therefore, understanding how engagement occurs in this type of interventions is fundamental. When the intervention technological tool presents low engagement, we might end up with high dropout rates (e.g., [Haque et al., 2020]).

3.3.3 User Engagement

Engagement is defined as:

> "Engagement is a category of user experience characterized by attributes of challenge, positive affect, endurability, aesthetic and sensory appeal, attention, feedback, variety/novelty, inter-activity, and perceived user control." [O'Brien and Toms, 2008]

Engagement is not a single phenomenon; the experience of product use over time contributes to an evolution of engagement [Kuru and Forlizzi, 2015]. According to [O'Brien and Toms, 2008], engagement occurs in three phases namely:

- *Point of engagement and reengagement*: "initiated by the aesthetic appeal or novel presentation of the interface, the users' motivations and interests, and users' ability and desire to be situated in the interaction and to perceive that there is sufficient time to use the application."
- *Engagement*: "when users are able to maintain their attention and interest in the application, and is characterized by positive emotions. Users want to customize the interface to meet their needs and receive appropriate and timely feedback from the application."
- *Disengagement*: "users disengage for many reasons such as the usability of the technology (i.e., challenge and interactivity), and distractions in their environments."

Kuru and Forlizzi go further by defining *engaging experience* as "the ability to inspire and motivate people, allowing repeated interaction with a thing over time" [Kuru and Forlizzi, 2015]. They studied the engaging experience in a physical activity tracking product and defined four characteristics of an engaging experience: *connectivity*, *curiosity*, *personalization*, and *motivation.*

- *Connectivity* is "the product's ability to communicate with the user, who expects to connect to the product whenever they desire." (e.g., people want to see their data instantly, but if there are usability problems that prevent them from connecting to the product they loose interest in their use or users expect the system to connect by providing analyzed data along with meaningful suggestions and achievable goals).
- *Personalization* is: "the product's ability to allow the user to make changes in the functionality, interface, information content or distinctiveness of a product to best support individuals' needs". (e.g., users want the system to make suggestions for behavior change based on analyzing their data.)

- *Curiosity* is: "the desire to learn about and keep interest in product data." (e.g., giving users instant access to their data, as people understand their data, the action of accessing it turns out more repetitive).
- *Motivation* is: "the product's ability to stimulate people's interest in order to make them continually interested in using the product for reaching a specific goal." (e.g., the user want to see more than a record of their activity or providing positive feedback about their behavior seems to bolster motivation and system use.)

This literature informs three necessary aspects to answer our research question: *what are the BPNs feature combinations that yield the most positive effects on the individuals' physical activity behavior?*. First, we acknowledge past research that has explored the use of the SDT to develop physical activity interventions. Second, we examine the state-of-the-art behavior change app features relevant to the design scope of our research app. Third, we acknowledge the concept of engagement and its relevance in technological interventions. In the following section, we present our SDT-informed research prototype.

3.4 Research Prototype

Our goal was to create a steps-tracking app with a simple set of features supporting the SDT Basic Psychological Needs (BPNs). In our previous research [Villalobos-Zúñiga et al., 2021] we examined the user perceptions of hypothesized BPNs supportive app features. This past research app was the first of its iterations, and in this present chapter, we introduce a second iteration by incorporating the findings of this previous research. We found the following features to provide support to the users' BPNs: personalized goal increments, detailed performance indication, and coopetitive challenges. In particular this most recent app version further develop these design suggestions in the *autonomy* and *relatedness* supportive features of *goal-setting* and *peers-comparison* respectively; and updated the *competence* supportive feature of *steps-history* to a *rewards* feature. In the coming paragraphs, we describe the design process and rationale for each of the new Basic Psychological Needs supporting features.

Autonomy support feature

On the previous version of the *goal-setting* feature participants could see a dialogue indicating a suggestion for a weekly goal (see Figure 3.1). In this suggestion, the step-goal was increased by 5% concerning last week's average. Participants could accept or deny this proposition. From our findings, participants expressed there were occasions in which incrementing 5% was too much of an effort to them. Therefore they denied the proposition. To improve this feature, we provide participants with more *control* over their step goal-setting. We offer them information about their average steps counts (decisions with information support autonomy [Ryan and Deci, 2017, p. 461]). We also modified the phrasing of the text by not offering a suggestion; instead, we pose the question: "How many steps do you want to increase your step-goal this week?". This question phrasing will emphasize the decision weight on the participant because they will have to explicitly choose how much they want to increase the daily goal and not just accepting or denying the app's weekly goal suggestion. Then we offered more choices by allowing the individuals to choose between increasing their weekly steps goal by 0%, 5%, or 10% (See Figure 3.1 (a)). To aid the individual with the decision-making, we placed the equivalent of each increase in number of steps as a unit (e.g., 10%: 193 additional steps by day) next to each option. When the user selects an option, it gets highlighted (See Figure 3.1 (b)), and after tapping on the SAVE button, a small dialogue with a time indicator announces the system is saving the goal selection (See Figure 3.1 (c)) . When the saving is complete, the app displays a text field with instructions for the current week and a summary of the steps goal (See Figure 3.1 (d)). Finally, the app displays its main screen.

Competence support feature

On the previous version of the *competence* support feature, participants could observe a two-column table: the first one indicated the total number of steps for a given day, and the second one indicated the date in year-month-day format. Participants could see their steps history for 2-months back. From our findings, participants conveyed that there was no weeks indicator on the list, nor a sum of total weekly steps. Therefore, they made steps computations to know if they achieved their weekly goal; this effort produced a cognitive overload and did not facilitate the feeling of being competent at the walking activity. To improve these features and provide participants with a better *competence* need satisfaction, we replaced the steps history feature with rewards.These

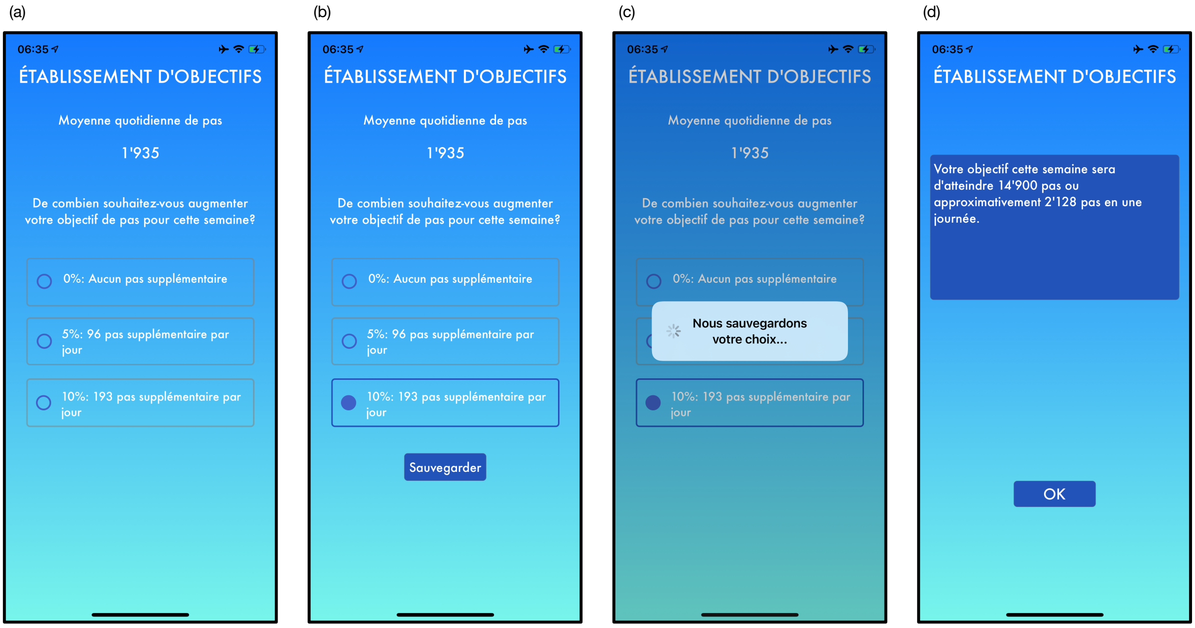

Figure 3.1: (a) Goal setting feature with options to increase the weekly goal. (b) Goal setting feature with the 10% increase option highlighted. (c) Waiting dialogue indicating the chosen goal option is being saved. (d) Summary screen with the weekly goal.

were represented in the form of yellow stars that *remained* on the screen allowing participants to visualize their progress better and reduce the cognitive workload (See Figure 3.2 (a)). Specifically, for every daily and weekly goal achieved, a star appears on the main screen. Similarly, for every weekly competition won, a trophy appears (See Figure 3.2 (b)). When the participant has more than 5 stars or trophies, the app displays the total number of steps in parenthesis next to the array of icons. Including these persisting rewards released the strain participants had to put in calculating their step totals and keeping in mind how many instances they had achieved the objective.

Relatedness support feature

On the previous version of the *relatedness* support feature, participants could observed a table with two columns: the first indicating participants' pseudonyms, and the second column with their daily steps. This list did not have any particular order and it's intention was to foster comparison between participants. They could not see further information from the other participants except for their first names, this caused participants to question if the people on the peers-comparison list were bots or "real" humans. To improve this feature iteration and provide participants with

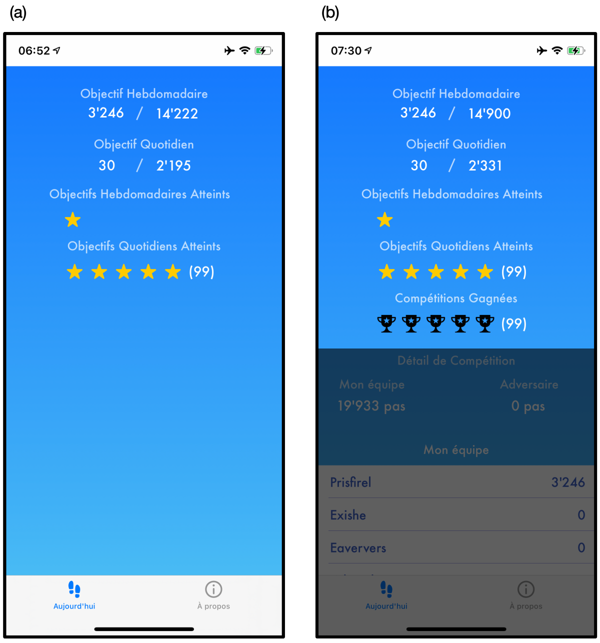

Figure 3.2: (a) Competence feature displaying one daily objective achieved. Weekly objective achieved displaying five stars and ninety nine total. (b) Competence feature displaying five trophies and ninety nine total. Lower part of the screen is grayed because it corresponds to the *relatedness* feature.

a better relatedness need satisfaction, we introduce a weekly competition in which they competed against other participants enrolled in the study (See Figure 3.3 (a)). Also we added a participant information section on the app that when tapping on the participant’s pseudonym, the app displayed a screen with their avatar, pseudonym, hobbies and age (See Figure 3.3 (b)). We included this functionality to give a more humane feeling to the participants in the study, whose names appeared on the leader boards. Including a hint that other people on the app are similar to the app user is fundamental for the relatedness satisfaction as it aids in creating this sense of connection and importance to another person or group [Ryan and Deci, 2017, p.297] to Additionally, we improved the peers-comparison feature by presenting the steps ordered in a decreasing fashion. The steps displayed were weekly totals. We now discuss our research methodology.

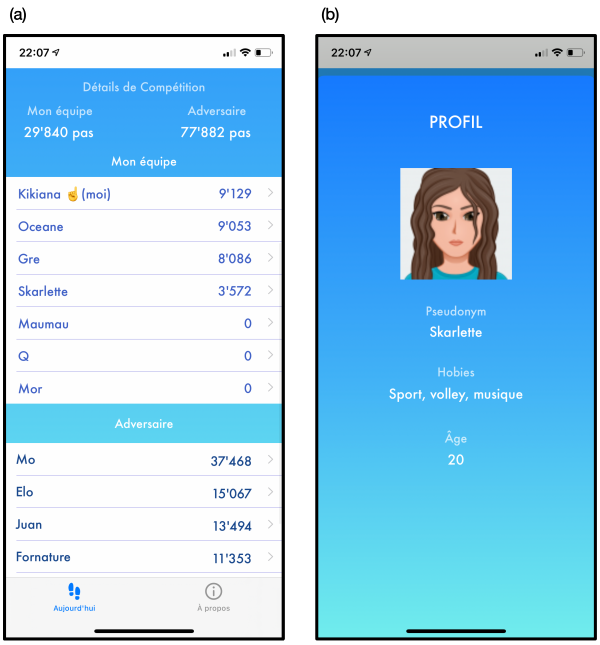

Figure 3.3: (a) Peers comparison feature, on top the sum of steps for the participants team and the rival. Below the leader-board for the participant's team and below the leader-board for the rival team. (b) When tapping on a pseudonym on the leader-bard the profile window pops up with participant's information.

3.5 Methodology

To answer our research question, we conducted a longitudinal study given that we intended to examine the individual's behavioral outcomes triggered by the BPNs supporting app features. These behavioral reactions concerned physical activity and intrinsic motivation. Therefore, we conducted a 5-month, randomized controlled trial. We had 4-months of observations (experiment) and 1-month of follow-up (post-phase). Also, we had a 7-month baseline of physical activity data (pre-phase). Having this time structure allowed us to have a complete set of observations not limited to experimental-window-only observations. Past studies on human motivation used this methodology (e.g., [Hanus and Fox, 2015]; [Oga-Baldwin et al., 2017]; [Reinboth and Duda, 2006]).

In addition to time considerations when running a physical activity study, we know that it is relevant to take a holistic approach, which means acknowledging that there are multiple *forms* of physical activity (e.g., yoga, swimming, climbing, cycling). Unfortunately, there is not a standardized metric in which we can measure physical activity that spans all possible activities. One could consider heart rate level as a metric to measure physical activity. However, we discarded it because heart rate is not exclusively affected by physical activity levels. Energy expenditure (EE) could be considered a promising metric to measure physical activity using wrist-worn devices. However, most of these devices poorly estimate EE [Shcherbina et al., 2017]. With this said, in this study, we focused on walking (and running) for 3 reasons:

- It is a universal activity that is naturally performed by most people, even those who do not do sport. For instance, going to work or school involves walking.
- It is performed everywhere, even outside of sports facilities.
- Software and hardware solutions for logging walking (or running) have improved in the last years and are now used to conduct scientific experiments.

In the following section, we provide the details of our study design.

3.5.1 Sampling Procedures

We had a total of 628 potential participants who enrolled in the study, and we assessed for eligibility. These people belong to a pool of subject volunteers (approximately 8K) for behavioral experiments at our university. LABEX, a specially designed unit at our university, manages this subject pool. They take charge of participants' randomization, enrollment, the transfer of financial incentives, and keeping all participants' personal information safe. We used a screener to filter out respondents who did not meet the study requirements. It also served to capture demographic data and self-report physical activity level (Global Physical Activity Questionnaire - GPAQ) [McCallum et al., 2018]. The purpose of this tool is to yield valid and reliable estimates of physical activity. With respect to the filtering of participants, this is our exclusion criteria:

1. Foresee moments in which they will remain without Internet connection during the following 6 months or more than 7 consecutive days.
2. Have a planned surgery or any circumstance that will prevent them from walking during the next 6 months.
3. Did not own an iPhone 6s Plus or newer model.
4. Did not practice any sport that involved walking or running.

After the eligibility selection, we excluded 173 respondents: 128 owned an Android Phone, 2 planned to buy an ANDROID or WINDOWS phone, 18 owned an IPHONE 6 or older, and 25 had an IOS version older than 13.7. We then invited 455 potential participants to install the research app, and we capture the satisfaction in exercise of the Basic Psychological Needs (PNSES) and the intrinsic motivation level [Ryan, 1982]. The Intrinsic Motivation Inventory (IMI) is a multidimensional measurement tool to evaluate the participants' subjective experience concerning a target activity. The IMI has been used in multiple experiments involving intrinsic motivation and self-regulation (e.g., [Deci et al., 1994]; [Plant and Ryan, 1985]; [Ryan, 1982], [Ryan et al., 1990]; [Ryan et al., 1991]). This tool generates 7 sub-scale scores by assessing the participants' interest/enjoyment, perceived competence, effort, value/usefulness, felt pressure and tension, perceived choice, and relatedness while performing a given activity. For our purposes, we use the sub-scale of interest/enjoyment to measure intrinsic motivation. The validity of the IMI has been submitted to scrutiny by [McAuley et al., 1989] finding strong support for its effectiveness. We invited 455 invited participants, 177 did not complete the questionnaires (IMI and PNSES), 35 did not install the research app. Therefore we ended up with 243 participants, which we assigned to the different experimental conditions. During the study 2 participants drop the study. Figure 3.4 shows the a flowchart with the progress of participants through the study as of June 30th, 2021.

3.5.2 Participants Characteristics and Recruitment

The participants' sample included 241 students from the University of Lausanne, Switzerland. They were 18 years of age or older (M = 20.7, SD = 2.32); 170 were women (71%) with average age IMI level of 4.8 (SD = 1.10). Students belonged to all seven faculties of the university, ensuring a mix of technical and non-technical backgrounds. We used the screener to measure additional baseline characteristics like the self-report physical activity level using the Global Physical Activity Questionnaire (GPAQ) which allowed us to classify participants that practice *vigorous* or *moderate* physical activity. With the Autonomy and Competence in Technology Adoption (ACTA) Questionnaire [Peters et al., 2018] we determined willingness to adopt new technology it provided us with a *Relative Autonomy Index* which we used for the randomization. We randomly assigned participants to 7 experimental conditions plus

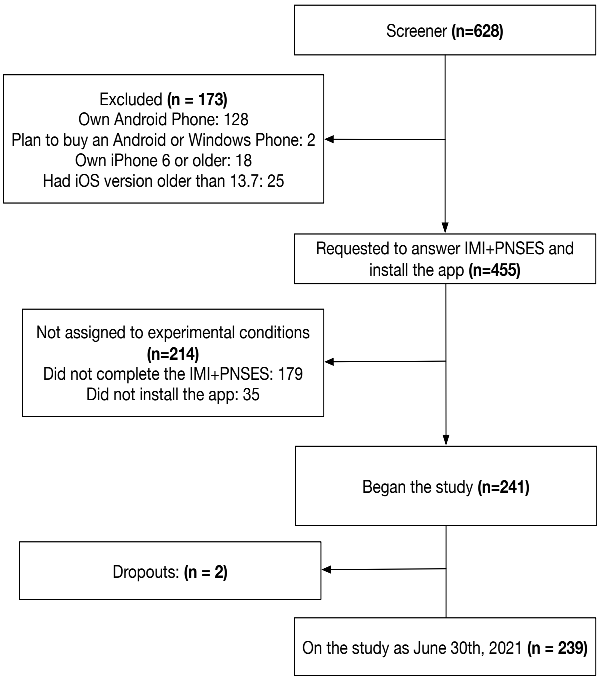

Figure 3.4
Flowchart showing the recruiting process of participants.

a control condition using the BLOCK_RA R function. We checked that there was an even distribution of the participants' characteristics across the conditions.

3.5.3 Experimental Design and Interventions

We used a between-subject factorial design, with the above mentioned 7 conditions plus a control condition. We detail each condition in the following lines.

1. Autonomy (A): Participants in the *Autonomy* condition had three options to select their weekly step goals. After choosing an option, they were taken to the home screen to visualize their weekly and daily goals and how many steps they have accumulated for the day.

2. Autonomy-Competence (AC): Participants in the *Autonomy-Competence* condition had the same options for goal-setting as participants in the *Autonomy* condition and the same home screen. In addition, the home screen displayed an indicator of daily and weekly goal *achievements*.

3. Autonomy-Relatedness (AR): Participants in the *Autonomy-Relatedness* condition had the same options for goal-setting as participants in the *Autonomy* condition

and the same home screen. In addition, after setting their goal, they received instructions indicating that they will be assigned to a team and that they will compete against other participants. The home screen displayed their team leaderboard and the rival one; also their weekly and daily goals, and how many steps they have accumulated for the day.

4. Competence (C): Participants in the *Competence* condition did not choose their goals. Instead, the system randomly chose and assigned them. The home screen displayed an indicator of daily and weekly goal *achievements*, their weekly and daily goals, and how many steps they have accumulated for the day.

5. Competence-Relatedness (CR): Participants in the *Competence-Relatedness* condition did not choose their goals. Instead, the system randomly chose and assigned them. In addition, they received instructions indicating that they will be assigned to a team and that they will compete against other participants. The home screen displayed their team leaderboard and the rival one, plus an indicator of daily and weekly goal *achievements*, their weekly and daily goals, and how many steps they have accumulated for the day.

6. Relatedness (R): Participants in the *Relatedness* condition did not choose their goals. Instead, the system randomly chose and assigned them. In addition, they received instructions indicating that they will be assigned to a team and that they will compete against other participants. The home screen displayed their team leaderboard and the rival one; also their weekly and daily goals, and how many steps they have accumulated for the day.

7. Autonomy-Competence-Relatedness (ACR): Participants in the *Autonomy-Competence-Relatedness* condition had the same options for goal-setting as participants in the *Autonomy* condition and the same home screen. In addition, after setting their goal, they received instructions indicating that they will be assigned to a team and that they will compete against other participants. The home screen displayed an indicator of daily and weekly goal *achievements*, their weekly and daily goals, and how many steps they have accumulated for the day, in addition to their team leaderboard and the rival one.

8. Control (CON): Participants in the *Control* condition did not have any of the above functionalities. The home screen just displayed a message indicating that the app was working normally.

We describe the design details of each condition in section 3.4

3.5.4 Apparatus

There are multiple devices employed to track steps (e.g., bracelets[1], wearable sensors[2]) these devices bring advantages to the data gathering precision because people tend to wear them most of their day. However, they do present the risks of people forgetting to wear them or to charge them. If this situation occurs, it leaves gaps in the data harming its quality. Therefore, in this study, we leveraged the APPLE IPHONE capability to track steps. APPLE facilitates extracting the steps data from the IPHONE by providing an application program interface (API). With the support of the API we can get easy and rapid access to the steps data and at the same time reduce development costs that are necessary when building infrastructure for an experiment. Specifically, we used APPLE'S HEALTHKIT, a central repository for health and fitness data. Data stored in this repository is collected through the iPhone's accelerometer, encrypted, and saved. The main advantage of using this service is that the activity tracking is always running in the background, capturing steps data although our experimental app is not running. Moreover, HEALTHKIT allowed us to retrieve steps data from months before the launch of our study, which we used to compute our steps count baseline. Our research app filtered out all manually input steps, and we did not consider these steps in any aspect of the study.

3.5.5 Procedure

Selected participants received an email with the invitation to participate in the study; if they agreed to join, they signed a consent form which indicated the participation conditions and the monetary compensation. Also, this email included a link to the APPLE APP STORE where participants could download the Agon app and a link to a website to create their pseudonym, avatar and upload them to the system. In addition, in the email, participants were requested to answer the Intrinsic Motivation Inventory Questionnaire (IMI) and the Psychological Need Satisfaction in Exercise Questionnaire (PNSES), which were mandatory to participate in the study.

[1]E.g., https://www.fitbit.com, last retrieved August 2021.
[2]E.g., https://www.suunto.com, last retrieved August 2021.

We allowed 4 weeks for all the participants to answer these questionnaires and install the app. If they did not comply, then they were not considered for the study. During this period, participants logged into the system using the email address they input while registering for the study participation.

After they logged in, the app was in standby mode displaying a message indicating that the system was working correctly. After the 3 weeks of setup time, we randomly assigned participants to the experimental conditions and balanced them across all the experimental conditions considering the IMI (captured on the invitation email) and GPAQ (captured on the screener) results.

The study ran across the spring semester and summer of the academic year: First, in March (week 11), participants received the link to install the app, and we allowed a three-week window for them to install it. In early April 2021 (week 14), we launched the intervention and concluded in late August 2021 (week 35). On week 14, we enabled the treatment, the app's user interface changed according to each experimental condition, and we started collecting step counts data and logged the number of interactions they had with the app (e.g., tap on buttons, scroll up and down, opening and closing the app). We asked participants to complete the TENS-Interface and TENS-Task questionnaires on week 18. On week 26, they answered the IMI and the PNSES as a mid-experiment measure. On week 31, we stopped the treatment making the experimental app display the same screen as the launching of the study. On week 35, at the end of the study, we deployed the IMI and PNSES as a post-experimental measure. During the study we contacted by email participants that had not opened the app for more than two continuous weeks, inquiring if they had problems with the app or if they had left the study. Participants did not receive reminders or notifications.

3.5.6 Descriptive Analysis

"Quantitative descriptive analysis characterizes the world or a phenomenon by identifying patterns in data to answer questions about who, what, where, when, and to what extent. Descriptive analysis is data simplification." [Loeb et al., 2017]. Because this is an ongoing study, we approach making sense of the current data by performing a *descriptive analysis*. Therefore in the following paragraphs, we present preliminary data analysis based on data from October 5th, 2020, to June 30th, 2021. Although the study began in March 2021, we were able to extract step counts data up to October 2020 thanks to the capabilities of the APPLE HEALTHKIT, which we describe in more detail in the Apparatus section. Furthermore, we performed all the analysis computations using the statistical software R.

The purpose of this section is to present the descriptive analyses of our data which serves to describe the basic features of our data. Once the study concludes we will move forward with inferential statistics that will lead us to conclusions. The outcome variables of this analysis are:

- Number of Steps: computed as the total steps count over each week during the experimental-phase and post-phase.
- Intrinsic Motivation Index (IMI) [Ryan, 1982]: measured at the onset of the study(IMI1 - week 10) and during the experimental-phase (IMI2 - week 26).
- Recency: measured as the number of weeks since the last app interaction. Examples of app interactions are Tap on Today View, Tap on the Avatar, Bring app from Background to Foreground or vice-versa. Interactions are based on the event logs captured from the moment they installed the app (week 10) until the end of June (week 25).

We followed a similar methodology to Matsumoto et al. [Matsumoto and Takenaka, 2021]. Data analysis proceeded in the following stages: First, we performed data preparation by doing data profiling and quality checking, which involved checking for missing values and estimating internal consistency score for the IMI and PNSES using α coefficients [Cronbach, 1951]. Then we performed the exploratory analysis, where we computed descriptive statistics for the 4 previously defined outcome variables. We developed a statistical analysis to find the underlying patterns, relationships, and trends between the variables. We primarily applied the Box plot analysis to display relevant variables' distribution and detect outliers and similarities between groups. Finally, we performed a statistic test, which we present in the following section.

3.5.7 Treatment of Missing Values

When treating the missing data, we noticed that the observed number of steps for each participant on each day could be missing if the participant's smartphone was turned off or disabled the fitness tracking option from HealthKit. The data could also be incomplete if the participants did not carry the phone with them. Therefore the number of steps lower than 980 were considered unrealistic and treated as missing values. We used the simple moving average as an imputation technique to handle the missing steps data, considering each participant's steps records as an individual time series. We calculated each participant-day a moving averaged based on the previous 15 days, and then we replaced the missing values with the corresponding moving average.

3.5.8 Intrinsic Motivation Index Treatment

For the IMI computation, we followed the same methodology utilized by [McAuley et al., 1989] to calibrate the IMI instrument. McAuley et al. evaluated the internal consistency of the IMI subscales by computing the Cronbach's Alpha coeficient [Cronbach, 1951] to determine the reliability of these measures and conducted a factor analysis to obtain the minimum number of factors that will account for the correlations among the observed variables.

3.6 Preliminary Results

In the following section, we present our preliminary results, which correspond to data from the time frame of October 5^{th}, 2020, to June 30^{th}, 2021.

3.6.1 Recency

The metric of *recency* relates to engagement because it tells us how interested a person is in the app because they continue using it. We map participants with smaller recency to the phase *engagement* as defined by [O'Brien and Toms, 2008] and participants with higher recency to the phase of *disengagement*. In this analysis, we are interested in participants who continue using the app and have small recency. It could have occurred that some participants were engaged at the onset of the study but entered the *disengagement* phase later on. Therefore they are not being exposed to the treatment anymore. Participants' level of exposure to the treatment is fundamental to understand the intervention's general effect.

We measured *recency* as the number of weeks between the last week of app usage and the cut-off week (week 25). Given that we had 11 weeks of data (week 14 to week 25), we divide the recency into ranges of 0 to 2 weeks or *engaged* , 3 to 5 weeks or *starting to disengage*, or more than 6 weeks or *disengaged*. With this classification, we observed that from the total number of participants, 71% (150 participants) present recency between 0 to 2 weeks, meaning that they are using the app by the time of this analysis. We describe this group as participants that *are interested in physical activity and use our research app to support it*. Then, 13% (27 participants) have not interacted with the app from 3 to 5 weeks. We describe this group as participants that *were interested in the app, or study during some weeks but lost interest*. Finally, 16% (33 participants) who have 6 or more weeks since the last usage. We describe this group as participants that *Enroll in the study but might not engaged with the app or are apathetic about physical activity*. See Table 3.1

Experimental Condition	**Users' Recency**		
	0 to 2	**3 to 5**	**6 or more**
A	19	7	6
C	27	1	2
R	29	1	*0*
AC	15	7	8
AR	17	5	7
CR	23	4	2
ACR	20	2	8
Total	150	27	33

Table 3.1
Users Recency in groups of 0 to 2 weeks; 3 to 5 weeks; and 6 weeks or more. Data displayed for each experimental condition. Notice experimental conditions R, C, and CR show the lower recency, which means they continue interacting with the app. Participants with a recency of 3 or more weeks are considered as participants that abandoned the treatment.

As we mentioned in section 3.3, app engagement has a crucial role when designing behavior change interventions. Therefore, in the following sections, we analyzed the outcome variables *Number of Steps* and *Intrinsic Motivation Index* according to the participants' recency classification to avoid mixing the results of participants who were exposed to the intervention treatment and those who abandoned the study.

3.6.2 Weekly Steps

Based on the total steps by week, we calculated the average weekly steps per participant in the following four different periods namely:

- *Pre*: subset of the *pre-phase* presented in section 3.5. Corresponds to a 6-month range, from September 2020 to February 2021 (4 weeks prior the onset of the study). We computed the average weekly steps from this period and used it as our *baseline*.
- *Set-up*: subset of the *pre-phase* presented in section 3.5. This period corresponds to March 2021. It is the time-window where participants installed the research app on their IPHONES. We excluded this period from the analysis to isolate the effect of the experiment setup.
- *Exp1, Exp2, Exp3*: The following are three 4-week periods which are subsets of the *experiment* phase presented in section 3.5. These 12 weeks span from the launch of the study in April to the data cut-off in June. We computed the weekly averages of each period and compared them to the *baseline* to quantify the improvement on weekly steps for each recency range.

We aim at study the difference in the treatment effect versus the control group to confirm a causal effect of the app usage over the outcome variable of average weekly steps. Therefore when we looked at the average weekly steps by rencency groups across the *Pre*, *Exp1*, *Exp2*, and *Exp3* periods (see Table 3.2 and Figure 3.5). We observed the following:

Recency	**Pre**		**Exp1**			**Exp2**			**Exp3**		
	Mean	SD	Mean	SD	Δ	Mean	SD	Δ	Mean	SD	Δ
0 to 2	37,824	11,389	41,989	14,104	11.0%	41,401	14,672	9.5%	40,257	14,783	6.4%
3 or more	35,324	11,033	38,319	14,054	8.5%	37,772	14,845	6.9%	36,224	14,485	2.5%
Control	39,630	10,518	40,268	8,955	1.6%	41,922	16,537	5.8%	41,192	15,369	3.9%

Table 3.2
Variations at different recency categories between the subsets of the experiment periods. Δ is computed for each period with respect to the *Pre* period.

During *Exp1* the steps' growth rates in the recency ranges are higher (*0 to 2* and *3 or more* than *control* group, 11.0%, 8.5%, and 1.6% respectively.

For *Exp2* we observed that the recency range *0 to 2*, showed a growth rate of 9.5% while the *control* group increase at a lower rate by 5.8%.

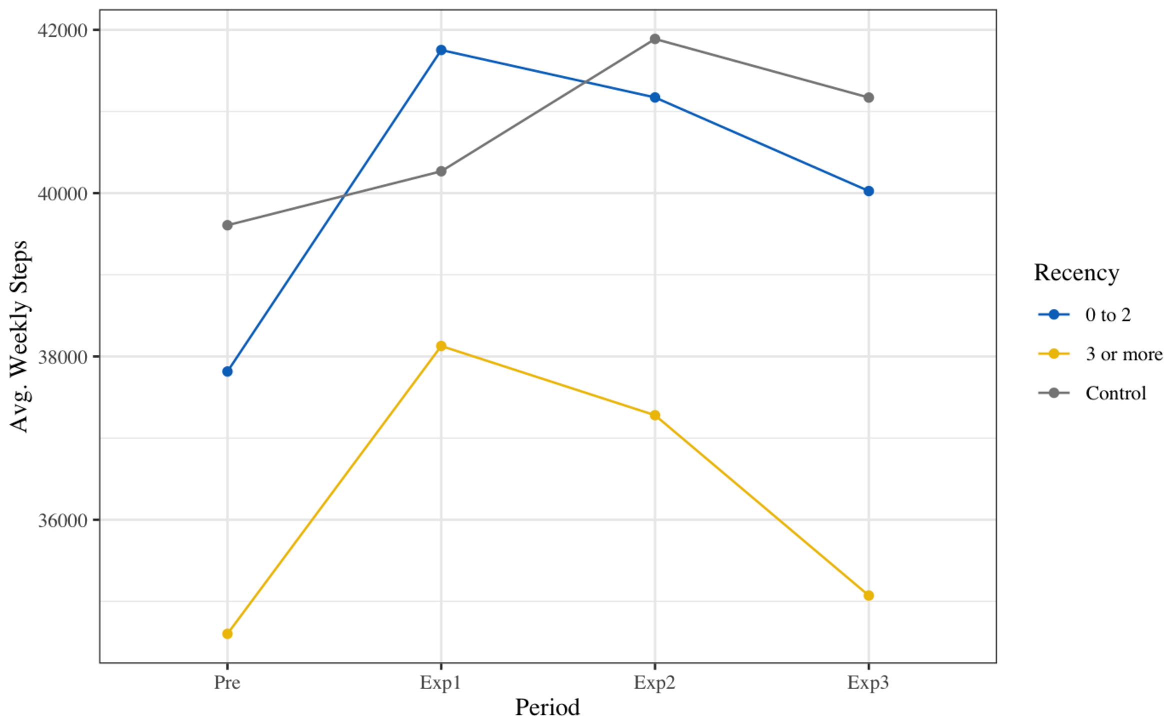

Figure 3.5
Difference in Difference Illustrative Graph showing, Average Weekly Steps displaying by recency ranges. X-axis includes the time periods: Pre equals six-month prior the launch of the study; Exp1 is 4 weeks into the study Exp3 is 12 weeks into the study, Y-axis includes the average weekly steps. Data corresponds to recency range *"0 to 2"*.

For *Exp3* we observed that despite the *Control* group showed recovery and *0 to 2* group decreased related to *Exp2* period. We observe that in comparison to the baseline period, the *0 to 2* group increased by 6.4%, and the *Control* group increased by 3.9%.

The above suggests a positive effect on the treatment, which we validated through a difference-in-difference analysis.

Now that we have seen the variations per recency ranges, we consider recency range *0 to 2* (engaged participants or participants receiving the intervention) and display in Figure 3.6 and Table 3.3 the average weekly steps per experimental condition.

To study a causal relationship between participants receiving the intervention and weekly steps, we conducted a difference in difference analysis (DID). DID is a widely used method to mitigate biases from groups or time-invariant factors [Abadie, 2005] [Blundell et al., 2004]. This method combines time-series difference, compar-

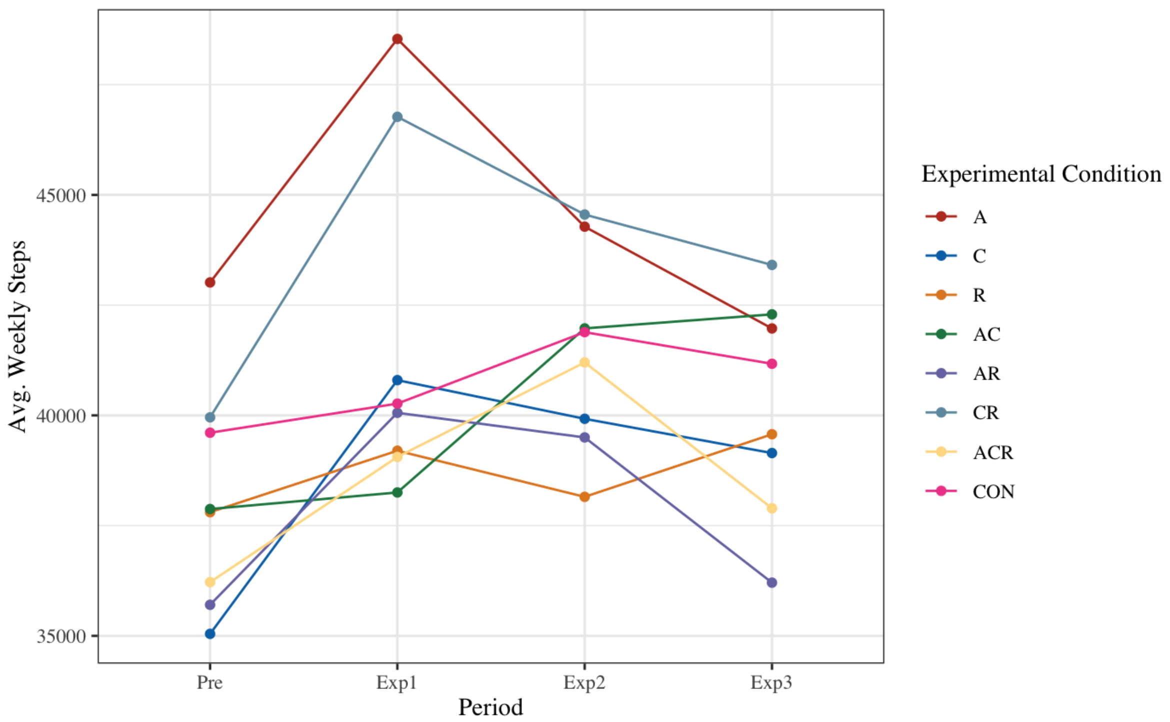

Figure 3.6
Diference in Diference Graph showing Average Weekly Steps by experimental conditions. X-axis includes the time periods: *Pre* equals six-month prior the launch of the study; $Exp1$ is 4 weeks into the study, $Exp2$ is 8 weeks into the study, $Exp3$ is 12 weeks into the study, Y-axis includes the average weekly steps. Data corresponds to recency range *"0 to 2"*

Experimental Condition	**Pre**		**Exp1**			**Exp2**			**Exp3**		
	Mean	SD	Mean	SD	Δ	Mean	SD	Δ	Mean	SD	Δ
A	43,016	11,433	48,561	16,456	12.9%	44,280	13,231	2.9%	41,972	15,143	-2.4%
C	34,841	12,905	41,129	14,583	18.0%	40,438	17,216	16.1%	39,018	11,621	12.0%
R	37,599	10,313	39,456	13,985	4.9%	38,290	12,910	1.8%	39,603	12,288	5.3%
AC	37,877	13,183	38,252	12,056	1.0%	42,110	20,992	11.2%	42,312	18,900	11.7%
AR	35,746	13,734	40,084	14,567	12.1%	39,561	15,975	10.7%	36,225	17,357	1.3%
CR	39,956	9,380	46,789	15,327	17.1%	44,597	13,685	11.6%	43,412	14,088	8.7%
ACR	36,466	9,071	39,471	8,777	8.2%	41,609	9,701	14.1%	39,326	17,324	7.8%
CON	39,630	10,518	40,268	8,955	1.6%	41,922	16,537	5.8%	41,192	15,369	3.9%

Table 3.3
Average weekly steps by period for each experimental condition. Δ is computed for each period with respect to the *Pre* period. Data corresponds to recency range *"0 to 2"*.

ing outcomes across $pre - phase$ and $experiment - phase$ periods and cross-sectional difference, comparing results between the experimental conditions to *control* group.

This analysis is valid under the assumption that the intervention is unrelated to the outcome at baseline and the composition of intervention and comparison groups is stable for repeated cross-sectional design.

For this purpose, a linear regression model estimates the effect on weekly step counts in participants for the different experimental conditions during the *experiment phase* in comparison to the *control* group.

We fitted the following model:

$$Y_{kt} = \beta_0 + \beta_1 P_t + \beta_2 T_k + \beta_3 (P_t * T_k) + u_{kt}$$

The independent variable Y refers to participants' weekly steps. P is a dummy variable that indicates if the week was during the *experiment phase* (0 in *Pre*, 1 in *Exp1, Exp2, and Exp3*). T_k is a dummy variable to identify the experimental condition k, k in $\{A, C, R, AC, AR, CR, ACR\}$ exposed to the treatment (0 for participants in the *control* group, 1 for participants in the corresponding experimental condition). $P * T$ is the interaction between the treated groups, called the "DID" interactions.

	Estimate	Std. Error	t value	Pr(>|t|)	
Intercept	39719.2	633.7	62.68	0.000	***
Phase					
P	1671.1	1140.2	1.466	0.143	
Experimental Condition					
A	3296.6	1042.4	3.163	0.002	**
C	-4732.6	941.7	-5.026	0.0000005	***
R	-1202.9	912.6	-1.318	0.188	
AC	-1842.6	1110.6	-1.659	0.097	.
AR	-2565.6	1062.7	-2.414	0.016	*
CR	236.7	952.8	0.248	0.804	
ACR	-3253.2	991.9	-3.28	0.001	**
Interaction					
P*A	303.4	1862.6	0.163	0.871	
P*C	3725.1	1684.2	2.212	0.027	*
P*R	-1013.6	1630.1	-0.622	0.534	
P*AC	1810.2	1980.7	0.914	0.361	
P*AR	1151.8	1896.4	0.607	0.544	
P*CR	3429.1	1703.9	2.013	0.044	*
P*ACR	2214.1	1773.1	1.249	0.212	

Table 3.4
Difference in Difference Linear Regression coefficients. Data corresponds to recency range *"0 to 2"*

The model was performed for each experimental condition. However, only CR and C groups showed a statistically significant effect. The coefficients and their $p - values$ are reported in Table 3.4.

This outcome shows a preliminary result, showing the positive effect of the intervention for participants in C and CR conditions.

We can interpret the regression model as for CR, the coefficient of the interaction term DID is 3429.1, (p<.05). That means that the participant of CR walked 3429 steps more than the *control* group during the intervention period. Similarly, C group showed a positive difference of 3253.2 steps over the *control* group.

3.6.3 Intrinsic Motivation Index

Given that we want to compute a difference between the baseline measurement of the IMI (IMI1) vs. the follow-up measurement (IMI2), we needed both measurements from the participants. However, 18.5% ($n_{missing} = 46$) of them did not complete IMI2, and we had to exclude them from the IMI analysis.

As an initial step, we analyzed the participant's IMI scores, following the methodology employed by [McAuley et al., 1989]. First, we conducted a simple Factor Analysis to calibrate the IMI model based on our specific data structure. This analysis let us select the items that best represent the IMI latent factor and validate the invariance across time for the selected items. Then Cronbach alpha was used to measure that the selected variables of IMI form a coherent and reliable factor.

For the seven items, the Cronbach coefficients are $\alpha = .908$, $CI(2.5\%) = .873$, $CI(97.5\%) = .932$ for IMI1, and $\alpha = .898$, $CI(2.5\%) = .863$, $CI(97.5\%) = .922$ for IMI2. High alpha coefficients indicate that all items in the scale Pre (IMI1) and Post (IMI2) were perceived very similar to each other by the participants.

In Table 3.5, we show factor analysis results, including a varimax rotation to maximize the sum of the variance of squared loadings.

	Item 1	**Item 2**	**Item 3**	**Item 4**	**Item 5**	**Item 6**	**Item 7**
IMI1	0.86	0.80	0.73	0.76	0.41	0.83	0.76
IMI2	0.90	0.84	0.67	0.66	0.82	0.84	0.74

Table 3.5
Factor Analysis Loadings for IMI1 and IMI2. Data include all recency ranges.

The higher a factor loading, the more relevant a variable is for the said factor. Based on the results on Table 3.5, we defined a loading cutoff of 0.8. This cutoff determines which variables belong to the IMI factor. Based on the loading, we kept the items that consistently showed a simple factor loading greater than .8. Finally, we obtained the alpha coefficients for the final model are $\alpha = .905$, $CI(2.5\%) = .864$, $CI(97.5\%) = .934$ for IMI1 and $\alpha = .897$, $CI(2.5\%) = .857$, $CI(97.5\%) = .929$ for IMI2, which means that using Item1, Item2, and Item 6, the IMI measure maintains its internal consistency reliability.

After we had adjusted the measurement instrument, we continue with the examination of the IMI scores.

Similar to the weekly steps analysis, for the IMI analysis, we look at the data exclusively for the recency range *0 to 2*, see Figure 3.7.

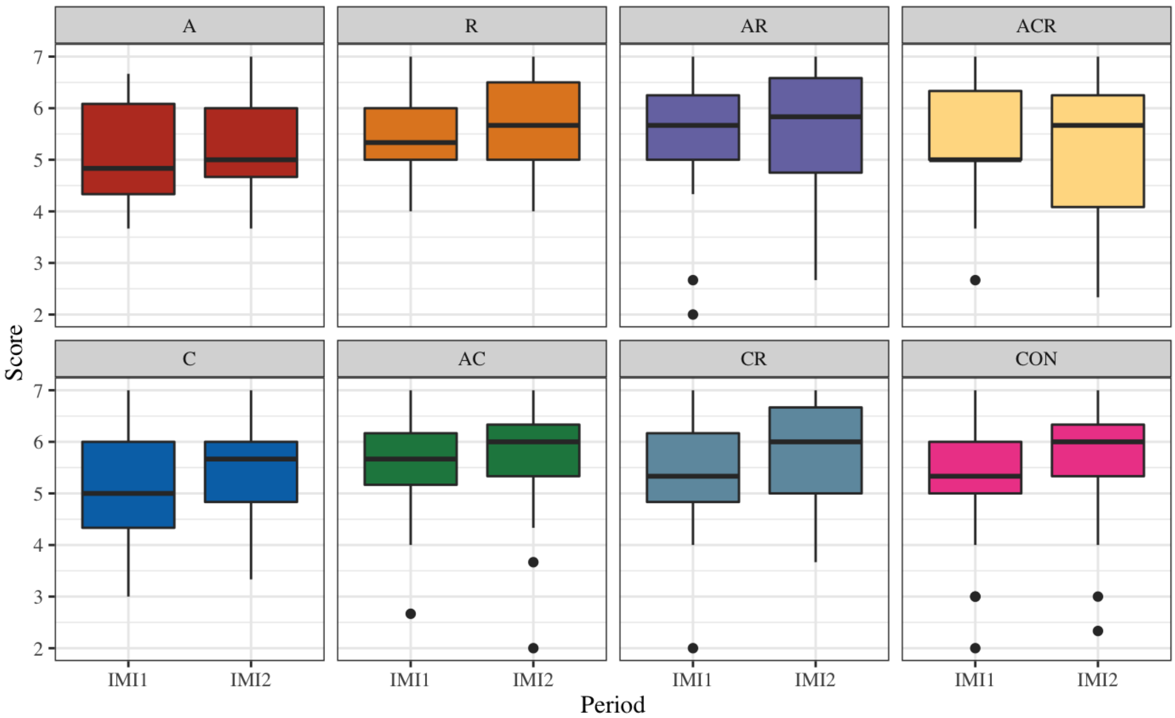

Figure 3.7
Box Plot Graph displaying the participant's IMI1 and IMI2 scores distributions by experimental condition. This boxplot illustrates the shift baseline IMI and the Follow-up. Data corresponds to recency range *"0 to 2"*.

To statistically validate the differences between IMI1 and IMI2, we performed a Paired Wilcoxon signed-rank test, a non-parametric test commonly used for analyzing data from a repeated-measures design with two conditions when the data are measured on an ordinal scale. A non-parametric test is appropriate since the variable's score is ordinal but not normally distributed. This test indicates whether a significant shift in the IMI scores has taken place. The hypothesis evaluated is whether or not there is a significant difference between the IMI1 or IMI2 scores of subjects on the dependent variable. See result in Table 3.6.

Experimental Condition	N	Mean	SD	Z	*p-value*	
A	16	0.21	0.90	-1.17	0.242	
C	19	0.37	0.73	-2.35	0.019	*
R	27	0.19	0.82	-1.87	0.062	
AC	15	-0.02	1.56	-0.77	0.444	
AR	14	0.07	1.01	-0.80	0.422	
CR	19	0.33	1.26	-2.16	0.030	*
ACR	18	-0.04	1.13	-0.45	0.653	
CON	25	0.36	1.29	-1.89	0.058	

Table 3.6
Paired Differences IMI1 IMI2. *N* is the number of participants. Mean is the mean of the score differences IMI1-IMI2. SD is the standard deviation f the score differences IMI1-IMI2. Data corresponds to recency range *"0 to 2"*.

For C and CR, the Wilcoxon one-tailed test showed a significant difference in their IMI scores ($Z = -2.35$, $p = .019$) ($Z = -2.16$, $p = .030$) respectively. We also evaluate the effect size by calculating Cohen’s d coefficient. Cohen’s d is defined as the difference between two means divided by a standard deviation for the data. For C, $d = 0.54$ and CR, $d = 0.54$ which represent a large effect, according to Cohen’s classification (d greater 0.5 are associated with large effect) [Cohen, 2013].

3.7 Discussion

3.7.1 Engagement in Relation to Physical Activity

As mentioned in section 3.3, *engagement* refers to the experience of using a product over time. We chose *recency* as a metric to measure engagement because it tells us about the continued use of the application. This third study taught us the importance of designing for engagement and aiming to obtain the lowest recency. This is important because participants of behavior change technological interventions should interact with the app that delivers the treatment to be exposed to that treatment. From our findings, we divided participants into two groups: (1) participants who have used the app at least once in the last two weeks; (2) participants who have spent three or more weeks without interacting with the app.

Our findings show a positive correlation between the participants' recency and the average weekly steps: the lower the recency, the higher the average weekly steps. More precisely, on period *Exp1* (4 weeks of intervention), participants improved their average weekly steps regardless of their recency level when we compare it to the increase of the control condition. We believe this occurred due to the *novelty effect* that appeared during this initial phase. Then, when we look at the next 4-week period in *Exp2* (8 weeks of intervention), we remarked a decrease in the average weekly steps for both recency groups when compared to the behavior of the control condition. However, the decrease is more substantial for people with higher recency. This effect is even more apparent for period *Exp3* (12 weeks of intervention). We observed that people with higher recency levels had low average weekly steps. Therefore, we believe that two potential causes for why participants in all recency groups increased their step counts and then decreased are: (1) the app did not succeed at engaging them for a sustained period; (2) participants were predisposed to low levels of physical activity given that they already had a low number of step counts at baseline.

Experimental Conditions and Engagement

When we look specifically at the experimental conditions with lowest recency ("0 to 2") (see Table 3.1) we noticed that the most engaged participants belong to condition *R*, followed by *C*, and then *CR*. The positive results from the *relatedness* feature might be caused by participants' interest in the weekly competition, the fact that they were part of a team, and that they could compare their performance with other teammates and rivals. In the case of the *competence* group, participants received feedback for their performance in the form of weekly and daily stars that appeared on the app screen. This might have triggered the interest of participants in the app. Then, the combination of these two features might have caused participants to feel interested not only because they were part of a competition but also because their effort was compensated at a group level by receiving trophies and at a personal level by receiving stars. On the contrary, it was *A* and *AC* the experimental conditions with the lowest number of engaged participants. A possible explanation for having more disengaged participants is that the *autonomy* support featured appeared once a week for a limited amount of time. Therefore, the feature could be easily ignored, or participants did not pay enough attention to it. Another possible reason is that participants were not interested in setting their weekly steps goals, either because they were not interested in walking or because the feature did not show choices that adapt to their needs.

Experimental Conditions and Physical Activity

When we take participants that have interacted the most with the app (lowest recency) and observe how they have behaved concerning their average weekly steps, we observe that during the *Exp1* period, all the experimental conditions presented an increase in their average weekly steps when compared to control condition (see Table 3.3. As we mentioned in previous paragraphs, we believe the cause of this increase is the novelty effect or the fact that they were enrolled in the study. In *Exp2* the experimental conditions that continue with a positive increase are *ACR*, *C*, and *CR*. In *Exp3*, *C*, and *CR* maintain their positive trend however, this is not the case for *ACR*. This decrease in the average weekly steps of conditions with *autonomy* support began to give us signs that we might be observing an adverse effect caused by the particular design of the *autonomy* support feature. Although our goal-setting feature yielded positive results concerning *autonomy* on chapter 2, the situation was not promising for the goal-setting feature in this chapter, as the preliminary results show.

Our preliminary results show a significant difference in the average weekly steps for *C* and *CR* experimental conditions. However, not for *AC* or *ACR*, this makes us reflect that the *autonomy* support provided by the goal-setting feature did not contributed to reaching significant difference to *AC* and *ACR*. A possible explanation to the effect of *autonomy* support could be that participants with this feature were requested to select how much they wanted to increase their weekly step goal. On the contrary, participants in experimental conditions without *autonomy* support received goals assigned by the app based on a random weekly goal increase between 0%, 5%, or 10% of their past week steps average. This means that both conditions had weekly goals; the only difference is that for experimental conditions without *autonomy* the goal was assigned automatically, meaning that participants did not have a choice on how much they wanted to increase it. It might occur that participants indeed experimented a feeling of *autonomy*, however, this was not enough to support an increase in physical activity levels.

We learned the importance of designing apps that are not monotonous and that evolve. Eventually, users will disengage, but recency groups can help determine which strategies to use to re-engage them. For instance, app designers could develop interactions or new features on the app that appear through time to call the users' attention. For instance, a new pop-up window or sound when they have reached a certain step-goal; or enable a feature to update their avatars once their team has won 3 competitions, or allow users to customize the look and feel of the app. For disengaged users, designers might consider employing reminders to bring them back to the app. The goal is to maintain attention and interest in the application.

3.7.2 Competence and Relatedness Support: the Winning Feature Combination

Overall we can say that from all the possible feature combinations, *competence* and *relatedness* was the one that yielded the most positive effects concerning the two primary outcome variables. Participants in this experimental condition presented the highest increase in the average weekly number of steps compared to the control condition and presented the highest positive difference when compared their baseline IMI score to the follow-up (IMI2). What might be causing *CR* to stand from the rest of other feature combinations?

CR experimental group results could be explained by the fact that the competition in which they were involved influenced them to walk more. These participants were part of the same team through the study, and on the app screen, they could see their weekly sum and their teammates sum. They also were ranked in each team, meaning that they could also compete for a first-place inside their team (coopetition) [Wolf et al., 2020]. Further than that, when the weekly competition was over, and if they beat their opponent, participants could see a trophy appeared on their screens, this was a reward for winning the competition, and it persisted on the screen along the course of the study. All these aspects correspond to the design of the *relatedness* feature.

In addition, the app screen for *CR* participants showed stars for each daily and weekly goal they achieved. This means that they were not only involved in the *team competitions*, but also they had *personal goals* to achieve. When comparing this design to the competence feature from chapter 2, we observe an improvement. Notably, having stars representing when a daily or weekly goal was achieved improved the participants' experience. Their experience improved because they did not need to compute their steps to know if they had achieved or not the goal as happened to participants in the previous study on chapter 2. Participants had two reasons to increase their steps: 1) contributing to their teams and 2) rewarding their personal goals with stars. Therefore, we believe this explains the increase in their average weekly steps.

Participants in the *R* condition could only see the leader boards of their team and their rivals, but they did not receive trophies when winning the competitions. This could have produced a motivation to compete, but maybe not as strong as the support individuals in *CR* received when their effort was compensated with a trophy. Individuals in *R* did not receive stars for winning their daily and weekly goals, which means that the only element that could push them to walk more was being part of a competition.

Participants in the *C* condition received stars when they achieved their daily and weekly goals. However, they were not part of *team* competition, which means that the only design element that could push them to walk more was the daily and weekly goals and receiving stars when achieving them, which could be observed as a *personal* competition. However, although the weekly goals were changing, the activity could turn monotonous. This is where being part of the team stands out as it causes expectation on participants of what would will the competition bring next.

Furthermore, because *competition* relates to the positive outcome of the *CR* feature combination, it must be noted that exposing individuals to constant competition might produce adverse effects. For example, people might continue to exercise despite acknowledging that this activity is developing physical and/or psychological issues; others might experience negative effects such as anxiety and irritability in the absence of physical exercise [Freimuth et al., 2011]. Also, recent articles show the effects of degrading body image caused by constant social comparison [James Vincent, 2021, Georgia Wells, Jeff Horwitz and Deepa Seetharaman, 2021]. Therefore, employing elements of *competition* might raise concerns and should be handled with care. To this end, some alternative models of competition could be implemented, for instance, *optimal challenge* [Deci, 1975] where the intended activity is tailored to the particular user characteristics. For instance, if a person is used to performing upper body training using 10-pound weights, setting a new weight target of 12 pounds is a reasonable goal. However, it might not be the case for a person who normally lifts 5-pound weights to set a goal of lifting 12 pounds, resulting in an unreasonable expectation. The *optimal challenge* concept is further examined in chapter 1.

3.7.3 What happened to ACR?

As SDT states, supporting the three needs will yield higher well-being and motivation for the individuals. We hypothesized that the *ACR* design would yield the best behavioral outcomes. However, this was not the case. Although we observed an increase on *Exp1* period, the statistical test did not yield a significant difference to *control*. Observing the positive outcomes of *CR*, and looking at *ACR*, made us think that the way the *autonomy* supportive features was designed did not contribute to increasing the average weekly steps but instead, it seems that it produced a detrimental effect.

A future design of the goal-setting feature might incorporate specifying the goal and not only increasing it, allowing participants more flexibility when they can not increase their weekly goal due to personal commitments. Furthermore, this feature might benefit from allowing participants to select the day of the week in which they want to start their physical activity and do not limit it to Mondays as in the present design. Also, the rewards feature may benefit by creating a screen that makes the stars and trophies accumulation more visual. In this way, users observe star and trophy icons increasing in number and not only a star counter as in the present design. Likewise, the peer comparison feature might benefit from including more relevant and diverse information to share. Also, it might benefit from functionalities that allow developing connections with other participants, such as the possibility of exchanging messages.

3.8 Conclusion to the Chapter

This chapter taught us the importance of designing engaging behavior change apps that increase app usage and more prolonged treatment exposure. Also, we learned that our particular combination of the *competence* and *relatedness* features produced the most effects on the physical activity level, intrinsic motivation, and app engagement in our participants when compared to *control* condition. The following chapter presents a study exploring motivational messages and monetary rewards to support physical activity improvements. We conducted this study before the other studies in previous chapters and consequently did not follow the temporal timeline of the rest of the thesis. In this study, we gathered data about alternative implementations of persuasive features. Comparing various feature implementations and combinations of these could reveal how we can translate theory into a design which is the ultimate goal of this thesis.

Third Chapter References

[Abadie, 2005] Abadie, A. (2005). Semiparametric Difference-in-Differences Estimators. *The Review of Economic Studies*, 72(1):1–19.

[Aldenaini et al., 2020] Aldenaini, N., Alqahtani, F., Orji, R., and Sampalli, S. (2020). Trends in Persuasive Technologies for Physical Activity and Sedentary Behavior: A Systematic Review. *Frontiers in Artificial Intelligence*, 3(April).

[Altmeyer et al., 2018] Altmeyer, M., Lessel, P., Sander, T., and Krüger, A. (2018). Extending a gamified mobile app with a public display to encourage walking. In *Proceedings of the 22nd International Academic Mindtrek Conference*, pages 20–29.

[Blundell et al., 2004] Blundell, R., Dias, M. C., Meghir, C., and van Reenen, J. (2004). Evaluating the employment impact of a mandatory job search program. *Journal of the European Economic Association*, 2(4):569–606.

[Cercos and Mueller, 2013] Cercos, R. and Mueller, F. (2013). Watch your steps: designing a semi-public display to promote physical activity. In *Proceedings of the 9th Australasian Conference on Interactive Entertainment: Matters of Life and Death*, pages 1–6.

[Chevance et al., 2020] Chevance, G., Baretta, D., Golaszewski, N., Takemoto, M., Shrestha, S., Jain, S., Rivera, D. E., Klasnja, P., and Hekler, E. (2020). Goal setting and achievement for walking: A series of n-of-1 digital interventions. *Physiology behavior*, 176(3):139–148.

[Cohen, 2013] Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic press.

[Cronbach, 1951] Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *psychometrika*, 16(3):297–334.

[De Cocker et al., 2016] De Cocker, K., De Bourdeaudhuij, I., Cardon, G., and Vandelanotte, C. (2016). The effectiveness of a web-based computer-tailored intervention on workplace sitting: a randomized controlled trial. *Journal of medical Internet research*, 18(5):e96.

[Deci, 1975] Deci, E. L. (1975). *Intrinsic Motivation*. Bibliographie pp.295. Plenum Publishing Company Limited, New York, USA.

[Deci et al., 1994] Deci, E. L., Eghrari, H., Patrick, B. C., and Leone, D. R. (1994). Facilitating internalization: the self-determination theory perspective. *Journal of personality*, 62(1):119–142.

[Edney et al., 2020] Edney, S. M., Olds, T. S., Ryan, J. C., Vandelanotte, C., Plotnikoff, R. C., Curtis, R. G., and Maher, C. A. (2020). A social networking and gamified app to increase physical activity: cluster rct. *American journal of preventive medicine*, 58(2):e51–e62.

[Freimuth et al., 2011] Freimuth, M., Moniz, S., and Kim, S. R. (2011). Clarifying exercise addiction: differential diagnosis, co-occurring disorders, and phases of addiction. *International journal of environmental research and public health*, 8(10):4069–4081.

[Georgia Wells, Jeff Horwitz and Deepa Seetharaman, 2021] Georgia Wells, Jeff Horwitz and Deepa Seetharaman (2021). Facebook Knows Instagram Is Toxic for Teen Girls, Company Documents Show.

[Hanus and Fox, 2015] Hanus, M. D. and Fox, J. (2015). Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Computers & Education*, 80:152–161.

[Haque et al., 2017] Haque, M. S., Abdullah, W. M., Rahaman, S., Kangas, M., and Jämsä, T. (2017). Persuasive health and wellbeing application: A theory-driven design in promoting physical activity. *1st International Conference on Medical Engineering, Health Informatics and Technology, MediTec 2016*.

[Haque et al., 2020] Haque, M. S., Kangas, M., and Jämsä, T. (2020). A persuasive mhealth behavioral change intervention for promoting physical activity in the workplace: feasibility randomized controlled trial. *JMIR Formative Research*, 4(5):1–14.

[James Vincent, 2021] James Vincent (2021). Instagram internal research: 'We make body image issues worse for one in three teen girls'.

[Kuru and Forlizzi, 2015] Kuru, A. and Forlizzi, J. (2015). Engaging experience with physical activity tracking products. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9186:490–501.

[Lacroix et al., 2009] Lacroix, J., Saini, P., and Goris, A. (2009). Understanding user cognitions to guide the tailoring of persuasive technology-based physical activity interventions. In *Proceedings of the 4th international conference on persuasive technology*, pages 1–8.

[Lloyd et al., 2017] Lloyd, G., Munro, S. D., and Arnott, J. L. (2017). Mobile Delivery of Health Information for People with Mild Cognitive Impairment. *Studies in Health Technology and Informatics*, 242:31–37.

[Loeb et al., 2017] Loeb, S., Dynarski, S., McFarland, D., Morris, P., Reardon, S., and Reber, S. (2017). Descriptive analysis in education: A guide for researchers. ncee 2017-4023. *National Center for Education Evaluation and Regional Assistance*.

[Matsumoto and Takenaka, 2021] Matsumoto, H. and Takenaka, K. (2021). Relationship Between Basic Psychological Needs and Exercise Motivation in Japanese Adults: An Application of Self-Determination Theory. *Japanese Psychological Research*.

[McAuley et al., 1989] McAuley, E., Duncan, T., and Tammen, V. V. (1989). Psychometric Properties of the Intrinsic Motivation Inventory in a Competitive Sport Setting: A Confirmatory Factor Analysis. *Research Quarterly for Exercise and Sport*, 60(1):48–58.

[McCallum et al., 2018] McCallum, C., Rooksby, J., and Gray, C. M. (2018). Evaluating the Impact of Physical Activity Apps and Wearables: Interdisciplinary Review. *JMIR mHealth and uHealth*, 6(3):e58.

[O'Brien and Toms, 2008] O'Brien, H. L. and Toms, E. G. (2008). What is user engagement? a conceptual framework for defining user engagement with technology. *Journal of the American society for Information Science and Technology*, 59(6):938–955.

[Oga-Baldwin et al., 2017] Oga-Baldwin, W. Q., Nakata, Y., Parker, P., and Ryan, R. M. (2017). Motivating young language learners: A longitudinal model of self-determined motivation in elementary school foreign language classes. *Contemporary Educational Psychology*, 49:140–150.

[Peters et al., 2018] Peters, D., Calvo, R. A., and Ryan, R. M. (2018). Designing for motivation, engagement and wellbeing in digital experience. *Frontiers in Psychology*, 9(MAY):1–15.

[Plant and Ryan, 1985] Plant, R. W. and Ryan, R. M. (1985). Intrinsic motivation and the effects of self-consciousness, self-awareness, and ego-involvement: An investigation of internally controlling styles. *Journal of Personality*, 53(3):435–449.

[Reinboth and Duda, 2006] Reinboth, M. and Duda, J. L. (2006). Perceived motivational climate, need satisfaction and indices of well-being in team sports: A longitudinal perspective. *Psychology of Sport and Exercise*, 7(3):269–286.

[Ryan, 1982] Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of personality and social psychology*, 43(3):450.

[Ryan et al., 1990] Ryan, R. M., Connell, J. P., and Plant, R. W. (1990). Emotions in nondirected text learning.

[Ryan and Deci, 2017] Ryan, R. M. and Deci, E. L. (2017). *Self-Determination Theory Basic Psychological Needs in Motivation, Development and Wellness*. The Guilford Press, New York, USA.

[Ryan et al., 1991] Ryan, R. M., Koestner, R., and Deci, E. L. (1991). Ego-involved persistence: When free-choice behavior is not intrinsically motivated. *Motivation and Emotion*, 15(3):185–205.

[Shcherbina et al., 2017] Shcherbina, A., Mikael Mattsson, C., Waggott, D., Salisbury, H., Christle, J. W., Hastie, T., Wheeler, M. T., and Ashley, E. A. (2017). Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. *Journal of Personalized Medicine*, 7(2):1–12.

[Spruijt-Metz et al., 2008] Spruijt-Metz, D., Nguyen-Michel, S. T., Goran, M. I., Chou, C.-P., and Huang, T. T. (2008). Reducing sedentary behavior in minority girls via a theory-based, tailored classroom media intervention. *International Journal of Pediatric Obesity*, 3(4):240–248.

[Stragier et al., 2015] Stragier, J., Evens, T., and Mechant, P. (2015). Broadcast yourself: an exploratory study of sharing physical activity on social networking sites. *Media International Australia*, 155(1):120–129.

[Villalobos-Zúñiga et al., 2021] Villalobos-Zúñiga, G., Rodríguez, I., Fedosov, A., and Cherubini, M. (2021). Informed Choices , Progress Monitoring and Comparison with Peers : Features to Support the Autonomy , Competence and Relatedness Needs , as suggested by the Self-Determination Theory. pages 1–18. ACM.

[Wolf et al., 2020] Wolf, T., Jahn, S., Hammerschmidt, M., and Weiger, W. H. (2020). Competition versus cooperation: How technology-facilitated social interdependence initiates the self-improvement chain. *International Journal of Research in Marketing*.

Chapter 4

Pearls of Wisdom

"Money is just an idea."
— Robert Kiyosaki

4.1 Introduction to the Chapter

In this chapter, we introduce a complementary research stream. We implemented a minimal version of a steps counter app with two main features: motivational messages and monetary rewards. In this research stream, we observed how this feature combination negatively affected people's physical activity behavior and motivation.

4.2 Introduction

We all know that exercising is good for us [Barnes, Patricia M., Schoenborn, 2012]. Yet, many fail at giving it the importance it deserves [for Disease Control and Prevention, 2014, Mozaffarian et al., 2015]. We often do not choose how to optimally invest our time because our brains are tricked by, among others, relativity of choices[1], emotions, and social norms [Ariely, 2010]. However, once we understand when and where we make erroneous decisions, we can teach ourselves to think differently and embrace

[1] Humans rarely choose in absolute terms. Rather we look at the relative advantage of one option over another.

healthier behavior. Technology can play a major role in empowering users to overcome their natural limitations and to opt for healthy choices.

Designers have taken this challenge and today we can find thousands of apps on the market that aim at supporting behavior-changing programs. These apps have various purposes: for instance, helping the user to lose weight, to learn a new language, to quit smoking, or to stay hydrated. Many of these applications are based on behavior-changing techniques that are supposed to help users establish healthy routines or cease negative ones. We still lack, however, systematic design guidelines for these persuasive features that can help designers produce more effective apps. For instance, it is not clear which of these techniques is better than the other and for which kind of user.

This challenge can be described with a single question: *what motivates people*? Some people are money driven, whereas others are completely uninterested in material incentives. Some people are extremely sensitive to awareness campaigns, e.g., global warming, and others are unresponsive to them. Theories on human motivation describe motivation in a continuum that goes from extrinsic to intrinsic. [Ryan and Deci, 2000a]. *Extrinsic motivation* supports actions that are executed because there is an element external to the activity that moves the individual to perform the activity (e.g., reading a book because it will be evaluated in an exam next week). Whereas, a person has *intrinsic motivation* towards specific actions that they might find to be satisfying (e.g., the pleasure of reading a book because the content is appealing).

Given these two extremes, app designers developed two classes of interventions: The first class involves *tangible rewards* that are meant to encourage a person by offering benefits external to the activity [Burns and Rothman, 2018]. In the second class, we find *motivational messages* that are supposed to grow the inner values that each person associates with performing an activity [de Vries et al., 2016]. Although the motivational messages might scale easily to support large-scale deployment, the tangible rewards might require subsidies from private companies or governmental agencies.

Examples of applications belonging to the first group are SweatCoin [SweatCoin, 2019], Helsana+ [Helsana+, 2019], ExerciseRewards [ExerciseRewards, 2019], or Healthy Wage [HealthyWage, 2019]. These apps assign virtual points to the users who achieve specific goals (e.g., walking a certain amount of steps, or exercising for a given amount of time). Then, these points can be converted in real money to buy goods or services. Other applications, such as Pact [GymPact, 2019] or DietBet [WayBetter, 2019], propose to the users that they bet on whether they will achieve the goal. If users win the bet, they gain actual money. Other applications follow the other strategy described earlier: they send motivational messages to the users. Examples of apps in this second group include Runkeeper [Runkeeper, 2019], Fitbit [Fitbit, 2019], and Food stand [FoodStand, 2019]. These apps send informative messages at a regular frequency to the users. These messages are designed to provide convincing evidence about why performing the activity could be beneficial and recommendations that can help the recipient set up routines to achieve the target behavior (e.g., exercise more often).

Unfortunately, research is still inconclusive about whether these interventions provide consistent adherence to healthy routines and whether they are equally supportive to all kinds of users. Concerning the financial incentives, researchers still debate the long-term benefits of providing tangible rewards to support changes in behavior. [Mitchell et al., 2013, Barte and Wendel-Vos, 2017]. Furthermore, it is still unclear how to properly design the interventions that involve financial payouts [Patel et al., 2016, Gneezy and Rustichini, 2000]. Similarly, research on motivational messages has not yet contributed strong evidence that regularly providing this information to users leads to the consistent adoption of the target behavior [Kinnafick et al., 2016, Thompson et al., 2014]. Unfortunately, as of this writing, little work has been conducted on the effects that these techniques have on the attitudes of people towards their desired outcomes (i.e., how motivated people are to perform a certain activity). Therefore we posed the following research question: *Will tangible rewards and motivational messages help or hinder people in attaining a more autonomous level of motivation to perform physical activity?*

We conducted a randomized controlled trial to compare the effectiveness of tangible rewards and motivational messages in motivating people to adopt healthy exercise routines. For an application domain, we choose that of *physical activity*, given that a large group of persuasive apps are typically built for this category of use. We compared three experimental conditions: two conditions where participants were offered money to increase their physical activity and one condition where participants were persuaded by motivational messages. The two conditions offering tangible incentives varied in the salience of the prizes being offered to participants. We provide evidence that shows the detrimental effects of the rewards and motivational messages on the users' motivation to perform physical activity. We discuss the implications of these results on the design of persuasive apps. Next, we frame the presented research within the larger body of work conducted in this domain.

4.3 Related work

To explain users behavior, scholars studying the effect of technological interventions on the increase of physical activity[2] refer to psychological theories of human motivation and to behavioral economics studies. Before reviewing studies that are specific to our research, we reviewed the seminal research on which these studies are built.

4.3.1 Behavioral economics studies and behaviour change

As briefly described in the introduction, people generally know that physical activity is good for their health. However, as revealed by surveys, a large portion of the population does not participate enough in these activities (see the CDC report in the U.S. [for Disease Control and Prevention, 2014]). In the last few years, several studies have revealed systematic mistakes that people make when taking decisions and this could explain why people do not engage in physical activity (or PA) as much as they should. These predictable "irrationalities" are often produced by cognitive biases. For instance, researchers found that people tend to favour *immediate* gratification over *delayed* gratification [O'Donoghue and Rabin, 2000]. For example, as the benefits of exercising, for the most part, are perceived only after a certain amount of time (days, weeks, months), the immediate returns of eating a meal might be preferred. An alternative explanation for why people avoid PA is offered by the *projection bias* [Loewenstein et al., 2003]: the tendency to overestimate the normality of our beliefs in comparison to others' beliefs and our future beliefs. Basically, we tend to think that we will feel and act in the future the same as we do now. In the context of PA, this bias makes us prefer options that satisfy current desires or emotional states, instead of pursuing long-term goals.

Other scholars have tried to counter these cognitive biases by designing incentives and other interventions in order to change the perceived benefits of engaging in certain activities. For instance, Loewenstein et al. [Loewenstein et al., 2013], study the design of different incentive schemes to encourage patients to take better care of themselves and find that small and frequent payments are more effective than discounts on monthly premiums or bonuses on their paycheck. Other studies conducted in this direction show that lotteries can be more effective than fixed payments, when nudging

[2] In the remainder of the article we will sometimes refer to physical activity with the acronym PA.

people to adhere to target behaviors [Patel et al., 2016, Patel et al., 2018]. These results are explained by the *Prospect Theory* [Kahneman and Tversky, 1979] that shows that people tend to overestimate small probabilities, and by the *Saliency Theory* [Bordalo et al., 2012] that states that true probabilities might be distorted by consistent payoffs. Both theories contribute to the explanation of why larger wins, which are improbable, might be more motivating than smaller wins that are more likely to occur.

These theories typically fall short in explaining what happens in the post-intervention periods. In this critical phase of these experiments, inconclusive or contradictory findings were reported [Gneezy and Rustichini, 2000]. Therefore, scholars sometimes refer to exogenous factors that would depress the effects recorded during the intervention phase in the post-experimental phase, when the intervention is removed. This is referred to as the *crowding-out effect* [Burns et al., 2012, Moller et al., 2012, Deci et al., 1999]. In this regard, other studies conducted in the domain of psychology provide additional data to better understand what happens after the intervention is removed.

4.3.2 Psychological Theories Explaining Behavior Change

The four theories most used to frame technological interventions in the physical activity domain are the following: Social Cognitive Theory (or SCT), Theory of Planned Behavior (or TPB), The Transtheoretical Model (or TTM) and Self-Determination Theory (or SDT) [Buchan et al., 2012].

SCT [Bandura, 1986] emphasizes the reciprocal causation relationship –people, behavior, environment– in which individuals are actors, as well as products of their environment. It states that behavioural changes occur due to a personal sense of control that is affected by various constructs; and *self-efficacy* is the most important construct. Self-efficacy indicates the belief that a person is capable of executing actions that reflect a sense of control over their environment. SCT posit that to measure *self-efficacy* it is necessary to account for three dimensions: *strength* (a probabilistic judgment of how certain a person is of their ability to perform a specific task), *magnitude* (ordering the tasks by level of difficulty), and *generality* (the extent to which self-efficacy expectations about a particular situation or experience can be applied to other situations). Scholars who have applied SCT within the physical activity domain find that *self-efficacy* is a predictor of physical activity [Dishman et al., 2004, McAuley et al., 2003, Resnick and Spellbring, 2000]. However, these studies only measured the dimension of *strength* but not the other dimensions, resulting in a misinterpretation of the utility of *self-efficacy* effects on PA. This suggests that the construct of self-efficacy might not be a suitable way to study and manipulate PA, which led us to consider other theories for this work.

TPB [Ajzen, 1985] explains the decision-making process individuals go through when setting up a new routine. The theory posits that *intention* is influenced by three constructs: *attitude*, *subjective norm* and *perceived behavioral control*. Attitude indicates the person's evaluation towards the behavior. A subjective norm represents the belief that significant others want the person to engage in the behavior. Perceived behavioral control indicates the person's perception that the behavior is under their control. These three elements determine how strong the *intention* will be and, consequently, the likelihood of the individual to engage in the behavior. After the *intention* is conceived, a period of time occurs before the individual performs the behavior. Hausenblas et al. [Hausenblas et al., 1997] suggest that this relation (intention-behavior) does not diminish over time, whereas Ajzen and Fishbein [Ajzen and Fishbein, 1980] recommend measuring it, as close as possible, from the beginning of the intended behavior. This ambiguity led us to consider that the suggested constructs were difficult to put in practice in the context of a behavior-changing intervention.

The TTM [Prochaska and Di Clemente, 1983] is a practical approach, in six stages, to explaining behavior changes. Each of these six stages represents a temporal period characterized by a behavior pattern and a different degree of the readiness of the person to change. The first stage, *pre-contemplation*, indicates that the individual does not intend to change their behavior in the subsequent months. In the second stage, *contemplation*, the individual has the intention to change their behavior, however does not act yet. In the next two stages, *preparation* and *action*, the individual is committed to engaging in the intended behavior and starts acting however discontinuously. In the fifth stage, *maintenance*, the individual has reached a moment of observable modification in their behavior and has achieved regular patterns. In the last stage, *termination*, the individual has full self-efficacy and is unlikely to return to an unhealthy behavior. The theory also describes the *processes of change*: activities people use to progress through the stages. Some of these include *consciousness raising*, *self-reevaluation* and *self-liberation*. Despite having this precise process of progression, in a systematic review on physical activity about TTM-based studies [Riemsma et al., 2002], the authors find no evidence of a positive effect of stage-based interventions, as opposed to alternative interventions not based on these stages.

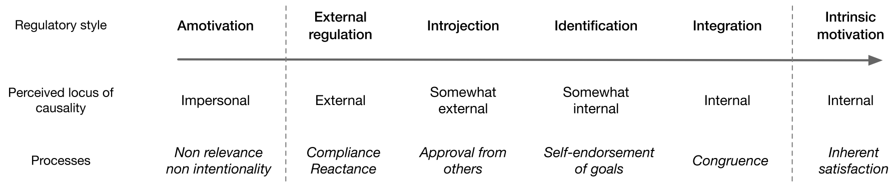

Figure 4.1
Taxonomy of human motivation postulated by SDT (adapted from [Ryan and Deci, 2000a]).

4.3.3 Self-Determination Theory

SDT is a meta-theory that is concerned with social conditions that facilitate or hinder human growth and well-being. The theory examines how specific contextual conditions "*either enhance or undermine the inherent human capacities for psychological growth, engagement, and wellness*" [Ryan and Deci, 2017, p. 3]. SDT research studies factors that facilitate or hinder human motivation, ultimately leading the individual towards vitality or, alternatively, to unhappiness. SDT is an *organismic perspective*, meaning that it offers a unified approach to psychological growth, integrity, and wellness. One of the basic statements of the theory is that "*people manifest intrinsic tendencies to take interest in, deeply learn about, and gain mastery with respect to both their inner and outer worlds*" [Ryan and Deci, 2017, p. 4]. These inclinations will be referred as *intrinsic motivation* in the remainder of this paper.

Furthermore, the theory posits that these tendencies are natural yet conditional: they require contextual conditions to satisfy three basic psychological needs: *autonomy*, *competence*, and *relatedness*. Autonomy is the need to self-regulate one's experience and actions. This provides the individual a form of functioning associated with feelings of volition, congruency, and integration [DeCharms, 1968, Ryan, 1993]. Competence refers to the basic need of feeling effective and having mastery [Deci and Moller, 2005]. Finally, relatedness concerns the feeling of being socially connected and cared for by others [Deci and Ryan, 2014].

Motivation is what moves people to action. SDT postulates that people have not only different amounts of motivation towards a certain activity but also –and more importantly– different *types* of motivation, specifically different orientations with underlying attitudes and goals that give rise to action [Deci and Ryan, 1985]. The most basic distinction is between *intrinsic* and *extrinsic* motivation. The former refers to doing something because a person finds it inherently interesting or enjoyable, whereas the latter refers to doing something because it leads to a separate outcome.

Furthermore, SDT proposes that there are several types of extrinsic motivation that differ in the degree of internalization (i.e., the degree to which the behavioral regulation is autonomous versus controlled). As described by Ryan and Deci "*behaviors can be externally regulated, meaning they are directly controlled by external and self-alien forces; or they can be controlled through introjection, in which case the person has taken but not fully accepted external controls*" [Ryan and Deci, 2017, p. 14]. This source of the regulation of the autonomous behavior plays a very important role in moderating the basic need of autonomy and is often referred to as *perceived locus of causality* (or PLOC, see [DeCharms, 1968]). Therefore, SDT organizes the different types along this *autonomy–control continuum* (see Figure 4.1): On the far left we find *amotivation*, the complete absence of intention to act. Next, we find *external regulation* that describes behaviors that are performed to obtain an externally imposed reward. Farther to the right, we find *introjected regulation* that describes behavior that are performed to avoid guilt. Then we find *identification* that describes a state where the person has accepted the regulation as their own (but not yet integrated). Integration occurs in the last two stages of the continuum: Although in the *integration* stage regulations have been evaluated and are considered congruent with the values, goals, and needs that belong to the self, in the last stage, these become fully assimilated to the self.

According to SDT, extrinsic motivation types can encourage a person to behave a certain way in the short-term, but fail to maintain the behavior over time [Deci and Ryan, 1985]. Furthermore, according to SDT when the needs are satisfied, the individual reaches higher levels of self-determined motivation, consequently moving towards the intrinsic motivation segment of the continuum. In the next section, we will review specific techniques that can be used to satisfy the basic needs described by the theory, and later specifically in the domain of sport and exercise.

4.3.4 Incentives and motivational messages in SDT

SDT research has extensively focused on how contextual conditions affect intrinsic motivation. During the 1970s and 1980s there was a burst of experimental studies on rewards, punishments, feedback, and other extrinsic events. Essentially, these contextual episodes can support a person's sense of *autonomy* and *competence*; this support enhances intrinsic motivation, whereas episodes that affect the same basic needs negatively diminish intrinsic motivation.

From an SDT standpoint, incentives are a form of extrinsic and tangible *rewards*. SDT scholars demonstrated that, under well-specified conditions, rewards can yield detrimental effects on intrinsic motivation. In fact, externally administered rewards can be perceived as coercive and controlling, hence disrupting the basic need of autonomy. In an early study, Deci [Deci, 1971] demonstrate that participants who received money for solving puzzles (i.e., task-contingent reward) showed a decrease in their subsequent intrinsic motivation (measured as a free-choice persistence of the target behavior). In comparison, in a later study, researchers test the use of a monetary reward that is provided regardless of the successful completion of the activity (i.e., task non-contingent reward) and find that this kind of reward does not have any negative effect on intrinsic motivation [Deci, 1972]. Later, Deci and Ryan [Deci and Ryan, 1980, Deci and Ryan, 1985] argue that offering an extrinsic reward (e.g., money) for an activity that was already intrinsically motivated can prompt people to experience an *external perceived locus of causality* [DeCharms, 1968] in their behaviour hence to feel controlled. People exposed repeatedly to extrinsic rewards might come to see the activity as something they did only in order to receive the reward. Furthermore, using rewards to motivate people can signal that the activity is not worth doing for its own sake [Ryan and Deci, 2017, p. 128]. A meta-analysis conducted by Ryan et al. [Ryan et al., 1983] reports results of over 50 experiments that used *completion-contingent rewards*, specifically rewards that were offered when the target activity was completed. The analysis reveals that completion-contingent rewards undermine intrinsic motivation.

Concerning positive *feedback*, SDT researchers found that this can enhance peoples' sense of competence [Deci and Ryan, 1980]. Especially when positive feedback is not expected, people are less likely to think that they completed the activity in order to get the feedback, thus not affecting the PLOC. However, other research shows that when feedback is experienced as an evaluation, pressure or control, it prompts a shift in the PLOC hence undermining intrinsic motivation [Smith and Sarason, 1975]. Additionally, Ryan [Ryan, 1982] demonstrates that simply providing positive feedback is not enough to motivate people if the participants do not also experience autonomy. A more recent study conducted by Hewettand Conway [Hewett and Conway, 2016] finds that, overall, providing positive feedback is not detrimental to intrinsic motivation. However, if the feedback is very salient, it can be perceived as controlling hence hindering intrinsic motivation. By salient, it is implied here that the feedback must be *task-inherent* to enable recipients to gain a sense of their effectiveness.

In summary, external events can affect intrinsic motivation, depending on whether their functional significance is *controlling* (i.e., experienced as an external pressure towards a given goal) or *informational* (i.e., experienced as affirming autonomy and competence). Next, we look into studies that focus specifically on increasing PA.

4.3.5 Incentives and motivational messages aiming at increasing physical activity

Sport and outdoor recreation activities are most often intrinsically motivated because, by practicing them, people satisfy inherent psychological needs [Ryan and Deci, 2007]. Whereas, exercising is often less enjoyable than playing a sport. Yet, people often persist at it perceiving the activity as instrumental to achieving extrinsic goals (e.g., looking attractive) [Ryan and Frederick, 1997]. Several researchers in the exercise domain [Mullan et al., 1997,Standage et al., 2008,Thøgersen-Ntoumani and Ntoumanis, 2006,Teixeira et al., 2012,Owen et al., 2014,Wilson and Rodgers, 2004] find a predictive relation between autonomous types of motivation and physical activity. As we have seen in the previous section, these types of motivation can be assessed with self-reports, such as the *Intrinsic Motivation Inventory* (or IMI [Ryan et al., 1983]). Therefore, we can pose our first hypothesis: *People with a high intrinsic motivation will make more steps than participants with low intrinsic motivation* (*H1*).

Early studies on the use of rewards to motivate athletes showed that these can readily undermine intrinsic motivation [Orlick and Mosher, 1978]. Trophies, prizes and other rewards can either enhance or diminish intrinsic motivation, depending on factors such as the nature of the contingency (i.e., whether delivery is based on performance, or activity completion) and the functional significance they foster (i.e., being perceived as informative or controlling). Unfortunately, often these rewards are used precisely to select the best players but yield negative consequences for the non-selected players. Recently, the use of incentives as an intervention strategy to increase PA has garnered renewed interest. A meta-review by Mitchell et al. [Mitchell et al., 2013] reports the results of several randomized controlled trials where several types of tangible rewards were tested [Charness and Gneezy, 2009,Finkelstein et al., 2008,Daryanto et al., 2010]. Although most of the studies report positive effects of these incentives on PA, there is limited evidence to draw conclusions regarding long-term effects. An exception is the study of Jeffery et al. [Jeffery et al., 1998]: they recorded sustained exercise adherence for more than a year. Unfortunately, in their study they did not isolate the effect of the incentive over training. Furthermore, the most notable weakness of these studies is the lack of observation of the post-intervention behavior. An exception is the study of Charness and Gneezy [Charness and Gneezy, 2009] who find that the increase in gym attendance persisted for 16 weeks (with a 47 USD incentive per week) following a 5-week intervention. Another study that focuses on the post-intervention period conducted Patel et al. [Patel et al., 2016] shows no effect of the intervention in the post-treatment phase; this indicates that more research is necessary to assess the effects of tangible rewards when the money are no longer offered to participants. In particular, there is a strong evidence to support the so-called *crowding out* effect [Burns et al., 2012, Moller et al., 2012, Deci et al., 1999]: the potential for new, external motivators to depress intrinsic motivation and harm post-intervention behavior. SDT states that it is precisely the effect of the intervention that harms intrinsic motivation. Rewards when salient and potent, can clearly motivate *immediate* behavior and still have detrimental effects on people's *subsequent* motivation [Ryan and Deci, 2000b]. Therefore, we pose our second hypothesis: *At the end of the experimental phase, the intrinsic motivation of the participants receiving a tangible incentive will be lower than the intrinsic motivation measured at the beginning of the experiment* [*H2a*]. If the detrimental effect of the manipulation would be validated in the experiment, then this would have in turn an effect on PA. Therefore we posed a related hypothesis focusing on the amount of activity: *During the experimental and post experimental phases participants receiving a tangible incentive will be less physically active than participants in other conditions* [*H3a*].

Furthermore, a body of research focuses on positive feedback provided to people under physical activity training. These works find that information given to trainees can promote autonomous motivation and provide greater vitality thus enhance intrinsic motivation [Moustaka et al., 2012]. Using a technological intervention, Duncan et al. [Duncan et al., 2012] were able to enhance people's motivation with a series of verbal guided-imagery. Also, Carpentier and Mageau [Carpentier and Mageau, 2013] study the effect of *change-oriented feedback* in which constructive, effectance-relevant messages were provided to improve motivation. Their findings show that such feedback predicts positive outcomes for athletes. A few studies focus on SMS-based interventions to improve physical activity (e.g., [Hurling et al., 2007, Kim and Glanz, 2013, Fjeldsoe et al., 2010]). Unfortunately, these studies do not isolate the effect of messages alone, do not include follow-up periods, and lack a clear theoretical underpinning. An exception is provided by the study of Kinnafick et al. [Kinnafick et al., 2016]; they study the effect of supportive text messages on PA. Their study, based on

SDT, tests the effect of sending, on a weekly basis, supportive messages to individuals embarking on exercise programs. The study reveals an effect of the treatment on the levels of intrinsic motivation; it increases from pre- to post-intervention. Furthermore, there is some partial evidence of increase of physical activity (self-reported) between the treatment group and the control group. We therefore pose our research hypotheses, complementing H2 and H3 described above: *At the end of the experiment phase, the intrinsic motivation of the participants receiving motivational messages will be higher than the intrinsic motivation measured at the beginning of the experiment* [*H2b*] and *During the experimental and post-experimental phases, participants receiving motivational messages will be more physically active than participants participating under other conditions* [*H3b*].

In the rest of this paper, we will use interchangeably the terms feedback, informative messages, and motivational messages. We acknowledge that the term feedback is usually used when information is provided in response to a person's performance, whereas informative and motivational messages might not bear any connection with the outcome of the task for which they are given. We will come back precisely to this point in Section 4.6.

4.4 Methodology

To answer our research questions, we chose a longitudinal study because we wanted to study the long-term effects of the tangible rewards and motivational messages. Also, we choose this research method because intra-subject variability is substantially less than inter-subject variability, consequently it is usually more powerful than a cross-sectional study and it provides a more sensible statistical test [Hedeker and Gibbons, 2006]. Therefore, we conducted a ten-month, randomized, controlled trial that included a three-month baseline (pre-phase), four-month intervention (experiment) and three-month follow-up (post-phase). This method enabled us to obtain objective measurements, and to mitigate the effects of possible biases due to the novelty effect of the technology involved in the intervention or the experimental-window-only observations.

Various research studies on human motivation were conducted using the above mentioned methodology (e.g., [Oga-Baldwin et al., 2017, Reinboth and Duda, 2006, Hanus and Fox, 2015]). Studying human PA would require taking a holistic approach, as a person might perform a variety of activities (yoga, swimming, climbing, etc.) as part of their weekly routines. Unfortunately, reliably measuring the amount of PA that span all possible activities is a complex engineering problem for which a solution is yet to be found. In this experiment, we focused on walking (and running) for several reasons:

- It is a universal activity that is naturally performed by most people, even those who do not do sport. In fact, even going to work (or school) involves walking.
- It is a most basic activity that can be performed everywhere even outside of sport facilities.
- Software and hardware solutions for logging walking (or running) have greatly improved in the last few years and are now used to conduct scientific experiments.

In this section, we provide the details of the design of our experiment.

4.4.1 Sampling procedures

A total of 461 potential participants enrolled online for the study and were assessed for eligibility. These people volunteered to be part of a subject pool (consisting of approximately 8K subjects) for behavioural experiments at the University of Lausanne. A specialized unit at our institution, called `Labex` managed the subject pool, took care of the randomization and enrollment processes, automated the transfers of financial incentives, and kept secure contact information of the participants of the study. We captured demographic data through a screener that also served to check whether respondents qualified for the study. Respondents were excluded from the experiment if they were participating (or had in the last six months) in a study involving PA, if during the expected duration of the study they planned to remain without Internet connection for more than 7 consecutive days, or if they foresaw circumstances for which they were unable to move for more than 7 consecutive days (e.g., a planned medical intervention) because this would have produced substantial "holes" in the collected data. Also, we excluded participants if they did not have an `iPhone 5S` or a newer version. The screener was also further used to measure their *Intrinsic Motivation Index* (or IMI, see [Ryan, 1982]) at the beginning of the experiment[3].

[3]In the results, we will refer to this initial measure of the intrinsic motivation as IMI_1; it was recorded at week 14.

This inventory is a multidimensional measurement device intended to assess the participant's subjective experience related to a target activity. It has been used in several experiments related to intrinsic motivation and self-regulation (e.g., Deci et al. [Ryan, 1982], Plant and Ryan [Plant and Ryan, 1985], Ryan [Ryan et al., 1990], Ryan et al. [Ryan et al., 1991, Deci et al., 1994]). The instrument is composed of seven sub-scale scores: the participants' interests/enjoyment, perceived competences, efforts, values/usefulness, pressure and tension felt, perceived choices, and relatedness while performing the activity. Recently McAuley et al [McAuley et al., 1989] examined the validity of the IMI and found strong support for its validity. After the eligibility selection, 179 respondents were excluded due to non-compliance with inclusion criteria, leaving 282 participants that were randomly assigned to each experimental condition. A CONSORT diagram showing participant progress through the study is shown in Figure 4.2. During the first weeks of the experiment, we discovered that older models of the iPhone that we originally considered compatible with the software needed for the study were, in fact, unable to track steps. Specifically, the iPhone 4S, 5 and 5C do not have the M7 Motion Sensor co-processor that tracks steps. Therefore, we had to exclude 53 participants, as they could not take part in the experiment. This number is listed in 'Technical issues' in Figure 4.2. Also, during the analysis of the data, we noticed that 12 participants did not correctly log their steps. In fact, these users had around $\sim$40% of missing data. We contacted some of these participants and learned that, in some cases, they were not in the habit of bringing their mobile phones on campus and, in other cases, they owned multiple mobile phones and had installed the research app on a secondary device. Therefore, we decided to exclude these participants from the analysis (see Figure 4.2). In addition to participants that we had to exclude for technical reasons, there were also 9 people who dropped out during the Experiment phase and 23 who stopped participating during the Post-phase, which results in a 14% attrition rate. The study was therefore conducted with a total of 220 participants, whereas the analysis was conducted with 208 participants. The sample size is discussed in Section 4.4.8 below.

4.4.2 Participants Characteristics

The participants sample included 220 students,[4] from the University of Lausanne, Switzerland, who were 18 years of age or older (M = 21.6, SD = 2.29), among whom

[4]Participants sampling is one of the limitations of our study. This is discussed in Section 4.6.2.

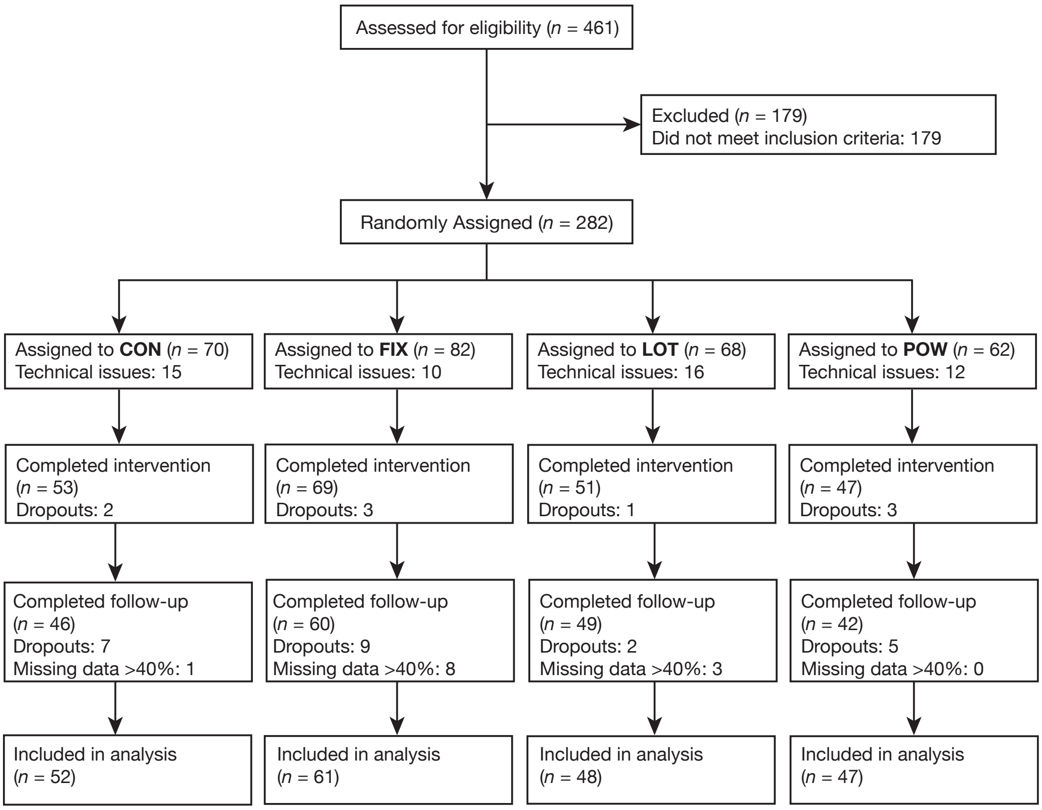

Figure 4.2
CONSORT flowchart showing participants progress through the study.

131 were women (59.5%), with an average IMI level of 4.8 (SD = 1.26)[5], see Table 4.1. To ensure a mix of technical and non-technical backgrounds, the students came from all seven faculties of the university. Through the enrollment questionnaire, we also measured additional baseline characteristics of the sample, such as the number of fitness activities they performed[6], whether they used a fitness tracker, whether they smoked and, finally, we measured a Socio-Economic Status indicator (SES)[7]. After randomly assigning the students to the experimental conditions, we checked an even distribution of the different characteristics across the conditions.

[5] IMI is expressed in a scale from 1 to 7. Low scores indicate a lack of intrinsic motivation.

[6] This is expressed on a scale from 1 to 7 and it was based on the amount of PA practiced by the respondent. Low numbers indicate a lack of fitness activities.

[7] This is a compound metric measured with 5 standardized items. For a full description of the instrument the reader can refer to Oesch [Oesch, 2006].

[8] IMI reported in the table was measured at week 0 and it is later referred to as IMI_1.

Table 4.1
Characteristics of the study participants

	Control Group	FIX	LOT	POW	Total
Number of participants n (%)	52 (25)	61 (29.33)	48 (23.08)	47 (22.60)	208 (100)
Women n (%)	28 (13.46)	35 (16.83)	29 (13.94)	32 (15.38)	124 (59.62)
Age (years) mean (sd)	21.60 (2.11)	21.92 (2.35)	21.65 (2.18)	21.28 (2.52)	21.63 (2.29)
IMI[8] mean (sd)	4.62 (1.16)	4.90 (1.15)	5.11 (1.29)	4.79 (1.47)	4.85 (1.27)
Fitness Level mean (sd)	2.53 (0.86)	2.53 (0.71)	2.55 (0.71)	2.36 (0.73)	2.50 (0.75)
SES mean (sd)	2.55 (0.68)	2.51 (0.66)	2.53 (0.75)	2.61 (0.83)	2.55 (0.72)
Fitness Tracker n(%)	4(1.92)	11(5.29)	9(4.32)	8(3.85)	35(15.9)
Smoker n(%)	6(2.88)	10(4.81)	0(0)	9(4.32)	27(12.3)

4.4.3 Experimental Design and Interventions

We used a between-subject design, with three conditions (FIX, LOT and POW) plus a control condition (CON). Participants were electronically randomly assigned to the control group or to one of the three intervention groups. The final number of participants in each group can be found in Table 4.1. The imbalance is due to exclusions, as explained in Section 4.4.1.

Similarly to Patel et al. [Patel et al., 2016], FIX and LOT conditions were designed to provide tangible rewards. Participants of these conditions were assigned money as detailed below. Accrued incentives could be claimed at the end of the Experiment phase. Similarly to Kinnafick et al. [Kinnafick et al., 2016], the POW condition was designed to deliver motivational messages. All participants were required to install a research app on their mobile phone (described in 4.4.6). During the experiment phase, regardless of the condition in which they participated, all participants received a pop-up notification at 8am on their smart phones, reminding them the goal for the day (see an example in Figure 4.3a). The experimental conditions differed in terms of the additional pop-up notifications that were sent to communicate to participants about the incentives specific to their treatments. Incentives, and their related notifications, were offered only during the 19 weeks of intervention. However, the monitoring of the PA continued for the entire 30 weeks.

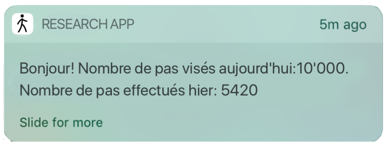

Figure 4.3a
Daily Goal Notification Message (Good morning! Target steps for today: 10000. Yesterday you did: 5420)

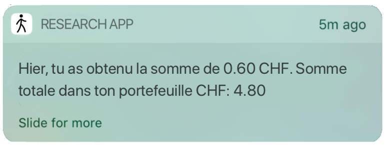

Figure 4.3b
Wallet Status Notification Message (Yesterday, you have been assigned 0.60 CHF. Total CHF in your wallet: 4.80)

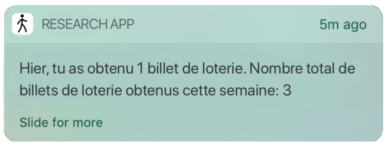

Figure 4.3c
Lottery Notification Message (Yesterday, you have been assigned one lottery ticket. Total number of tickets this week: 3)

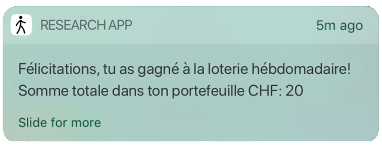

Figure 4.3d
Lottery Price Notification Message (Congratulations, You won the lottery! Total CHF in your wallet: 20)

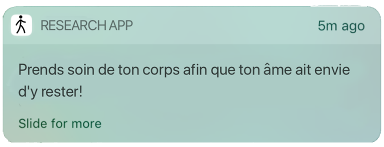

Figure 4.3e
Example of Motivational Message Notification (Take care of your body, it's the only place your soul can live in.)

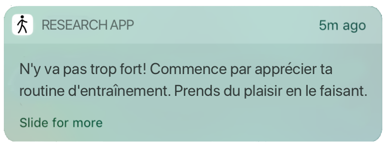

Figure 4.3f
Example of Motivational Message Notification (Don't push yourself too hard! Start off by getting a feel for your exercise routine. Make the exercise itself a joy!)

From the SDT standpoint, rewards can detrimentally affect a person's intrinsic motivation. Also, externally administered rewards can be perceived as coercive and controlling hence hinder the basic need of autonomy.

Lottery Condition (or LOT): Participants in this group were assigned a weekly lottery ticket every time they reached their goal of daily steps. At the end of each week, a winning ticket was drawn from all the tickets issued that week. The owner of the ticket was credited 10 CHF on their *digital wallet*. Participants who were assigned a lottery ticket received a notification the following day at 10am; this reported the total number of tickets for the week (see Figure 4.3c). Winners of the lottery were notified each Monday morning at 10:30am (see Figure 4.3d). This lottery design was tested successfully in prior work [Porter and Whitcomb, 2003, Patel et al., 2016, Patel et al., 2018]. Similarly to the FIX condition above, a lottery ticket can be considered a form of reward that hinders the basic need of autonomy.

Motivational Message Condition (or POW): Every day at 8:30 am, participants in this group received a motivational message pulled sequentially from a pre-arranged corpus (see two examples in Figure 4.3e and 4.3f). These messages were delivered as smart phone pop-up notifications. The method with which these messages were created is described in Section 4.4.5 below. These messages were delivered to all participants in the condition POW every morning, regardless of whether they had achieved their goal of daily steps. From an SDT perspective, these messages can help the users reflect on the reasons they want to engage in the specific activity hence have the potential to support the basic need of autonomy.

Control Condition (or CON): Participants in the control group did not receive any other notifications, aside from the daily morning message announcing the goal for the day, described at the beginning of this section (see Figure 4.3a). Participants in this group did not receive either any additional incentives for achieving their goal of daily steps.

4.4.4 Daily Goal

The main task all participants were asked to complete was to walk 10K steps a day. This target reflects several deliberate design elements. This value has been proven by medical research to provide important health benefits such as lowering blood pressure, increasing exercise capacity and fostering an active life [Iwane et al., 2000, Tudor-Locke and Bassett, 2004]. This value was also used in prior studies aimed at increasing the daily number of steps (see for instance [Bravata et al., 2007]). Furthermore, this value was also set from a pilot study involving a sample of 20 beta-testers, where we found that the average steps testers walked was 6K (with a standard deviation of $\sim$3K steps), making it reasonable to request that they walk 4K more steps than their average number of steps (this corresponds to an additional 40 minutes of walking at regular walking speed). As FIX and LOT required a threshold to assign prizes to the participants, we set the same target for all conditions (i.e., walking 10K steps/day) in order to avoid possible co-founds that affect only some conditions and not others. Indeed, providing a specific goal to participants is in itself another motivational strategy that was shown to influence performance [Locke and Latham, 2002]. Goal-setting theory posits that goals should be, among other qualities, adaptable, challenging but feasible, and self-set by the individuals. In the context of this experiment, it is important to notice that (i) the goal of walking 10K steps a day was designated by the organizers rather than self-set, and (ii) it was not adaptable on the specific competences of each participant; basically it was the same for everybody. Using SDT-specific language, we find that setting a goal the way we did in the experiment did not satisfy the basic needs of Autonomy, and Competence. Hence, we expected a negligible effect on the experimental design. Of course, it must be noted that setting an unobtainable goal might be demotivating. We will come back to this point in the results and discussion sections.

4.4.5 Motivational Messages Corpus

As we could not identify any publicly available corpus of motivational messages based on SDT, we looked for other sets that could be of use for the experiment.

We started from the set developed by de Vries et al. [de Vries et al., 2016]. They devised a method to generate motivational messages to increase physical activity by using Mechanical Turk workers. To obtain motivational messages from the Mechanical-Turk workers, the authors developed five scenarios, each about a different stage of change described by the *Trans-theoretical Model (TTM)* [Prochaska and Di Clemente, 1983], then they asked $\sim$500 workers to generate six motivational messages for each scenario. The resulting corpus contained $\sim$3000 messages. The messages of the original corpus were organized according to the stages of change (see Table 4.2). This model is typically employed in physical training when new activity routines are set up [Geller et al., 2011].

Table 4.2
The stages of change of the TTM with a short description (from de Vries et al. [de Vries et al., 2016])

Stage of change	**Description**
Pre-contemplation (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 of the pros of a change in behaviour, 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 their lack of activity.
Action (A)	The individual has changed, but for or since no more than 6 months. Individuals in this stage have 'changed', but have not reached the duration that exemplifies a real change in behaviour.
Maintenance (M)	The individual has changed, longer than 6 months. Individuals in this stage have changed and are working not to relapse.

The first two authors of this paper coded each message in the original corpus, according to how motivating they considered the message to be, on a scale from 1 ("Very demotivating") to 5 ("Very motivating") with a 3 as neutral. The two measures were then averaged. We used this information to rank the original messages, according to their effectiveness (as perceived by the first two authors).

Next, we selected messages for our corpus from the messages in the original corpus that were ranked highest in each stage of change. From these, we further excluded messages that were not relevant to walking/running.

We decided to include pre-contemplation messages in the corpus as these could have been relevant to participants who might have elected to participate in the study only for obtaining the economical incentive but that might have not been willing to change their routines around walking/running.

Then we organized the messages in chronological order so that *Pre-contemplation* messages were listed at the beginning of the set, *Contemplation* messages followed, and so forth. We did so in order that, during the first weeks, the participants would receive encouraging messages to establish a routine and that, later in the experiment, the messages would be aimed at supporting the participant in maintaining their routine. The original corpus also contained another code (orthogonal to the first), describing the processes of change, i.e., ten processes usually experienced when successfully progressing through the stages of change (for the details, we defer the reader to the description reported in de Vries et al. [de Vries et al., 2016]). When selecting the messages belonging to each stage of change, we ensured that collectively they covered all 10 different processes of change. This procedure provided us a diversified set of messages that followed the different stages of change. Incidentally, in the final set, we ended up selecting the 50 messages used by de Vries et al. [de Vries et al., 2016] in the second part of their study [de Vries et al., 2016, p. 303].

Next, the messages were translated from English to French. The translation was curated by two independent translators who double checked the meaning of each message and the accuracy of the translations. The resulting selection contains 150 motivational messages that ensured full coverage of the experiment phase. Before deployment, we considered the differences between the audience used by de Vries et al. [de Vries et al., 2016] to generate the original corpus and our target audience. Mechanical-Turk workers were all based in the US, their average age was 31 years (SD = 9 years), and the large majority had a college degree. Whereas, our participants were slightly younger and were all based in Switzerland. To make sure the selected messages could be used with our audience, we asked two students from the university to look through the selected messages and flag any content that did not make sense to them or that they would otherwise consider inappropriate. All messages in the corpus were validated by these two students.

4.4.6 Apparatus

There is a growing number of specialized devices that have been developed to sense human activity. These include bracelets[9], last retrieved June 2018., clips[10], last retrieved June 2018. or wearable sensors[11]. As consumer products, these devices are great, however, as research devices, they suffer serious limitations. For instance, users might forget to consistently wear them, or their batteries can run out of power if not systematically recharged. These points, among others, persuaded us to consider employing smart phones, without any additional hardware, as a tool for tracking steps. Smart phones can accurately predict walking activity [Nolan et al., 2014] and are used to measure physical exercise [Liu and Chan, 2016, Harries et al., 2016, Patel et al., 2016]. They have the following advantages: They act as silent observers, letting participants carry on with their tasks without explicit reminders that their activity is being tracked, thus making the data capturing less intrusive. Furthermore, people are willing to carry their phones with them during the majority of their daily activities and are also aware of the battery level because they want to remain active in their social networks and be reachable via instant messaging and calls. The fact that most of daytime period users carry their phones diminishes the negative factors that usually come with the usage of wearable trackers. Finally, another methodological advantage of using smart phones as sensors is that they do not require participants to use additional hardware, thus increasing the *ecological validity* of the study.

Of the different hardware solutions, we focus on the `iPhone` because `Apple` has standardized both the hardware and the API that can be used to collect activity data, thus reducing the development costs required to build the experimental infrastructure for a study. Apple's `HealthKit` is a platform with a repository of physical activity data collected from the iPhone's internal accelerometer. The data collected is stored in an encrypted database called `Healthkit Store` from which step counts are retrieved[12]. Researchers have made use of this platform in the past to implement step tracking applications [Frank, 2017, Jeong et al., 2017, Xu et al., 2017]. One of the advantages of using this service is that activity tracking always runs in the background and is activated by default on the iPhone (unless the user changes this preference in the

[9] See for instance https://www.polar.com/en/products/lifestyle/loop

[10] E.g., https://www.fitbit.com/zip

[11] E.g., https://www.suunto.com/, last retrieved July 2018.

[12] HealthKit also collects data from a variety of other sensors. However, in the context of this paper we focus only on step count measurements.

phone's settings). As HealthKit also permits steps to be added manually, this could have led some participants to cheat. Fortunately, the HealthKit store enabled us to distinguish between readings coming from the sensor and manually entered steps[13]. Steps that were manually added to the HealthKit store through the iPhone `Health` app or 3rd party apps were ignored by the research app described in this paper and not reflected in the daily stepcount.

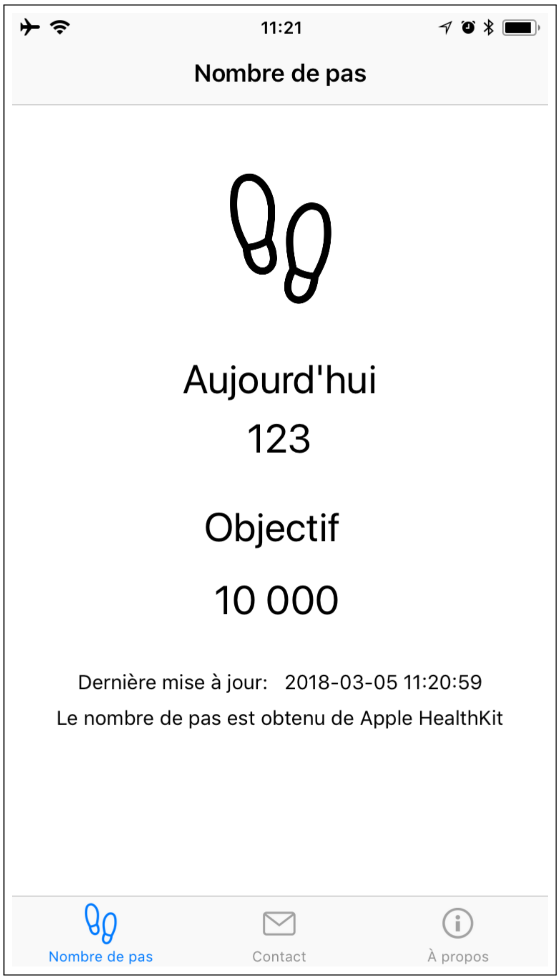

Figure 4.4a
Home screen of the Research App (Number of steps; Today 123; Objective 10000; Last update; The number of steps is obtained from Apple HealthKit)

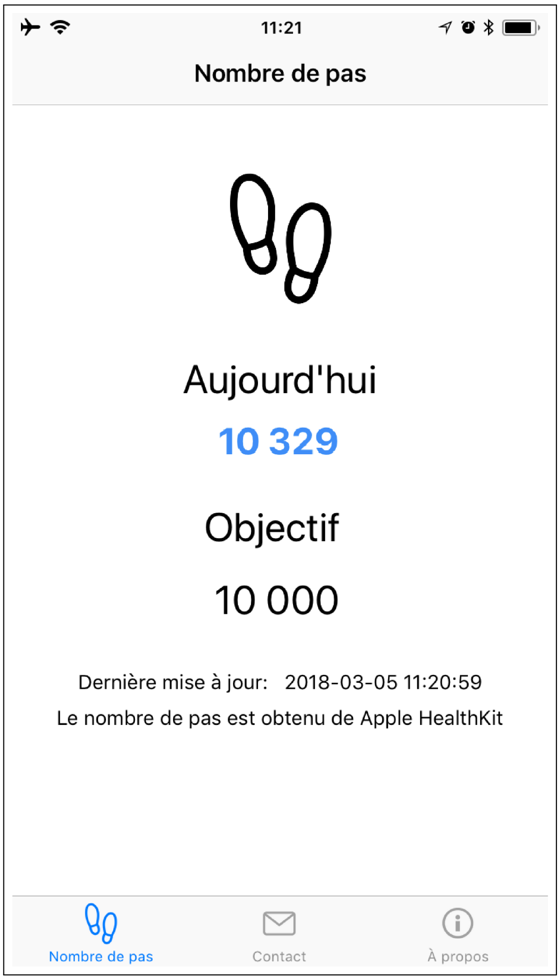

Figure 4.4b
Home screen when reaching the daily goal of 10K steps (Number of steps; Today 10329; Objective 10000; Last update; The number of steps is obtained from Apple HealthKit)

To implement the different experimental conditions, we developed an app that ran on `iPhone 5s` or a later version, with an OS 8.0 or later version. The app provided three views that were accessed through the tab bar at the bottom of the screen (see Figure

[13] For the technical reader, entries were filtered using the `NSPredicate` logical condition.

4.4a). Right after installation, the app presented two requests for permissions from the user to ensure that it could display pop-up notifications (the main communication channel for receiving information about the incentives accrued in the study) and that it could retrieve step data from the Healthkit Stores. The main view of the app provided information on the user's current number of steps for the current day (see Figure 4.4a). When the daily goal was reached, the number of steps became bold and its color changed to blue (see Figure 4.4b). As HealthKit does not constantly refresh the step counts, we could not instantly update the number on our research app. To avoid misunderstandings with participants, the app indicated at the bottom of the screen the last date and time when the steps were refreshed. The second view displayed a contact e-mail to communicate with the research team, and the third view showed the version and copyright information about the app. We intentionally chose a minimalist design to keep the UI elements similar across experimental conditions, thus controlling for confounding effects of these elements. The purpose of the research app was to deliver the notifications and retrieve the activity logs from the participants' phones. Daily step counts were synced with a secure database on servers at our institution. Once the app was installed, participants were not required to reopen the app, although they could as often as they wished to check their daily progress. We also implemented a dashboard, as part of the overall architecture, through which we could monitor whether specific participants turned off notifications or stopped sending data; in this case, we could preemptively attempt to solve the emerging technical issues.

Before deploying the final version of the app, we ran a month-long pilot involving 20 beta testers. This study was useful for resolving bugs, for testing the delivery of the notifications and the accuracy of the step-count measurements.

4.4.7 Procedure

All selected participants received an invitation e-mail, together with an informed consent. The consent form was adapted for the experimental condition the participant was assigned to. This document specified the conditions for participation and for winning a reward (i.e., simply making 10K steps for participants in FIX, or winning the lottery for LOT), if any. All participants were requested to download from Apple App Store the research app described in the previous section. They were also asked to carry their mobile phone on them during their daily activities (for example in a arm band, pocket, or belt clip).

The study was designed to be run across various moments of the academic year (i.e., fall semester, winter holidays and spring semester), which allowed various walking behavior observations. It began early September 2017 (week 15) and finished late March 2018 (week 44). Towards the end of the experiment phase (week 26), participants were asked to fill an intermediate questionnaire that presented the same standardized inventory used in the screener. The compound score was labeled IMI_2. The same instrument was again deployed at the end of the study (week 44) through which we collected the IMI_3 scores. At the onset of the Experiment phase, the research app extracted and recorded the step count from 1st of June 2017 (week 1). The data of week 1 to 14 served as a baseline and is presented in the remainder of this paper as the Pre-Experiment phase.

All participants were paid 10 CHF for participating in the experiment, specifically for installing and maintaining the research app on their phones for the duration of the study. In addition to this incentive, the participants who filled out the questionnaires through which we collected intrinsic motivation measurements were given the opportunity to participate in a lottery draw, with the possibility of winning a prize of 100 CHF ($\sim$ 100 USD). They were allowed to withdraw from the study at any point in time and without any consequences. The participants were identified through an ID, and their identity and contact information were separated from the steps data collected through the experiment. The University of Lausanne Institutional Review Board approved the study.

4.4.8 Statistical Analysis

Several analyses were performed. Their main characteristics are highlighted below. Data were analyzed with generalized mixed-effect regression models that are now widely used for the analysis of longitudinal data (see for example Fitzmaurice et al. [Fitzmaurice et al., 2008]). For each outcome and model presented below, the results, reported in Section 4.5, include the unadjusted effects –means without controlling for the model factors– and the adjusted ones –controlling for the model factors.

The Outcomes

The first primary outcome was the participant IMI index measured at the onset of the study (IMI_1, weeks 0), in the experiment phase (IMI_2; weeks 12), and post-phase (IMI_3; weeks 30). The model for IMI index is presented in Section 4.4.8 below.

The second primary outcome was the average step count per day over each week during the experiment phase and post-phase. The measurements during the pre-phase were considered as secondary. The model for step counts is presented in Section 4.4.8 below.

Another secondary outcome is the binary indicator of whether a given participant in a given day reached the goal of 10K steps (we will refer to this as 'participant-day'). This enables us to analyze the influence of the IMI score and conditions on the proportion of participant-day, where they reach the predefined goal, and to see if it is in line with the results for the step counts. The model for the goal achievement is presented in Section 4.4.8 below.

Missing values

For each participant on each day (or participant-day), the observed daily number of steps could be missing if the participant turned off their smart phone for the entire day, left the smart phone at home, or if the application did not grant the permission to read the steps from HealthKit. In addition, the number of steps per participant-day that were larger than 54K or lower than 600 were considered unrealistic and replaced by a missing value code (i.e., `NA`). Although the upper-bound corresponded to running a marathon and rarely occurred (6 times, for 5 people), the lower-bound was caused by a number of unspecified factors (e.g., placing the phone in a purse, or overload of the CPU of the phone).

The measurement errors of physical activity by the accelerometers embedded in the mobile phones was also identified by prior research [Patel et al., 2017]; it demonstrated that, over the course of a several months-long study, these measurements errors have negligible effects.

Therefore, we conducted an analysis of the missing data by comparing each condition in each phase of the experiment. However, we did not conduct any statistical test on missing data because any test based on such a large number of observations would result significant, and this could mislead the reader. We thus conducted the analysis by simply computing the proportions. The results are reported in Table 4.3. In the Pre-phase, we see that the proportion of missing participant-days is similar between groups CON, FIX and LOT; whereas, in POW this proportion is slightly lower. In the Experiment phase, the proportions are overall similar between conditions and smaller than those in the Pre-phase. In the Post-phase, the proportions increased to larger values, especially in the group POW. The differences between the Pre-phase and the Experiment phase are expected, as several participants had intentionally disabled the automatic capturing of their steps, hence creating more missing data (similarly across all conditions). Similarly, the differences between the Experiment and the Post-phase are due to the participants who dropped out from the experiment, which resulted slightly more frequent in the POW condition.

Table 4.3
Analysis of the missing data in each condition and phase of the experiment. SMP represents the number of observations. These numbers are computed from the 208 participants included in the full analysis reported in this paper (cf. Sec. 4.4.1).

	Pre-phase (wk 1–14)			Experiment (wk 15–33)			Post-phase (wk 34–44)		
Condition	SMP	NAs	Prop	SMP	NAs	Prop	SMP	NAs	Prop
CON	5044	910	0.18	6656	376	0.06	3848	710	0.18
FIX	5917	990	0.17	7808	506	0.06	4514	1206	0.27
LOT	4656	844	0.18	6144	411	0.07	3552	851	0.24
POW	4559	628	0.14	6016	431	0.07	3478	1123	0.32

IMI model

The adjusting factors are the participant effect (random intercept), the condition (categorical: CON, FIX, POW, LOT), the phase (categorical: experiment, post-), and their interaction. The research hypotheses are assessed using multiple comparisons between conditions at each phase, and between phases at each condition.

Step counts model

Because the trend in time (week effect) is nonlinear, the step counts were analyzed in two different stages. First, a smoothing spline was fitted to estimate the weekly average step counts. The smoothing parameter was selected by cross-validation. Second, the spline was subtracted from the average step counts before the regression analysis was run. The adjusting factors are the participant effect (random intercept), the IMI_1 level (numerical), the condition (categorical: CON, FIX, POW, LOT), and the interaction between IMI_1 and the condition. The research hypotheses were assessed using multiple comparisons between conditions at several IMI_1 values (i.e., 2, 3.5, 4.5, 6). Also, the IMI_1 effect (the slope) was assessed at each condition.

Goal model

A generalized linear model with binomial family was fitted. The adjusting factors were the same as for the step-count model. The estimates are given in *odds ratios* (for the unfamiliar reader, see for example Szumilas2010 [Szumilas, 2010]).

Statistical software

All the analyses were performed using the `R` packages `lme4` and `lmerTest` [Bates et al., 2015]. The mixed-effect models were fitted using the `REML`. All the multiple comparisons were obtained using the R function `lsmeans` [Lenth, 2016]. Dunnett's correction was applied for comparison to the reference level. The smoothing spline was fitted using the `R` function `smooth.spline`.

Limitations of the statistical analysis

For the step counts, separating the analysis into two stages can bring bias in the results because the adjusted analysis does not incorporate the uncertainty of the smoothing spline estimates. We chose this for clarity of exposure and assumed that this bias was small. The results were checked using a generalized additive mixed model (R function `gamm`) [Wood, 2017]) that can incorporate both analyses at once, but whose presentation of results is more complex.

For the adjusted analysis, using the weekly averages of the daily counts, instead of the daily counts directly, can bias the analysis in the sense that the number of observations (i.e., full data) for building the averages are not taken into account. This choice was made for ease of presentation. Results were checked on the full data, using the day of the week (Monday through Sunday) as an additional factor. The models lead to the same conclusions, but some cases could not be treated because of numerical limits (non-convergence).

For the mixed-effect models, all the inferential results (p-values and confidence intervals) assumed the normality of the residuals. In addition, missing at random is assumed.

Power Analysis

Similarly to [Patel et al., 2016], we made an approximate power analysis for the step counts, using a two-sided paired t-test based calculation with a Bonferroni adjustment of the α level to 0.017 (CON versus FIX, POW and LOT). With a power of 80% and the participant design in each condition (see Table 4.1), an effect size of 0.60 can be detected. With an estimated pooled within-group standard deviation of 3306, a true contrast of 1984 steps between two conditions can be detected.

For the IMI analysis, the Bonferroni adjustment implied an α level of 0.025 (IMI_1 versus IMI_2 and IMI_3). With a power of 80% and 112 responses (see Table 4.6), an effect size of 0.295 can be detected. With an estimated pooled within-group standard deviation of 1.23, a true contrast in IMI of 0.36 can be detected.

4.4.9 Physical Activity and Outdoor Weather Conditions

It is reasonable to consider the weather conditions as a predictor for the physical activity of the participants. Indeed, we attempted to use them in the models to explain the step counts and the daily-goal achievements. In particular, we considered the daily mean temperature (in Celsius)[14]. We retrieved the data from the Federal Office of Meteorology and Climatology in Switzerland[15].

Although the data carries some predictive power, this was not useful to estimate the differences in physical activities between groups in the experiment. This can be explained by two observations: (1) all the participants are exposed to the same weather conditions; (2) the weather conditions are highly collinear with the time trend

[14]We also considered the total daily precipitations (in millimeters), however as the set contained mostly zeroes we decided to exclude these from the analysis.

[15]See https://gate.meteoswiss.ch/idaweb, last retrieved March 2019.

Table 4.4
Unadjusted daily average number of steps

	Control Group	FIX	LOT	POW	Total
Pre-					
mean	7383.3	6916.1	7193.8	7793.1	7299.2
sd	5592.8	5053.8	5517.6	5996.2	5531.7
95% CI	(7212.8-7553.8)	(6775.0-7057.2)	(7018.6-7369.0)	(7605.6-7980.5)	(7215.6-7382.8)
Experiment					
mean	7392.8	6707.8	7055.7	7151.5	7060.2
sd	4731.2	4110.0	4759.0	4575.2	4535.7
95% CI	(7275.8-7509.8)	(6613.5-6802.1)	(6932.5-7178.9)	(7031.5-7271.5)	(7003.8-7116.5)
Post-					
mean	6696.3	6317.1	6602.3	6887.4	6604.3
sd	4617.8	4340.4	4582.1	4535.6	4518.3
95% CI	(6534.8-6857.9)	(6169.2-6465.1)	(6429.5-6775.1)	(6704.1-7070.6)	(6521.7-6686.9)

Table 4.5
Unadjusted proportions of participant-days reaching 10′000 steps

	Control Group	FIX	LOT	POW	Total
Pre-					
mean	0.231	0.206	0.234	0.252	0.231
sd	0.425	0.405	0.424	0.434	0.421
95% CI	(0.223-0.249)	(0.195-0.218)	(0.221-0.248)	(0.238-0.265)	(0.224-0.237)
Experiment					
mean	0.231	0.189	0.205	0.194	0.204
sd	0.422	0.391	0.403	0.395	0.403
95% CI	(0.221-0.242)	(0.180-0.198)	(0.194-0.215)	(0.184-0.204)	(0.199-0.209)
Post-					
mean	0.171	0.155	0.173	0.184	0.170
sd	0.377	0.362	0.378	0.388	0.375
95% CI	(0.158-0.184)	(0.143-0.168)	(0.158-0.187)	(0.169-0.200)	(0.163-0.176)

component of the LMER analysis. In order to keep the models simple, we did not include weather conditions in the models described above.

Table 4.6
Unadjusted IMI measured at week 0 (IMI_1), week 12 (IMI_2), and week 30 (IMI_3)

	Control Group	**FIX**	**LOT**	**POW**	**Total**
IMI_1					
n	52	61	48	47	208
mean (sd)	4.62 (1.16)	4.90 (1.15)	5.11 (1.29)	4.79 (1.47)	4.85 (1.27)
95% CI	(4.30-4.94)	(4.61-5.19)	(4.75-5.48)	(4.36-5.21)	(4.68-5.03)
IMI_2					
n	46	55	41	43	185
mean (sd)	4.96 (1.16)	4.57 (1.15)	4.92 (1.22)	4.48 (1.40)	4.73 (1.24)
95% CI	(4.62-5.30)	(4.26-4.88)	(4.55-5.30)	(4.06-4.90)	(4.55-4.90)
IMI_3					
n	35	28	20	18	101
mean (sd)	5.13 (1.41)	4.89 (0.99)	4.64 (1.18)	5.15 (1.32)	4.97 (1.24)
95% CI	(4.65-5.60)	(4.52-5.27)	(4.11-5.17)	(4.52-5.78)	(4.73-5.21)

4.5 Results

The Participants' baseline characteristic means and standard deviations were similar across the four experimental groups, which implies that the groups were well balanced after randomization and exclusions (see Table 4.1).

During the Pre-Experiment phase, the average daily step-count for all participants was 7299.2 steps (95% CI 7215.6 to 7382.8). During the the Experiment phase, the same average was 7060.2 steps (95% CI 7003.8 to 7116.5). Finally, in the Post-Experiment phase, this average was 6604.3 steps (95% CI 6521.7 to 6686.9). See Tables 4.4 for the details. Figure 4.5 shows the average daily step-counts per week and the non-linear trend fitted as described in Section 4.4.8. The figure shows a global downward trend during the Experiment phase with a drop around weeks 30-32 (end of December 2017, beginning of January 2018). The overall trend can be attributed to the fact that the experiment was held during different times of the academic year. The Pre-phase began during the exams of the spring semester, then it overlapped the vacation time for most students. The Experiment started with the beginning of the fall semester and culminated at Christmas and the winter-exam period. We observed that students typically tend to exercise less during the Christmas break and around the exam periods. During the Experiment, the mean proportion of days that the participants achieved the

10K steps goal was 0.204 (95% CI 0.199 to 0.209), see details in Table 4.5. These statistics confirm that the 10K-step goal was indeed a reasonable target that could be attained by the majority of participants in the sample, with roughly 30 additional minutes of walking/running per day[16]. Below, we assess our hypotheses by using the adjusted estimates from the models and inferential results. All the statistical significances are meant with an alpha level of 0.05.

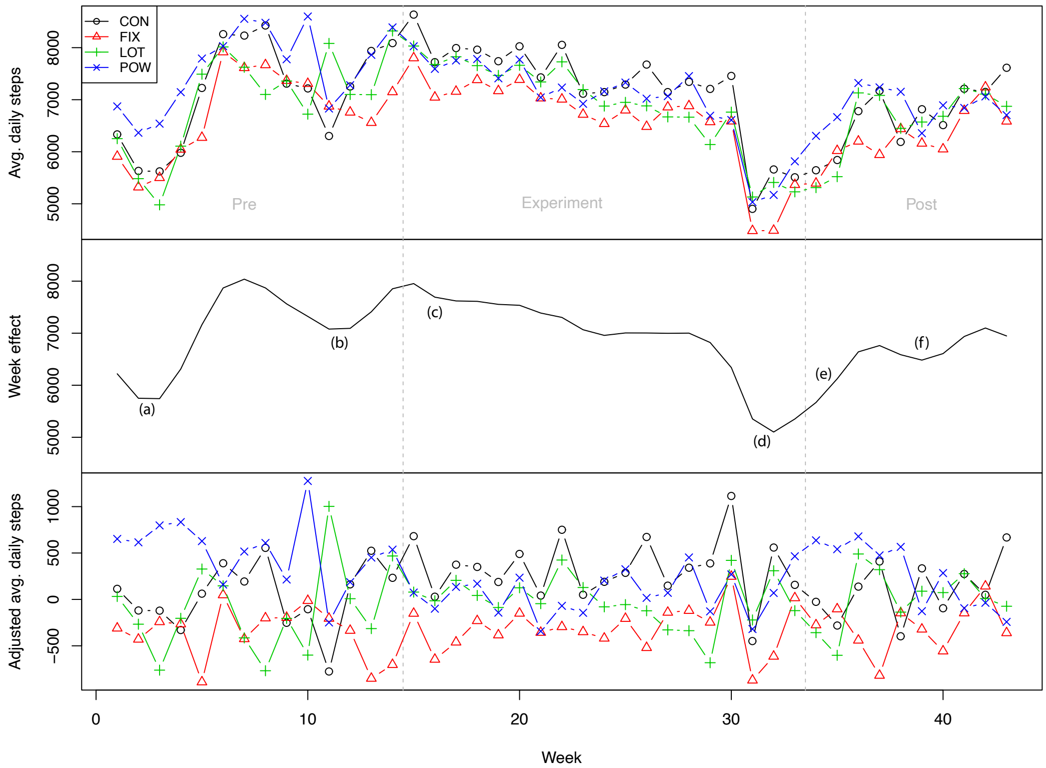

Figure 4.5
Average daily step-counts per week. Top: unadjusted numbers. Middle: weekly non-linear trend. The minimas correspond to specific periods of the academic calendar: (a) summer exam session; (b) summer retake session; (c) beginning of the fall courses; (d) end of courses and holidays season; (e) winter exam session; (f) beginning of the spring courses. Bottom: numbers after adjustment for the weekly trend.

4.5.1 H1. Participants with a high IMI_1 score will make more steps than participants with a low IMI_1 score

The results of the step counts model described in Section 4.4.8 are reported in Table 4.7. The model shows a linear increase in the step counts with IMI_1 in each phase for the different experimental conditions: for the control group, in the Pre-Experiment phase

[16]See http://www.wellocracy.com/2013/10/how-many-steps/, last retrieved March 2019.

step-counts increased of 1129.7 steps for each unit of IMI_1 (95% CI, 620.3 to 1639.0; $p < .01$), in the Experiment phase, step counts increased by 798.1 steps for each unit of IMI_1 (95% CI, 304.0 to 1292.2; $p < .01$), and in the Post-experimental phase step counts increased by 1118.8 steps for each unit of IMI_1 (95% CI, 620.7 to 1616.9; $p < .01$).

During the Pre-Phase, the linear increase described for the control condition was also found in the other conditions, which indicates that the IMI_1 score predicts the amount of PA a participant will do. In the Experimental phase, this linear increase remained mildly positive, whereas, for the control group, this remained significant, for the other conditions this was not statistically significant[17]. Similar results were found in the Post-Experiment phase. Although for FIX and LOT, the linear increase of the step counts with IMI_1 remained non-significant, we found a significant relationship in the POW condition where the step count increased by 571.8 steps for each unit of IMI_1 (95% CI, 126.4 to 1017.2; $p < .01$).

If we look at the goal model described in Section 4.4.8, we find the same results described above (reported in Table 4.8). Therefore, H1 is *partially validated*: Although the relation between steps and IMI_1 score was significant throughout the different phases of the experiment for the control group, the same relation was significant for the other conditions in the Pre-phase, then it became non-significant during the experiment and, after the experiment, returned to being significant but only for the POW condition. This finding reveals that the association between the IMI_1 score and the amount of steps was stable in the control group, whereas it was perturbed by the intervention. Therefore, we looked more specifically at the influence of the experimental manipulation on the intrinsic motivation with H2.

Table 4.7
Adjusted increase in daily step counts for one IMI_1 level by experimental condition and measurement phase.

Increase in daily step count for 1 point change in the IMI score

	Pre-phase (wk 1–14)			Experiment (wk 15–33)			Post-phase (wk 34–44)		
Condition	Estimate	95% CI	Adj. P	Estimate	95% CI	Adj. P	Estimate	95% C.I.	Adj. P
CON	1129.7	(620.3 ; 1639.0)	**<0.01**	798.1	(304.0 ; 1292.2)	**<0.01**	1118.8	(620.7 ; 1616.9)	**<0.01**
FIX	497.1	(39.1 ; 955.0)	**0.03**	437.9	(-18.6 ; 894.5)	0.06	445.6	(-22.6 ; 913.9)	0.06
LOT	466.0	(3.48 ; 935.4)	**0.05**	265.9	(-196.9 ; 728.7)	0.26	436.8	(-38.3 ; 911.9)	0.07
POW	502.7	(93.0 ; 912.4)	**0.02**	372.2	(-36.9 ; 781.2)	0.07	571.8	(126.4 ; 1017.2)	**0.01**

[17] In the remainder of this paper, we will use term *significant* to mean *statistically significant*.

Table 4.8
Adjusted slope of the IMI $_1$ in the goal model by experimental condition and measurement phase. The unit is the odds ratios.

Increase in odds ratio for 1 point change in the IMI score

	Pre-phase (wk 1–14)			Experiment (wk 15–33)			Post-phase (wk 34–44)		
Condition	Estimate	95% CI	Adj. P	Estimate	95% CI	Adj. P	Estimate	95% C.I.	Adj. P
CON	1.73	(1.37 ; 2.17)	**<0.01**	1.64	(1.27 ; 2.12)	**<0.01**	1.78	(1.36 ; 2.33)	**<0.01**
FIX	1.27	(1.04 ; 1.55)	**0.02**	1.22	(0.97 ; 1.54)	0.10	1.26	(0.99 ; 1.61)	0.06
LOT	1.22	(0.99 ; 1.49)	0.06	1.17	(0.92 ; 1.48)	0.20	1.32	(1.02 ; 1.70)	**0.03**
POW	1.24	(1.04 ; 1.48)	**0.02**	1.22	(0.98 ; 1.50)	0.07	1.36	(1.08 ; 1.73)	**0.01**

Table 4.9
Adjusted contrasts between IMI levels in each condition: measurements at week 26 (IMI_2) and 44 (IMI_3), against the reference measured at week 14 (IMI_1).

Change of IMI score during the Experiment and Post– phases

	IMI_2 - IMI_1			IMI_3 - IMI_1		
Condition	Estimate	95% CI	Adjusted P	Estimate	95% C.I.	Adjusted P
CON	0.30	(-0.06 ; 0.67)	0.12	0.44	(0.03 ; 0.85)	**0.03**
FIX	-0.33	(-0.67 ; 0.01)	0.06	-0.01	(-0.45 ; 0.44)	>0.99
LOT	-0.26	(-0.65 ; 0.13)	0.24	-0.37	(-0.90 ; 0.15)	0.20
POW	-0.29	(-0.67 ; 0.10)	0.18	0.12	(-0.43 ; 0.66)	0.84

4.5.2 H2a. During the experimental phase, the IMI score of participants in FIX and LOT will decrease

Adjusted differences in the IMI scores recorded at week 26 and at week 44, against the reference measured at week 14, are displayed in Table 4.9. IMI scores of the participants in the control group increased by 0.44 IMI points at the end of the Post-phase, when compared to the scores registered at the beginning of the experiment (95% CI, .03 to .85, $p = .03$). Observing the trends on the IMI scores, we note that, while these increased during the experiment and post-experiment phase for participants in the control condition, they decreased for participants in the FIX and LOT conditions. To mitigate the effects of extraneous factors (e.g., seasonal effects) or selection bias, we conducted a difference-in-difference analysis (or DID). DID is a common analysis technique used for example in Abadie [Abadie, 2005] or Blundell [Blundell et al., 2004].

We compared the average change over time in the IMI score in FIX and LOT groups, compared to the average change over time of the IMI score for the CON group. This analysis is valid under the assumption that individuals in any treatment group (FIX, LOT, or POW) would behave the same way as participants in the control group. In particular, we assume that in average the IMI scores in any treatment group would change over time by the same amounts of the IMI scores of the control

Table 4.10
Difference in differences analysis for H2. Fixed effect coefficients of the regression of the IMI in function of the phase and of the condition. Reference levels are CON group and IMI_1, respectively. The subject effect is controlled with a random intercept (not shown below).

	Estimate	Std. Error	df	t value	Pr(>|t|)	CI (2.5%)	CI (97.5%)
Intercept	4.74	0.15	398.05	31.69	0.00	4.45	5.03
condition							
FIX-CON	0.18	0.21	398.05	0.88	0.38	-0.22	0.59
LOT-CON	0.39	0.21	398.05	1.84	0.07	-0.02	0.81
POW-CON	0.11	0.22	398.05	0.52	0.60	-0.31	0.54
phase							
IMI_2-IMI_1	0.30	0.16	319.58	1.85	0.07	-0.02	0.62
IMI_3-IMI_1	0.44	0.18	326.78	2.40	**0.02**	0.08	0.80
interactions							
(FIX-CON):(IMI_2-IMI_1)	-0.63	0.22	316.24	-2.83	**0.01**	-1.07	-0.20
(LOT-CON):(IMI_2-IMI_1)	-0.56	0.24	323.23	-2.36	**0.02**	-1.03	-0.10
(POW-CON):(IMI_2-IMI_1)	-0.59	0.24	320.02	-2.47	**0.01**	-1.05	-0.13
(FIX-CON):(IMI_3-IMI_1)	-0.45	0.27	328.41	-1.65	0.10	-0.98	0.08
(LOT-CON):(IMI_3-IMI_1)	-0.82	0.30	334.74	-2.74	**0.01**	-1.39	-0.24
(POW-CON):(IMI_3-IMI_1)	-0.33	0.31	334.22	-1.07	0.29	-0.92	0.27

group. Though we do not have data to test it, this is a reasonable assumption to make as we do not have reasons to believe that extraneous factors could have unevenly affected the individuals in different experimental conditions. Therefore, the estimates of the DID study are obtained from the interaction terms in the underlying regression used previously in the within group study. These coefficients and their p-values are reported in Table 4.10[18]. The difference of IMI scores between the FIX and CON groups decreased by 0.63 points between the beginning and the end of the experiment[19] (95% CI, -1.07 to -0.20, $p < .01$). Similarly, that difference between the LOT and CON groups decreased by 0.56 points between the beginning and the end of the experiment (95% CI, -1.03 to -0.10, $p = .02$). When looking at the differences between the Pre- and Post-experiment phases, the decrease of the IMI scores between the FIX and CON groups is no longer significant (95% CI, -0.98 to -0.08, $p = .10$, ns), while that decrease of the difference between the LOT and CON group, estimated to -0.82, is still significant (95% CI, -1.39 to -0.24, $p < .01$). In summary, these results *enable us to validate* H2a. The negative trends on the IMI scores in FIX and LOT are indicative of a

[18]Scripts for reproducing our analyses, as well as full `R` outputs are available on https://osf.io/m28xe/.

[19]The unfamiliar reader might wonder whether a difference of half-point in the IMI score is clinically significant. In a similar study, Tsigilis [Tsigilis, 2005] found that the IMI score significantly predicted the performance obtained on a 20m shuttle field test. They found that a a difference of .50 in the IMI score resulted in the variation of the VO_2 max, or maximal oxygen consumption, between 30.5 and 33.5 (mL/Kg $\cdot$ minute) for a population of students 19 years old, which results in a difference between a poor an fair fitness level [Guyton and Hall, 2015].

detrimental effect of the tangible rewards over the intrinsic motivation of the participants. Next, we looked at the effect of the motivational messages on the IMI scores.

4.5.3 H2b. During the experimental phase, the IMI score of participants in POW will increase

For this analysis we refer to the same DID study reported in Section 4.5.2 and in Table 4.10 above. Contrary to our expectations, the difference in the IMI scores between the POW and CON groups decreased by 0.59 points between the Pre- and Experiment phases (95% CI, -1.05 to -0.13, $p < .01$). That decrease between the Pre- and Post-experiment phase is no longer significant (95% CI, -0.92 to -0.27, $p = .29$, ns). These *results do not enable us to validate* H2b. Not only we do not see the expected increase of the IMI scores, but we observe a negative trend similar to those obtained in the FIX and LOT intervention. These negative trends are indicative of a detrimental effect of the motivational messages, which we did not anticipate. The motivational messages sent in POW did not increase the intrinsic motivation of the participants. Instead, these messages had a negative effect on the IMI scores. Therefore, we checked the step counts to see whether these trends produced visible effects on the overall PA.

4.5.4 H3a. During the experimental and post-experimental phase, participants in FIX and LOT walked less than participants in the control condition

Adjusted differences in the mean daily step-counts are displayed in Table 4.11. The analysis of the goal study is reported in Table 4.12 and confirms the results of the steps study[20]. The results reported in these two tables were obtained using models where IMI_1 is a numerical factor. This accounts for the effect of IMI_1 on the number of steps for Table 4.11 and on the probability of reaching their goal for Table 4.12. As this effect varies from one group to another, an interaction between IMI_1 and condition was included. Hence, to compare the conditions, because of the presence of interactions, we fixed the value of IMI_1 (because the difference between two conditions varies with

[20] As an odds ratio below 1 indicates a decrease in the proportion, we see that these trends follow the results of the step counts.

Table 4.11
Adjusted contrasts of the step counts between conditions for several IMI_1 levels.

Change in daily step count for fixed IMI_1 levels

	Pre-phase (wk 1–14)			Experiment (wk 15–33)			Post-phase (wk 34–44)		
	Estimate	95% CI	Adj. P	Estimate	95% CI	Adj. P	Estimate	95% CI	Adj. P
IMI_1 = 2.0									
FIX - CON	1167.3	(-1284.3 ; 3618.9)	0.53	170.8	(-2243.7 ; 2585.3)	0.99	1299.3	(-1162.4 ; 3760.9)	0.45
LOT - CON	1393.5	(-1191.6 ; 3978.6)	0.43	909.2	(-1621.1 ; 3439.4)	0.71	1446.9	(-1117.2 ; 4010.9)	0.40
POW - CON	2131.9	(-211.8 ; 4475.6)	0.08	767.6	(-1532.1 ; 3067.4)	0.75	1786.8	(-602.9 ; 4176.5)	0.19
IMI_1 = 3.5									
FIX - CON	218.5	(-1176.1 ; 1613.0)	0.94	-369.5	(-1746.8 ; 1007.8)	0.83	289.6	(-1122.4 ; 1701.5)	0.90
LOT - CON	398.0	(-1120.4 ; 1916.4)	0.84	110.8	(-1378.7 ; 1600.4)	0.99	423.9	(-1085.1 ; 1932.8)	0.81
POW - CON	1191.5	(-180.4 ; 2563.4)	0.11	128.7	(-1221.8 ; 1479.1)	0.98	966.4	(-441.4 ; 2374.2)	0.25
IMI_1 = 4.5									
FIX - CON	-414.1	(-1383.9 ; 555.6)	0.60	-729.7	(-1686.2 ; 226.8)	0.18	-383.5	(-1374.9 ; 607.8)	0.66
LOT - CON	-265.7	(-1329.6 ; 798.3)	0.85	-421.4	(-1465.6 ; 622.8)	0.64	-258.1	(-1323.7 ; 807.5)	0.86
POW - CON	564.5	(-449.6 ; 1578.7)	0.41	-297.3	(-1299.5 ; 704.9)	0.80	419.4	(-638.8 ; 1477.6)	0.65
IMI_1 = 6.0									
FIX - CON	-1363.0	(-2774.7 ; 49.8)	0.06	-1270.0	(-2648.8 ; 108.7)	0.08	-1393.2	(-2813.3 ; 26.8)	0.06
LOT - CON	-1261.2	(-2671.5 ; 149.2)	0.09	-1219.7	(-2593.8 ; 154.3)	0.10	-1281.1	(-2698.3 ; 136.1)	0.09
POW - CON	-375.9	(-1817.6 ; 1065.9)	0.84	-936.2	(-2354.4 ; 482.0)	0.28	-401.0	(-1905.1 ; 1103.0)	0.83

the value of IMI_1). We then selected $IMI_1 = 2$, 3.5, 4.5, and 6 in an ad-hoc way, to globally cover the range of possible IMI_1 values.

The experimental conditions were not significantly different at the various IMI_1 levels. We now turn to the description of trends estimated by the model. These results should be taken with caution, as the lack of significance does not allow us to imply causal relations between the variables and should be cross-validated in future research. During the experiment, for FIX we observe a negative trend in the step counts for $IMI_1 \geq 3.5$, and for LOT we observe a similar negative trend for $IMI_1 \geq 4.5$. However, these differences were visible also in the Pre-experiment phase, hence we were not able to derive any conclusion on the effects of the experimental manipulation. Therefore, these *results do not enable us to validate* H3a: The tangible rewards did not change the amount of steps of the participants during and after the experiment.

Finally, we look at whether motivational messages have an effect on the main outcome.

Table 4.12
Adjusted contrasts between conditions in the goal model, for several IMI_1 levels. The unit is the odds ratios.

Change in odds ratio for fixed IMI_1 levels

	Pre-phase (wk 1–14)			Experiment (wk 15–33)			Post-phase (wk 34–44)		
	Estimate	95% CI	Adj. P	Estimate	95% CI	Adj. P	Estimate	95% CI	Adj. P
IMI_1 = 2.0									
FIX - CON	1.95	(0.66 ; 5.80)	0.33	1.80	(0.52 ; 6.30)	0.54	2.40	(0.64 ; 9.08)	0.28
LOT - CON	2.40	(0.76 ; 7.57)	0.18	1.96	(0.53 ; 7.32)	0.48	2.01	(0.50 ; 8.10)	0.49
POW - CON	2.87	(1.01 ; 8.13)	**0.05**	1.73	(0.52 ; 5.73)	0.56	2.62	(0.72 ; 9.55)	0.20
IMI_1 = 3.5									
FIX - CON	1.23	(0.66 ; 2.29)	0.74	1.15	(0.56 ; 2.36)	0.91	1.43	(0.66 ; 3.08)	0.54
LOT - CON	1.42	(0.72 ; 2.78)	0.47	1.18	(0.54 ; 2.56)	0.90	1.28	(0.56 ; 2.91)	0.79
POW - CON	1.74	(0.95 ; 3.19)	0.08	1.10	(0.54 ; 2.22)	0.96	1.76	(0.82 ; 3.76)	0.20
IMI_1 = 4.5									
FIX - CON	0.91	(0.59 ; 1.38)	0.87	0.86	(0.52 ; 1.40)	0.78	1.01	(0.60 ; 1.72)	>0.99
LOT - CON	1.00	(0.63 ; 1.59)	>0.99	0.84	(0.49 ; 1.44)	0.75	0.95	(0.54 ; 1.68)	0.98
POW - CON	1.25	(0.80 ; 1.93)	0.49	0.81	(0.48 ; 1.37)	0.65	1.35	(0.77 ; 2.35)	0.45
IMI_1 = 6.0									
FIX - CON	0.57	(0.31 ; 1.04)	0.08	0.55	(0.27 ; 1.11)	0.12	0.60	(0.29 ; 1.25)	0.25
LOT - CON	0.59	(0.32 ; 1.08)	0.10	0.50	(0.25 ; 1.02)	0.06	0.60	(0.29 ; 1.25)	0.25
POW - CON	0.76	(0.41 ; 1.40)	0.56	0.52	(0.25 ; 1.07)	0.09	0.90	(0.42 ; 1.96)	0.96

Table 4.13
Adjusted contrasts between odds ratio of exceeding 11K steps among the day-participants exceeding 10K steps

Odds ratio of exceeding 11K steps once the daily goal has been achieved

	Pre-phase (wk 1–14)			Experiment (wk 15–33)			Post-phase (wk 34–44)		
Contrast	Estimate	95% CI	Adj. P	Estimate	95% CI	Adj. P	Estimate	95% CI	Adj. P
CON - FIX	1.16	(0.88 ; 1.52)	0.42	0.89	(0.72 ; 1.08)	0.35	1.01	(0.71 ; 1.42	1
LOT - FIX	0.95	(0.71 ; 1.26)	0.92	0.79	(0.64 ; 0.99)	**0.03**	0.97	(0.67 ; 1.39)	0.98
POW - FIX	0.85	(0.64 ; 1.13)	0.41	0.71	(0.57 ; 0.89)	**<0.01**	0.96	(0.66 ; 1.38)	0.97

4.5.5 H3b. During the experiment and post-experiment phases, the number of steps of the participants in POW will be higher than those of participants in the other conditions

We find that all the differences between the POW group and the control group, in terms of daily step counts, are not significant (see Table 4.11). Similar results can be observed in the goal study, reported in Table 4.12. Therefore, we turned to the description of trends, as per H3a above. During the experiment, we observed a negative trend for participants with $IMI_1 \geq 4.5$. However, these trends are not indicative of any effect on the PA of the participants during and after the experiment, as the same trends were present in the Pre-phase.

Therefore, these *results do not enable us to validate* H3b, as they contradict the effect we expected for motivational messages.

Instead of encouraging participants to walk more, motivational messages had no effect on the number of steps of the participants.

We then wondered whether the discrepancies we observed between our experimental results and the effects predicted by the theory could have been artifacts of our experimental design.

Therefore, we look at the two aspects of our experiment that could have influenced the results: (1) the choice of 10K steps with regard to the levels of PA before the beginning the experiment; (2) the incentive mechanism of the experimental conditions with regard to the goal of reaching the 10K steps. Finally, given the results obtained in the IMI_1 study, reported in Section 4.5.2 and 4.5.3 and Table 4.9, we conduct an additional analysis to evaluate the effect of the experiment on the *most* and *least* motivated people in the groups of participants.

4.5.6 Additional Subgroup Analysis for the Steps Study

In addition to testing the hypotheses derived from the theory, we explored whether the interventions worked for participants for whom the 10K steps was a reasonable challenge better than for those for whom this was a target too difficult to achieve. Therefore, we performed a post-hoc split of participants, based on their level of PA as recorded in the 3 weeks that preceded the beginning of the experiment (i.e., week 12 to 14). We chose this particular time window to limit the seasonal effects of summer during the baseline. We then organized the participants in three groups, based on the following definitions:

- **Challenging**: those for whom the target of 10K steps was above the mean plus 1.5 times the standard deviation (or SD) of the steps calculated on the three weeks before the beginning of the experiment.
- **Feasible**: those for whom the target fell between the mean and the mean plus the SD.
- **Achieved**: those for whom the target was below the mean.

We then computed the frequencies in each PA activity group by experimental condition. We did not observe any significant difference between the experimental groups and the control group during the experiment and Post- phase. We now look at the analysis of incentive mechanism, with regard to the the goal of reaching 10K steps.

4.5.7 Additional Subgroup Analysis for the Goal Study

In addition to the aforementioned analyses suggested by the literature, we looked at whether setting a specific goal at 10K steps was perceived as an implicit contract by participants and provoked a *ceiling effect* on the step counts. This could have been particularly true for participants in the FIX condition, for whom attaining this specific goal would lead to the most immediate form of reward. Extrinsic incentives might cause participants to stop striving once their goal is reached, or might lead them to stop trying if they cannot even come close to attaining it. Therefore, we computed in each phase, in each group, and for each level of IMI_1, the number of day-participants exceeding 11k steps among the day-participants exceeding 10k steps. We then ran a logistic regression on this indicator with factor Condition and Phase, including the interaction, and controlling for the IMI_1. The odd ratio comparisons were done using a Dunnet's correction of the group FIX against any other group. The results of this analysis are reported in Table 4.13. We found that indeed, during the experiment only, the FIX group had an odd ratio smaller than the POW group (0.71, 95% CI 0.57 to 0.89, $p < .01$) and the LOT group (0.79, 95% CI 0.64 to 0.89, $p < .05$). The other contrast FIX - CON is not significantly different. In the Pre- and Post-phases no significant differences emerge. This result confirms that setting up a static goal, combined with offering a monetary incentive, can limit the chances that participants of this intervention will exceed this limit. We now look at the effects of the interventions on the most and least motivated people among the participants.

4.5.8 Additional Subgroup Analysis for the IMI Study

The analysis reported in Section 4.5.2 suggests that the experimental manipulations had a negative effect on the intrinsic motivation of the participants. To study further the trends reported in Table 4.9, we looked more closely at the least and most motivated participants in the sample.

We conducted a sub-group analysis of the IMI model among the 23 participants with IMI_1 smaller or equal to 3. For this analysis, we chose this particular threshold because selecting and IMI smaller or equal to 2 would have left us with too few participants. Hence, we conducted the analysis with 5 participants in CON, 6 in FIX, 6 in LOT, and 6 in POW. The results are shown in Table 4.14. The analysis revealed a significant effect on the FIX condition in the experimental phase: The IMI score increased by 1.15 units (95% CI, .02 to 2.33, $p = .05$). The estimates of the model for the other conditions and in the Post-phase are not significant.

Concerning the most motivated participants, an analysis of the sub-group of the IMI model was carried out among the 63 participants with IMI_1 greater or equal to 6 (specifically for 13 participants in CON, 17 in FIX, 19 in LOT, and 14 in POW). The results are shown in Table 4.14. The analysis reveals a significant effect on the FIX condition in the experimental phase: The IMI score decreased by 1.48 units (95% CI, -2.08 to -0.89, $p < .01$). The estimate for the FIX condition during the Post-phase was not significant (95% CI, .06 to -2.11, $p = .07$, ns). Furthermore, the analysis reveals significant effects for the LOT condition: The IMI score decreased by .99 units in the Experiment-phase (95% CI, -1.55 to -0.43, $p < .01$) and -1.25 in the Post-phase (95% CI, -2.00 to -0.50, $p < .01$). Similarly, we find significant effects on the POW condition: The IMI score decreased by -1.42 units in the Experiment phase (95% CI, -2.06 to -0.78, $p < .01$) and -1.04 in the Post-phase (95% CI, -1.95 to -0.13, $p < .05$). Finally, the estimates for the CON condition were not significant.

These results show that the experimental manipulations have an effect on the intrinsic motivation of the participants. The least motivated participants in the FIX condition increased their IMI score during the experiment. The most motivated participants are negatively affected by the experimental manipulations, as their IMI score decrease during the experiment and continue decreasing for those in the LOT and POW condition in the Post-phase.

The results should be taken with caution, as this sub-group analysis was decided a posteriori. In particular, the IMI level threshold of 6 and 3 were selected based on the results of the step counts and goal analysis. There was no a priori justification for these thresholds, thus these results should be confirmed in future research.

For a final step, to find possible qualitative accounts for the quantitative trends observed in the data, we looked at the results of the questionnaires delivered at week 26 and week 44.

Table 4.14
Subgroups analysis. Adjusted contrasts between IMI levels measured at different phase in each condition.

Change of IMI score for fixed levels of IMI_1

	IMI_2 - IMI_1			IMI_3 - IMI_1		
Condition	Estimate	95% CI	Adj. P	Estimate	95% CI	Adj. P
$IMI_1 \leq 3$						
CON	0.88	(-0.33 ; 2.10)	0.18	1.11	(-0.10; 2.33)	0.07
FIX	1.15	(-0.02 ; 2.33)	**0.05**	1.21	(-0.07 ; 2.48)	0.06
LOT	0.61	(-0.67 ; 1.89)	0.44	0.58	(-0.84 ; 2.01)	0.54
POW	0.36	(-0.75 ; 1.47)	0.66	-0.40	(-2.68 ; 1.87)	0.87
$IMI_1 \geq 6$						
CON	-0.51	(-1.30 ; 0.28)	0.26	-0.30	(-0.33 ; 2.10)	0.63
FIX	-1.48	(-2.08 ; -0.89)	**<0.01**	-0.84	(-0.33 ; 2.10)	0.07
LOT	-0.99	(-1.55 ; -0.43)	**<0.01**	-1.25	(-0.33 ; 2.10)	**<0.01**
POW	-1.42	(-2.06 ; -0.78)	**<0.01**	-1.04	(-0.33 ; 2.10)	**0.02**

4.5.9 Participants' perception of the experimental manipulations

As we explained in the methodology section, we deployed two short questionnaires at week 26 and at week 44 of the study. The goal was to collect qualitative feedback on the application, and about the design of the persuasive features of the research app. Specifically, we asked the following questions:

1. *Que pensez-vous de la récompense qui a été donnée lorsque l'objectif quotidien a été atteint?* [What do you think about the reward that was given when your daily goal was reached?]
2. *Que pensez-vous des messages de motivation envoyés chaque matin?* [What do you think about the motivational messages that were sent every morning?]

The first question was asked exclusively to participants in the conditions FIX and LOT, and the second was asked only to respondents in the condition POW. Two researchers analyzed the answers provided to the questions above, and they looked specifically for recurring comments that could account for the perceptions that participants had about the interventions. We summarize the main findings in the subsections below.

Perceptions of the Tangible Rewards

The participants with a low IMI felt generally positive about the tangible rewards that were offered during the experiment phase: e.g., *"Oui je marchais et courrais lègérement plus"* [Yes I walked and ran a little more] (Id: 292, IMI: 2.57, Condition: FIX); *"Ça m'a un peu motivé mais pas au point de toujours aller faire du sport car je n'ai pas le temps."* [It did motivate me a bit, but not to the point of doing sport every day because I do not have the time] (Id: 297, IMI: 2, Condition: FIX). However, the participants with a high IMI did not perceive the reward to be sufficient to motivate them to increase their physical activity level: e.g.,*"une plus grande récompense m'aurait encouragé davantage"*, (Id: 244, IMI: 4.71, Condition: FIX) [a bigger reward would have encouraged me more.]; *"La récompense fait plaisir mais me pousse pas à faire davantage d'effort"* [The reward is fun but not pushing me to make more of an effort.] (Id: 241, IMI: 3.75, Condition: FIX); *"Au début j'étais très motivée, mais les motivations financières n'étant pas assez élevées, j'ai arrêté de faire des efforts."* [At first I was very motivated, but the financial rewards were not high enough, I stopped making an effort."] (Id: 287, IMI: 4.86, Condition: FIX). Finally, participants with an IMI equal or greater than 6 expressed rather negative sentiments about the rewards: e.g., *"Je n'ai pas fait cette étude pour gagner de l'argent. Je fais assez activité physique pour moi-même"* [I did not participate in this study to make money. I do enough physical activity for my own benefit] (Id: 176, IMI: 6.77, Condition: LOT); *"Je ne modifiais pas particulièrement mon activité physique en fonction de cette "récompense", cela dépend plutot du temps libre que j'ai."* [I did not particularly modify my physical activity according to this "reward", it depends rather on the free time that I have.] (Id: 252, IMI: 7, Condition: FIX). In summary, we found that participants with a low IMI felt generally positive about the rewards, while those with an higher IMI had rather negative comments.

Perceptions of the Motivational Messages

Similarly to what we saw in the other conditions, the participants assigned to the condition POW and with low IMI reported being influenced by the messages and perceived them as encouraging: e.g., *"Oui je suis influencée par les commentaires que l'on reçoit."* [Yes I am influenced by the comments we receive.] (Id: 148, IMI: 3.44, Condition: POW); *"Parfois ça envoie des messages encouragants"* [Sometimes it sends encouraging messages.] (Id: 120, IMI: 2.86, Condition: POW).

On the contrary, participants with high IMI felt rather negative about the usefulness of the messages: e.g., *"Commentaires pas spécialement motivant, j'ai arrêté de les lire après quelques jours."* [Not particularly motivating comments, I stopped reading them after a few days.] (Id: 131, IMI: 5.86, Condition: POW); *"Je n'aime pas beaucoup les commentaires qu'envoit l'application du genre "vos amis vous préfèrent lorsque vous faite du sport" etc.* [I do not like very much the comments that come with the application like "your friends prefer you when you play sports" etc.] (Id: 309, IMI: 6.58, Condition: POW).

Interestingly, some participants reported that the messages were not personalized enough hence were not useful for them: e.g., *"Les messages de motivation ne me touche pas... ce n'est pas assez personnalisé je trouve."* [the motivational messages do not affect me ... they are not personalized enough, I think.] (Id: 154, IMI: 4, Condition: POW).

Also, some participants indicated that they received too many notifications and consider them annoying: e.g., *"Les notifications journalières son très pénibles"*["Daily notifications are very painful"] (Id: 196, IMI: 5.86, Condition: POW); *"Certaines notifications etaient un peu trop agressive et/ou intrusive"*["Some notifications were a little too aggressive and/or intrusive"] (Id: 151, IMI: 4.57, Condition: POW)

Other participants expressed being more attentive to the messages at the beginning of the intervention: *"Elle m'a fait changer au tout début, quand je voyais les notifications. C'est vrai que je faisais plus attention. Par contre, les notification sont devenus habituelles et a un moment donné je faisais plus attention."* [It made me change at first, when I saw the notifications, it's true that I was paying more attention. On the other hand, the notifications became habitual and at a certain moment I did not pay attention.] (Id: 122, IMI: 5.14, Condition: POW); *"Oui, au début, j'ai vraiment fait l'effort de prendre moins les transports en commun et de marcher plus dans ma vie de tous les jours. Puis les messages sont devenus agaçants."* [Yes, at first, I really made the effort to take less public transportation and walk more in my everyday life, then the messages became annoying.] (Id: 155, IMI: 6.57, Condition: POW).

4.6 Discussion

One of the main outcomes of this work is that we learned that tangible rewards do not help establish lasting healthy routines. During the course of the experiment, we observed that there was not any significant difference in the amounts of steps walked, whether we offered participants money or not to perform this activity. The estimates of the step model are not significant for each contrast and for each phase of the study. Hence, we cannot validate H3a and cannot confirm the predictions of the SDT theory [Ryan and Deci, 2017] or those of empirical studies [Orlick and Mosher, 1978, Patel et al., 2016] for which a reduction of PA in FIX and LOT was expected. However, these results do not contradict the theory either, as it is evident that the monetary incentives we tested in this experiment did not increase the level of PA of the participants.

A second – and more important – outcome of this research is that although PA did not substantially change, the intrinsic motivation of the participants indeed changed as a consequence of the experimental manipulations. At the end of the experiment phase, the DID analysis indicate that the IMI scores of participants in the control group increased, whereas those of the participants in FIX and LOT lowered. In the Post-phase, these negative trends recovered a bit but remained negative (see Table 4.10 and the analysis reported in Sec. 4.5.2). The subgroup analysis reported in Section 4.5.8 further reveals that though for the least motivated participants the rewards increased the IMI scores at the end of the experiment, for the most motivated participants it was just the opposite. In other words, an economic incentive can motivate people who are not yet interested in walking or running, but it can have the opposite effect on people who are already have the habit of walking (or running).

The extent of these negative effects largely depends on the *intrinsic motivation* of the person receiving the rewards: The more the person is already self-motivated about performing the activity, the higher the chance is that rewards will have a detrimental effect.

Offering money to a person who is already motivated to perform a certain activity can confuse the person about their real motives. This finding is in line to previous research conducted by Orji et al. [Orji et al., 2014] who found that monetary rewards do not work as a behavior-change strategy as "*people tend to view the rewards and the values they get from them as the only benefit of adopting a healthy behavior" [Orji et al., 2014, p. 29]*. Hence, failing to establish lasting benefits. This is also signaled by the need of some participants to re-state their original interest in walking or running (see qualitative examples reported in Section 4.5.9). This finding was also predicted by SDT; the theory connected the detrimental effects of the rewards to the basic need of *autonomy* [Ryan and Deci, 2000b]. However, little research to date has provided an empirical account of the theory that demonstrates the effect of the intervention on the intrinsic motivation. Although most studies infer the effect on intrinsic motivation by using a *free-choice* paradigm (i.e., the frequency with which the participants engage with the target activity after the intervention has been removed), we also demonstrated the negative effects through a direct measurement of the IMI score.

It is important to also highlight that the findings we present apparently contradict those that report positive effects of rewards in increasing PA, (e.g., as discussed in the literature review section [Finkelstein et al., 2008, Charness and Gneezy, 2009, Daryanto et al., 2010], among others). We argue that this contradiction is simply apparent because SDT also posits that when tangible rewards are salient to the activity for which they are given, they can indeed propel action. We speculate that if we had chosen a larger economic incentive, we could have produced a visible increase of PA, as posited by SDT. However, the theory also cautions (and these findings reveal) that though these interventions might have positive effects in the short term, they might – in the long term – harm the intrinsic motivation of the participants. In our findings, this is clearly evident, as the participants' IMI scores reduced during the experiment and post-phase. Therefore, our findings provide empirical evidence of this "downside" of tangible rewards, by showing that paying people to do more PA can harm the inner resources (for self-determined action) people might already possess. Furthermore, even if salient rewards work with people who are not yet motivated, these interventions will not encourage people to progress in the motivation continuum posited by the SDT, hence people will not ultimately become autonomous in performing the target behavior. Therefore, this type of intervention will always be dependent on the delivery of the rewards, as we expect that as soon as the incentive is suspended, the person will fall back into their previous behavior.

Another important finding is that we do not find any difference between the tangible rewards provided by the fix incentive and the lottery condition. During the experiment, people who were submitted to the LOT condition did not behave differently than those in the FIX condition: We observe the same negative trends on physical activity overall. We also observe the same detrimental effect on the IMI scores during the experiment. We designed the LOT condition to offer an incentive more salient than the FIX condition. Therefore, we expected the reward with higher salience to yield a larger effect, as predicted by the Salience Theory [Bordalo et al., 2012]. However, obtaining the reward of the LOT intervention required two steps: i.e., gaining a ticket and winning the lottery. This further delayed the gratification of the participants (see the work of O'Donoghue et al. [O'Donoghue and Rabin, 2000]), which can explain the presented results.

Comparing these results with those obtained by Patel et al. [Patel et al., 2016], we noticed several differences in the experimental design: (1) The maximum prize for winning the lottery was 50 USD, an incentive five-times larger than the incentive we offered; (2) the daily goal was 7K steps, instead of the 10K in our study; and (3) the participants of their study included obese people, which could result in lower IMI scores at the onset of the experimental phase of the study. In particular, this last point can account for the differences in efficacy of the lottery treatment in our study and in the study conducted by Patel et al. [Patel et al., 2016]: If we consider the subgroup analysis for the IMI study reported in Section 4.5.8, we notice that the trends on the IMI scores are all positive for the the least motivated people in the sample. A salient reward can be effective in motivating participants who are not intrinsically motivated to perform the activity. However, this positive trend might eventually stall, because once participants raise their intrinsic score, this mechanism will not be sufficient to make them progress towards autonomous behaviour (see Figure 4.1). This comparison between the two studies and its conclusion is speculative and should be confirmed by future research.

The third intervention that we tested in this experiment was related to sending daily motivational messages (condition POW). We find that this intervention does not increase the physical activity of the participants. Similar to what we describe above for the FIX and LOT conditions, we do not observe any differences in the control group, in terms of the number of walked steps. However, we observe a detrimental effect on the IMI scores measured at the end of the experiment phase. These findings contradict our expectations that were based on SDT and prior empirical research on motivational messages [Kinnafick et al., 2016]. This difference, however, has to be interpreted by looking at how we translated the theory into design. In fact, SDT specifies that in order for feedback to provoke positive effects on the recipients' sense of competence, it has to provide information that can help recipients reflect on their own performance [Ryan and Deci, 2017, p. 154]. If we consider the corpus of messages more carefully, we observe that the messages did not provide information related to the specific execution of the task by the participants (i.e., how much a participant was walking or running in regard to the goal of the 10K steps). Instead, these texts communicated generic information about why running or walking is beneficial for health and provided tips on how to setup healthy exercise routines. This information did not really provide useful information for the participants to assess progress from one week to the next, or to understand whether – aside from meeting the daily goal – the overall PA was increasing or decreasing. Therefore, the motivational messages that were delivered in this study fell very short of providing information that participants could use to reflect on their performances, hence the lack of effect on the basic need of competence. The corpus used in this experiment was modeled on the trans-theoretical model [Prochaska and Di Clemente, 1983]. The messages followed a specific chronological order that tailored the intervention to the stages of change: i.e., progressing stages of changing behavior. However, the same message was sent to all participants in the POW condition, regardless of the specific status of each participant. In a later work, the same researchers, who created the original corpus, conducted additional research and found that these motivational messages could be more effective if their delivery is tailored on the personality and gender of the recipient [de Vries et al., 2017], or even the designer of the message, i.e., an expert or a peer [De Vries et al., 2017]. This is one aspect that should be considered in future research involving motivational messages. A second aspect to consider is that repeatedly receiving the messages every day was experienced by participants as an external pressure about the daily goal hence was perceived as controlling. This aspect of the study could explain the negative effect that we observed on the IMI scores and is documented by the participants' comments reported in Section 4.5.9 where they express being annoyed by the repeated notifications. The frequency with which we sent the motivational messages to the participants could also explain a difference in the results of [Kinnafick et al., 2016]. These authors could not observe any significant difference of PA (however, they observed a positive trend). Yet, they measured an increase of the IMI scores at the end of the experimental phase. In their experimental design, they delivered the messages only twice a week. We can speculate that this low frequency of delivery could have been perceived by participants as less controlling, compared to the frequency of our messages (i.e., a message every day). Furthermore, our results reveal that, similarly to what we observed for the tangible rewards, the negative effect of sending these motivational messages affects particularly the most motivated participants in the sample. In summary, these results reveal that the most motivated participants are extremely susceptible to the interventions that we tested in our experiment.

4.6.1 Implications for Design

The most direct implication of this research is that "one solution fits all" does not work for persuasive design. Depending on the characteristics of the user, interventions can have detrimental effects. Therefore, we argue that interventions should be personalized in order to maximize their efficacy and to avoid negative outcomes.

Previous research demonstrated that profiling the personality of the participants [Orji et al., 2017b] or the persuasive strategy that works best for the participants [Orji et al., 2017a] can improve the overall effectiveness of the intervention.

The results presented in this work support the argument of designing personalized interventions in two ways: (1) in terms of establishing the level of the intrinsic motivation of the participant at the onset of the intervention, and (2) in terms of designing optimal challenges for the participant. Concerning point (1), in this study, we found that the level of intrinsic motivation towards a given activity is a predictor of the effect of the intervention, and we argue that this metric could be used to choose the most effective persuasive techniques. In particular, for users who have an IMI score lower than 4.5, tangible rewards could be considered a persuasive technique for encouraging a person's internalization process. However, after 3–5 weeks, this technique should be replaced by a different technique able to continue the introjection of the PLOC. Extending the intervention further might hinder subsequent improvements. Furthermore, for users with an IMI score higher than 4.5, using tangible rewards is altogether not recommended, as this technique might simply lead to outcomes that are contrary to the desired results. Unfortunately, at this stage, we are unable to advise which persuasive techniques should be used instead (or in combination with), as this goes beyond the scope of this paper.

One of the specific advantages of using the IMI score to personalize the user experience would be related to the design of practical instruments that capture the metric. Building personality profiles of participants usually involves long surveys[21],whereas capturing the IMI can typically be done with 6 questions, thus simplifying the data-capturing aspect.

Concerning point (2), we found that imposing the same goal (i.e., 10K steps/day) to all participants led to a ceiling effect on the outcomes that some participants could

[21] One of the standard to capture the personality profile of a person is the BigFive questionnaire that involves 50 questions in its shortest form [Goldberg, 1992].

have achieved. Therefore, this study suggests that participants should be offered personalized optimal challenges that could increase (or decrease), depending on their short-term achievements. For instance, if in a given week a person does 8K steps, a reasonable goal for the following week could be an increase between 5% and 10% (i.e., 8.4K to 8.8K steps). These adjustable targets could also take care of the seasonal effects we observed in the study (e.g., lower PA due to exam seasons, holidays, etc.). The concept of *optimal challenge* was already discussed within the SDT [Deci, 1975] and demonstrated empirically in game immersion [Qin et al., 2010, Rigby and Ryan, 2011]. Providing adaptable targets is also one of the key properties of effective interventions, as posited by the goal-setting theory [Locke and Latham, 2002].

A second relevant implication of this research deals with the use of tangible rewards to provoke behavior changes that last. Although many scholars still debate the usefulness of these rewards and how to best design the incentive schemes [Patel et al., 2016, Gneezy and Rustichini, 2000], our results highlight the importance of looking at the impact of these interventions in the long run. Providing well-designed and salient rewards can often boost compliance to the target behavior. However, this can create negative effects in the long-term perspective. Therefore as designers, we should reflect on what kind of changes we want to provoke. We argue that having a short-sighted approach to the problem might result in creating technology that acts as a *prosthesis* for humans but that will not ultimately empower users in *exploiting their full potential*. In other words, design solutions must not create dependency on the technology itself. Tangible rewards should be reconsidered in persuasive design, not only because in many cases they produce negative results even in the short term (i.e., for highly motivated people, as discussed above) but most importantly because even in the cases where they seem to work in the short-term they will discourage the users in their future endeavours. From this perspective, persuasive app designers should seek solutions that *empower* users rather than those that create dependence. Unfortunately, the two interventions tested in this study do not posses this characteristic.

Finally, this study also shed light on the design of motivational messages. We learned that providing informative messages does not work as a persuasive mechanism. Sending feedback at a high frequency might also be perceived negatively by users and demotivate them in their goals of establishing a healthy exercise routine. Feedback should be used by app designers to *satisfy the basic need of competence*. However, for feedback to work, it has to provide useful information that can help the users reflect on their own performance and gain a sense of their effectiveness. An example of this task-inherent feedback could be *Well done! Today you walked 12345 steps. This week you improved your step activity by 3%, compared to last week. Keeping up this level of activity for an additional three weeks will result on an overall improvement of your muscular tone*. This example demonstrates the following features of feedback as required by SDT: *a. personal* It provides content that is specific to the participant; *b. contextual* it provides task-inherent information that can help the user connect their performance of the activity with its outcomes; and *c. goal-oriented* it provides the next challenge to push their work further by being phrased in a way that is specific to the level of the user. We provide implications about how motivational messages should be delivered. Using the results of the study, we argue that the delivery of feedback should not follow a regular schedule, as this could have negative consequences for the user's basic need of autonomy. Instead, to avoid repetition and the related feeling of being controlled, it would be more beneficial to *deliver the feedback opportunistically* when the user performs a spontaneous activity, perhaps at a time or place where this did not occur in the past.

4.6.2 Limitations

Our study has several limitations. First, the participants were all university students, which limits the generalization of our findings. The physical activity of the students is largely influenced by the academic calendar that defines examination and class periods, and summer/winter holiday seasons. Class sessions demand long sitting-hours, whereas holidays can lead to an increase in physical activity. Second, the study was limited to one country and region. Swiss students living in the French-speaking region of Switzerland have exercise opportunities and transportation habits different than students of the same age living in denser metropolitan areas, or in locations with more severe winters or more constrained recreational facilities. Third, the study was limited to one type of physical activity, which does not allow us to measure the complete spectrum of physical activities that the participants were involved into (e.g., swimming, cycling, etc). The interventions we tested in this experiment might have had an effect in promoting the overall physical activity, but this might have not been necessarily reflected by the number of steps of the participants. Given that the design of the interventions was focused on the number of steps, we consider this last point unlikely but not impossible. In this paper, we are unable to provide any data on what types of PA participants might have been involved in (other than walking and running) and whether their overall PA changed during the experiment. Fourth, participants' physical activities were not tracked if they did not carry their smart phones. These last two points might account for under-reported levels of PA by the participants in this study. Additionally, the study was limited to iPhone users, thus excluding other market segments[22]. Furthermore, we did not test a combination of tangible rewards and motivational messages in the same intervention. Finally, we acknowledge that the procedure we used to create the corpus of motivational messages is not perfect, and that in order to generate a more effective corpus, it could be ameliorated in future research. We suggest two ideas for improvement: (1) the effectiveness of each message can be achieved through Mechanical Turks; (2) as reported in the discussion, effective feedback should be personalized, contextual, and goal-oriented.

4.7 Conclusions to the Chapter

In this chapter, we shed light on the elements of design of persuasive apps in their ability to provoke change in behavior. We have provided evidence that contributes to the body of knowledge concerning the use of tangible rewards and motivational messages in order to increase physical activity. To the best of our knowledge, this is the first study comparing, in the same longitudinal study, these two persuasive techniques. Also, this is one of the first studies focusing specifically on the detrimental effects of these interventions on the intrinsic motivation of participants in the post-intervention phase. Although most of the prior work measured experimental effects on the main outcome variable (e.g., steps) during the intervention, we have reported and discussed the negative effects that these interventions can have in the long-term on the intrinsic motivation of the participants of the intervention. The specific contribution of this work is to add to the current discussion by suggesting that in order to determine whether an intervention is effective, we should look at both the main outcome measurement of PA and, more importantly, at the intrinsic motivation. Indeed, the intrinsic motivation has often been neglected, yet it determines the development of people's inner resources that sustain their self-determined actions.

Furthermore, we argue that only by embracing a holistic approach to these behavioral interventions can we significantly advance the design of persuasive technology. By this, we mean there is a need for a systematic experimentation and comparison

[22]It has to be reported that the iPhone has one of the largest market share in Switzerland. See https://www.nzz.ch/wirtschaft/unternehmen/apple-bleibt-platzhirsch-1.18563593, last retrieved June 2018.

of different designs over a longitudinal segment of time spanning several months. Furthermore, we argue that careful testing of the post-intervention phase of these interventions might reveal unexpected side effects.

An additional benefit of designing this kind of study is that we can expect to eventually reconcile the apparent differences between studies conducted in the domain of behavioral economics and psychology (e.g., the differences of results around rewards, see for instance the discussion reported by Gneezy et al. [Gneezy and Rustichini, 2000]).

For future work, we plan to replicate this study with a different population. In particular, it would be valuable to study whether the same effects would be found with participants with an active job or with older participants who are retired. Furthermore, we plan to conduct additional systematic evaluations of design elements that might contribute to satisfying the basic needs identified by the Self-Determination Theory. Future studies could include other standardized measurements that would help us to comprehend the perception of autonomy, competence and relatedness (e.g., Locus of Causality for Exercise Scale [Markland, 1999], Physical Self-Perception Profile [Fox and Corbin, 1989] and Social Support for Exercise Questionnaire [Sallis et al., 1987]). Another idea that emerged in this work was related to the concept of an optimal challenge. Future work should examine whether exposing participants to adaptable goals could increase relevance and decrease the feeling of being controlled perceived by some when feedback was delivered.

We learned that different feature implementations support the individual's BPNs differently; for instance, we can support *competence* by designing a feature that displays *historical data* or by creating a feature that rewards when a task is complete. This is where comparing the results of the present chapter and the previous one (chapter 3) results relevant because we observe that it is not sufficient to provide BPNs support with different features. In other words, different feature implementations might provide different levels of support and could even occur that one specific implementation causes detrimental effects on the individual's behavior.

4.7.1 Acknowledgments

This work was supported by the UNIL HEC research fund and by the research fund of HES-SO, Switzerland. We would like to thank several colleagues that contributed directly and indirectly to the development of the ideas reported in this paper: Valérie Chavez, José Mata, Luís Santos-Pinto, Kashyap Todi, and Roelof de Vries. A special thank goes to Katja Schwab-Weis for their great help in coordinating communications with the participants of this study. Finally, we would like to thank Holly Cogliati-Bauereis for her precious comments.

Fourth Chapter References

[Abadie, 2005] Abadie, A. (2005). Semiparametric Difference-in-Differences Estimators. *The Review of Economic Studies*, 72(1):1–19.

[Ajzen, 1985] Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In *Action control*, pages 11–39. Springer, Salmon Tower Building New York City, USA.

[Ajzen and Fishbein, 1980] Ajzen, I. and Fishbein, M. (1980). *Understanding attitudes and predicting social behaviour*. Prentice-Hall, Upper Saddle River, NJ, US.

[Ariely, 2010] Ariely, D. (2010). *Predictably irrational : the hidden forces that shape our decisions*. Harper Perennial, HarperCollins Publishers, 195 Broadway, New York, NY, 10007.

[Bandura, 1986] Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall, Inc., Englewood Cliffs, NJ, US.

[Barnes, Patricia M., Schoenborn, 2012] Barnes, Patricia M., Schoenborn, C. A. (2012). Trends in Adults Receiving a Recommendation for Exercise or Other Physical Activity From a Physician or Other Health Professional.

[Barte and Wendel-Vos, 2017] Barte, J. C. M. and Wendel-Vos, G. C. W. (2017). A Systematic Review of Financial Incentives for Physical Activity: The Effects on Physical Activity and Related Outcomes. *Behavioral Medicine*, 43(2):79–90.

[Bates et al., 2015] 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.

[Blundell et al., 2004] Blundell, R., Dias, M. C., Meghir, C., and van Reenen, J. (2004). Evaluating the employment impact of a mandatory job search program. *Journal of the European Economic Association*, 2(4):569–606.

[Bordalo et al., 2012] Bordalo, P., Gennaioli, N., and Shleifer, A. (2012). Salience theory of choice under risk. *The Quarterly journal of economics*, 127(3):1243–1285.

[Bravata et al., 2007] Bravata, D. M., Smith-Spangler, C., Sundaram, V., Gienger, A. L., Lin, N., Lewis, R., Stave, C. D., Olkin, I., and Sirard, J. R. (2007). Using pedometers to increase physical activity and improve health: a systematic review. *Jama*, 298(19):2296–2304.

[Buchan et al., 2012] Buchan, D. S., Ollis, S., Thomas, N. E., and Baker, J. S. (2012). Physical activity behaviour: An overview of current and emergent theoretical practices. *Journal of Obesity*, 2012:1–11.

[Burns et al., 2012] Burns, R. J., Donovan, A. S., Ackermann, R. T., Finch, E. A., Rothman, A. J., and Jeffery, R. W. (2012). A Theoretically Grounded Systematic Review of Material Incentives for Weight Loss: Implications for Interventions. *Annals of Behavioral Medicine*, 44(3):375–388.

[Burns and Rothman, 2018] Burns, R. J. and Rothman, A. J. (2018). Comparing Types of Financial Incentives to Promote Walking : An Experimental Test. *Applied Psychology: Health and Well-Being*, 10(2):1–22.

[Carpentier and Mageau, 2013] Carpentier, J. and Mageau, G. A. (2013). When change-oriented feedback enhances motivation, well-being and performance: A look at autonomy-supportive feedback in sport. *Psychology of Sport and Exercise*, 14(3):423–435.

[Charness and Gneezy, 2009] Charness, G. and Gneezy, U. (2009). Incentives to Exercise. *Econometrica*, 77(3):909–931.

[Daryanto et al., 2010] Daryanto, A., de Ruyter, K., Wetzels, M., and Patterson, P. G. (2010). Service firms and customer loyalty programs: A regulatory fit perspective of reward preferences in a health club setting. *Journal of the Academy of Marketing Science*, 38(5):604–616.

[De Vries et al., 2017] De Vries, R., Zaga, C., Bayer, F., Drossaert, C., Truong, K., and Evers, V. (2017). Experts get me started, peers keep me going: Comparing crowd- versus expert-designed motivational text messages for exercise behavior change. In *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*, pages 155–162, United States. Association for Computing Machinery (ACM).

[de Vries et al., 2016] de Vries, R. A., Truong, K. P., Kwint, S., Drossaert, C. H., and

Evers, V. (2016). Crowd-Designed Motivation. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, pages 297–308, New York, New York, USA. ACM Press.

[de Vries et al., 2017] de Vries, R. A. J., Truong, K. P., Zaga, C., Li, J., and Evers, V. (2017). 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.

[DeCharms, 1968] DeCharms, R. (1968). *Personal causation: the internal affective determinants of behavior*. Academic Press, Cambridge, Massachusetts, US.

[Deci and Ryan, 1985] Deci, E. and Ryan, R. M. (1985). *Intrinsic Motivation and Self-Determination in Human Behavior*. Springer US, New York, USA.

[Deci, 1971] Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of personality and Social Psychology*, 18(1):105.

[Deci, 1972] Deci, E. L. (1972). The effects of contingent and noncontingent rewards and controls on intrinsic motivation. *Organizational behavior and human performance*, 8(2):217–229.

[Deci, 1975] Deci, E. L. (1975). *Intrinsic Motivation*. Bibliographie pp.295. Plenum Publishing Company Limited, New York, USA.

[Deci et al., 1994] Deci, E. L., Eghrari, H., Patrick, B. C., and Leone, D. R. (1994). Facilitating internalization: the self-determination theory perspective. *Journal of personality*, 62(1):119–142.

[Deci and Moller, 2005] Deci, E. L. and Moller, A. C. (2005). The Concept of Competence: A Starting Place for Understanding Intrinsic Motivation and Self-Determined Extrinsic Motivation. In *Handbook of competence and motivation.*, pages 579–597. Guilford Publications, New York, NY, US.

[Deci and Ryan, 1980] Deci, E. L. and Ryan, R. M. (1980). The empirical exploration of intrinsic motivational processes1. In *Advances in experimental social psychology*, volume 13, pages 39–80. Elsevier, Atlanta, USA.

[Deci and Ryan, 2014] Deci, E. L. and Ryan, R. M. (2014). Autonomy and need satisfaction in close relationships: Relationships motivation theory. In *Human motivation and interpersonal relationships: Theory, research, and applications.*, pages 53–73. Springer

Science + Business Media, Salmon Tower Building New York City, USA.

[Deci et al., 1999] Deci, E. L., Ryan, R. M., and Koestner, R. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125(6):627–668.

[Dishman et al., 2004] Dishman, R. K., Motl, R. W., Saunders, R., Felton, G., Ward, D. S., Dowda, M., and Pate, R. R. (2004). Self-efficacy partially mediates the effect of a school-based physical-activity intervention among adolescent girls. *Preventive Medicine*, 38(5):628–636.

[Duncan et al., 2012] Duncan, L. R., Hall, C. R., Wilson, P. M., and Rodgers, W. M. (2012). The use of a mental imagery intervention to enhance integrated regulation for exercise among women commencing an exercise program. *Motivation and Emotion*, 36(4):452–464.

[ExerciseRewards, 2019] ExerciseRewards (2019). Exercise reards: it pays to be active.

[Finkelstein et al., 2008] Finkelstein, E. A., Brown, D. S., Brown, D. R., and Buchner, D. M. (2008). A randomized study of financial incentives to increase physical activity among sedentary older adults. *Preventive Medicine*, 47(2):182–187.

[Fitbit, 2019] Fitbit (2019). Fitbit.

[Fitzmaurice et al., 2008] Fitzmaurice, G., Davidian, M., Verbeke, G., and Molenberghs, G. (2008). *Longitudinal Data Analysis*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. CRC Press, 6000 Broken Sound Pkwy NW #300, Boca Raton, FL 33487, USA.

[Fjeldsoe et al., 2010] Fjeldsoe, B. S., Miller, Y. D., and Marshall, A. L. (2010). Mobilemums: a randomized controlled trial of an sms-based physical activity intervention. *Annals of Behavioral Medicine*, 39(2):101–111.

[FoodStand, 2019] FoodStand (2019). FoodStand: Ditch dieting. Build healthy habits instead.

[for Disease Control and Prevention, 2014] for Disease Control, C. and Prevention (2014). The State Indicator Report on Physical Activity.

[Fox and Corbin, 1989] Fox, K. R. and Corbin, C. B. (1989). The physical self-perception profile: Devlopment and preliminary validation. *Journal of sport and Exercise Psychology*, 11(4):408–430.

[Frank, 2017] Frank, J. (2017). A personalized support tool for the training of mindful walking: The mobile "MindfulWalk" application. Master's thesis, Ulm University.

[Geller et al., 2011] 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.

[Gneezy and Rustichini, 2000] Gneezy, U. and Rustichini, A. (2000). Pay enough or don't pay at all. *The Quarterly Journal of Economics*, 115(3):791–810.

[Goldberg, 1992] Goldberg, L. R. (1992). The development of markers for the big-five factor structure. *Psychological assessment*, 4(1):26.

[GymPact, 2019] GymPact (2019). GymPact: Incentivize your exercise.

[Hanus and Fox, 2015] Hanus, M. D. and Fox, J. (2015). Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Computers & Education*, 80:152–161.

[Harries et al., 2016] Harries, T., Eslambolchilar, P., Rettie, R., Stride, C., Walton, S., and Van Woerden, H. C. (2016). Effectiveness of a smartphone app in increasing physical activity amongst male adults: A randomised controlled trial. *BMC Public Health*, 16(1):1–10.

[Hausenblas et al., 1997] Hausenblas, H. A., Carron, A. V., and Mack, D. E. (1997). Application of the theories of reasoned action and planned behavior to exercise behavior: A meta-analysis. *Journal of Sport and Exercise Psychology*, 19(1):36–51.

[HealthyWage, 2019] HealthyWage (2019). HealthyWage: weight loss challenges with cash prices.

[Hedeker and Gibbons, 2006] Hedeker, D. and Gibbons, R. D. (2006). Longitudinal data analysis.

[Helsana+, 2019] Helsana+ (2019). Helsana+ Programme de bonus.

[Hewett and Conway, 2016] Hewett, R. and Conway, N. (2016). The undermining effect revisited: The salience of everyday verbal rewards and self-determined motivation. *Journal of Organizational Behavior*, 37(3):436–455.

[Hurling et al., 2007] Hurling, R., Catt, M., De Boni, M., Fairley, B. W., Hurst, T., Murray, P., Richardson, A., and Sodhi, J. S. (2007). Using internet and mobile phone technology to deliver an automated physical activity program: randomized controlled trial. *Journal of medical Internet research*, 9(2):88–93.

[Iwane et al., 2000] Iwane, M., Arita, M., Tomimoto, S., Satani, O., Matsumoto, M., Miyashita, K., and Nishio, I. (2000). Walking 10, 000 steps/day or more reduces blood pressure and sympathetic nerve activity in mild essential hypertension. *Hypertension Research*, 23(6):573–580.

[Jeffery et al., 1998] Jeffery, R. W., Wing, R. R., Thorson, C., and Burton, L. R. (1998). Use of personal trainers and financial incentives to increase exercise in a behavioral weight-loss program. *Journal of consulting and clinical psychology*, 66(5):777.

[Jeong et al., 2017] Jeong, H., Kim, H., Kim, R., Lee, U., and Jeong, Y. (2017). Smartwatch Wearing Behavior Analysis. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(3):1–31.

[Kahneman and Tversky, 1979] Kahneman, D. and Tversky, A. (1979). On the interpretation of intuitive probability: A reply to jonathan cohen. *Cognition*, 7(4):409–411.

[Kim and Glanz, 2013] Kim, B. H. and Glanz, K. (2013). Text messaging to motivate walking in older african americans: a randomized controlled trial. *American Journal of Preventive Medicine*, 44(1):71–75.

[Kinnafick et al., 2016] Kinnafick, F. E., Thøgersen-Ntoumani, C., and Duda, J. (2016). The effect of need supportive text messages on motivation and physical activity behaviour. *Journal of Behavioral Medicine*, 39(4):574–586.

[Lenth, 2016] Lenth, R. V. (2016). Least-squares means: The `R` package `lsmeans`. *Journal of Statistical Software*, 69(1):1–33.

[Liu and Chan, 2016] Liu, C. T. and Chan, C. T. (2016). Exercise performance measurement with smartphone embedded sensor for well-being management. *International Journal of Environmental Research and Public Health*, 13(10):1–13.

[Locke and Latham, 2002] 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):705–717.

[Loewenstein et al., 2013] Loewenstein, G., Asch, D. A., and Volpp, K. G. (2013). Behavioral economics holds potential to deliver better results for patients, insurers, and employers. *Health Affairs*, 32(7):1244–1250.

[Loewenstein et al., 2003] Loewenstein, G., O'Donoghue, T., and Rabin, M. (2003). Projection Bias in Predicting Future Utility. *The Quarterly Journal of Economics*, 118(4):1209–1248.

[Markland, 1999] Markland, D. (1999). Self-determination moderates the effects of perceived competence on intrinsic motivation in an exercise setting. *Journal of Sport and Exercise Psychology*, 21(4):351–361.

[McAuley et al., 1989] McAuley, E., Duncan, T., and Tammen, V. V. (1989). Psychometric Properties of the Intrinsic Motivation Inventory in a Competitive Sport Setting: A Confirmatory Factor Analysis. *Research Quarterly for Exercise and Sport*, 60(1):48–58.

[McAuley et al., 2003] McAuley, E., Jerome, G. J., Marquez, D. X., Elavsky, S., and Blissmer, B. (2003). Exercise self-efficacy in older adults: social, affective, and behavioral influences. *Annals of behavioral medicine : a publication of the Society of Behavioral Medicine*, 25(1):1–7.

[Mitchell et al., 2013] Mitchell, M. S., Goodman, J. M., Alter, D. A., John, L. K., Oh, P. I., Pakosh, M. T., and Faulkner, G. E. (2013). Financial incentives for exercise adherence in adults: systematic review and meta-analysis. *American journal of preventive medicine*, 45(5):658–667.

[Moller et al., 2012] Moller, A. C., McFadden, H., Hedeker, D., and Spring, B. (2012). Financial motivation undermines maintenance in an intensive diet and activity intervention. *Journal of obesity*, 2012:810–827.

[Moustaka et al., 2012] Moustaka, F. C., Vlachopoulos, S. P., Kabitsis, C., and Theodorakis, Y. (2012). Effects of an autonomy-supportive exercise instructing style on exercise motivation, psychological well-being, and exercise attendance in middle-age women. *Journal of Physical Activity and Health*, 9(1):138–150.

[Mozaffarian et al., 2015] Mozaffarian, D., Benjamin, E. J., Go, A. S., Arnett, D. K., Blaha, M. J., Cushman, M., De Ferranti, S., Després, J.-P., Fullerton, H. J., Howard, V. J., et al. (2015). Heart disease and stroke statistics–2015 update: a report from the american heart association. *Circulation*, 131(4):e29–e322.

[Mullan et al., 1997] Mullan, E., Markland, D., and Ingledew, D. K. (1997). A graded conceptualisation of self-determination in the regulation of exercise behaviour: Development of a measure using confirmatory factor analytic procedures. *Personality and Individual Differences*, 23(5):745–752.

[Nolan et al., 2014] Nolan, M., Mitchell, J. R., and Doyle-Baker, P. K. (2014). Validity of the Apple iPhone®/iPod Touch®as an Accelerometer-Based Physical Activity Monitor: A Proof-of-Concept Study. *Journal of Physical Activity and Health*, 11(4):759–769.

[O'Donoghue and Rabin, 2000] O'Donoghue, T. and Rabin, M. (2000). The economics of immediate gratification. *Journal of Behavioral Decision Making*, 13(2):233.

[Oesch, 2006] Oesch, D. (2006). Coming to Grips with a Changing Class Structure: An Analysis of Employment Stratification in Britain, Germany, Sweden and Switzerland. *International Sociology*, 21(2):263–288.

[Oga-Baldwin et al., 2017] Oga-Baldwin, W. Q., Nakata, Y., Parker, P., and Ryan, R. M. (2017). Motivating young language learners: A longitudinal model of self-determined motivation in elementary school foreign language classes. *Contemporary Educational Psychology*, 49:140–150.

[Orji et al., 2017a] Orji, R., Mandryk, R. L., and Vassileva, J. (2017a). Improving the Efficacy of Games for Change Using Personalization Models. *ACM Transactions on Computer-Human Interaction*, 24(5):1–22.

[Orji et al., 2017b] Orji, R., Nacke, L. E., and Di Marco, C. (2017b). Towards personality-driven persuasive health games and gamified systems. In *Proceedings of CHI'17*, pages 1015–1027, New York, USA. ACM.

[Orji et al., 2014] Orji, R., Vassileva, J., and Mandryk, R. L. (2014). Modeling the efficacy of persuasive strategies for different gamer types in serious games for health. *User Modeling and User-Adapted Interaction*, 24(5):453–498.

[Orlick and Mosher, 1978] Orlick, T. D. and Mosher, R. (1978). Extrinsic awards and participant motivation in a sport related task. *International Journal of Sport Psychology*, 9(1):27–39.

[Owen et al., 2014] Owen, K. B., Smith, J., Lubans, D. R., Ng, J. Y., and Lonsdale, C. (2014). Self-determined motivation and physical activity in children and adolescents: A systematic review and meta-analysis. *Preventive medicine*, 67:270–279.

[Patel et al., 2016] Patel, M. S., Asch, D. A., Rosin, R., Small, D. S., Bellamy, S. L., Heuer, J., Sproat, S., Hyson, C., Haff, N., Lee, S. M., Wesby, L., Hoffer, K., Shuttleworth, D., Taylor, D. H., Hilbert, V., Zhu, J., Yang, L., Wang, X., and Volpp, K. G. (2016). Framing Financial Incentives to Increase Physical Activity Among Overweight and Obese Adults. *Annals of Internal Medicine*, 164(6):385.

[Patel et al., 2017] Patel, M. S., Foschini, L., GW, K., and et Al (2017). Using wearable devices and smartphones to track physical activity: Initial activation, sustained use, and step counts across sociodemographic characteristics in a national sample. *Annals of Internal Medicine*, 167(10):755–757.

[Patel et al., 2018] Patel, M. S., Volpp, K. G., Rosin, R., Bellamy, S. L., Small, D. S., Heuer, J., Sproat, S., Hyson, C., Haff, N., Lee, S. M., Wesby, L., Hoffer, K., Shuttleworth, D., Taylor, D. H., Hilbert, V., Zhu, J., Yang, L., Wang, X., and Asch, D. A. (2018). A Randomized, Controlled Trial of Lottery-Based Financial Incentives to Increase Physical Activity Among Overweight and Obese Adults.

[Plant and Ryan, 1985] Plant, R. W. and Ryan, R. M. (1985). Intrinsic motivation and the effects of self-consciousness, self-awareness, and ego-involvement: An investigation of internally controlling styles. *Journal of Personality*, 53(3):435–449.

[Porter and Whitcomb, 2003] Porter, S. R. and Whitcomb, M. E. (2003). The Impact of Lottery Incentives on Student Survery Response Rates. *Public Opinion Quarterly*, 44(4):389–407.

[Prochaska and Di Clemente, 1983] Prochaska, J. O. and Di Clemente, 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):390–395.

[Qin et al., 2010] Qin, H., Rau, P.-L. P., and Salvendy, G. (2010). Effects of different scenarios of game difficulty on player immersion. *Interacting with Computers*, 22(3):230–239.

[Reinboth and Duda, 2006] Reinboth, M. and Duda, J. L. (2006). Perceived motivational climate, need satisfaction and indices of well-being in team sports: A longitudinal perspective. *Psychology of Sport and Exercise*, 7(3):269–286.

[Resnick and Spellbring, 2000] Resnick, B. and Spellbring, A. M. (2000). Understanding what motivates older adults to exercise. *Journal of gerontological nursing*, 26(3):34–42.

[Riemsma et al., 2002] Riemsma, R., Pattenden, J., Bridle, C., Sowden, A., Mather, L., Watt, I., Walker, A., et al. (2002). A systematic review of the effectiveness of interventions based on a stages-of-change approach to promote individual behaviour change. *Health Technology Assessment*, 24:1–231.

[Rigby and Ryan, 2011] Rigby, S. and Ryan, R. M. (2011). *Glued to games: How video games draw us in and hold us spellbound.* New directions in media. Praeger/ABC-CLIO, Santa Barbara, CA, US.

[Runkeeper, 2019] Runkeeper (2019). Runkeeper: Everyone, every run.

[Ryan, 1982] Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of personality and social psychology*, 43(3):450.

[Ryan, 1993] Ryan, R. M. (1993). Agency and organization: Intrinsic motivation, autonomy, and the self in psychological development. In *Nebraska Symposium on Motivation, 1992: Developmental perspectives on motivation.*, Current theory and research in motivation, Vol. 40., pages 1–56. University of Nebraska Press, Lincoln, NE, US.

[Ryan et al., 1990] Ryan, R. M., Connell, J. P., and Plant, R. W. (1990). Emotions in nondirected text learning.

[Ryan and Deci, 2000a] Ryan, R. M. and Deci, E. L. (2000a). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *The American psychologist*, 55(1):68–78.

[Ryan and Deci, 2000b] Ryan, R. M. and Deci, E. L. (2000b). When rewards compete with nature. In *Intrinsic and Extrinsic Motivation*, pages 13–54. Elsevier, London, UK.

[Ryan and Deci, 2007] Ryan, R. M. and Deci, E. L. (2007). *Active human nature: Self-determination theory and the promotion and maintenance of sport, exercise, and health*. Human Kinetics Publishers, 107 Bradford Road Stanningley, Leeds LS28 6AT, United Kingdom.

[Ryan and Deci, 2017] Ryan, R. M. and Deci, E. L. (2017). *Self-Determination Theory Basic Psychological Needs in Motivation, Development and Wellness*. The Guilford Press, New York, USA.

[Ryan and Frederick, 1997] Ryan, R. M. and Frederick, C. (1997). On energy, personality, and health: Subjective vitality as a dynamic reflection of well-being. *Journal of* *personality*, 65(3):529–565.

[Ryan et al., 1991] Ryan, R. M., Koestner, R., and Deci, E. L. (1991). Ego-involved persistence: When free-choice behavior is not intrinsically motivated. *Motivation and Emotion*, 15(3):185–205.

[Ryan et al., 1983] Ryan, R. M., Mims, V., and Koestner, R. (1983). Relation of reward contingency and interpersonal context to intrinsic motivation: A review and test using cognitive evaluation theory. *Journal of personality and Social Psychology*, 45(4):736.

[Sallis et al., 1987] Sallis, J. F., Grossman, R. M., Pinski, R. B., Patterson, T. L., and Nader, P. R. (1987). The development of scales to measure social support for diet and exercise behaviors. *Preventive medicine*, 16(6):825–836.

[Smith and Sarason, 1975] Smith, R. E. and Sarason, I. G. (1975). Social anxiety and the evaluation of negative interpersonal feedback. *Journal of consulting and Clinical Psychology*, 43(3):429.

[Standage et al., 2008] Standage, M., Sebire, S. J., and Loney, T. (2008). Does exercise motivation predict engagement in objectively assessed bouts of moderate-intensity exercise?: A self-determination theory perspective. *Journal of Sport and exercise Psychology*, 30(4):337–352.

[SweatCoin, 2019] SweatCoin (2019). SweatCoin: Sweat for coins.

[Szumilas, 2010] Szumilas, M. (2010). Explaining odds ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, 19(3):227–229.

[Teixeira et al., 2012] Teixeira, P. J., Carraça, E. V., Markland, D., Silva, M. N., and Ryan, R. M. (2012). Exercise, physical activity, and self-determination theory: a systematic review. *International Journal of Behavioral Nutrition and Physical Activity*, 9(1):78.

[Thøgersen-Ntoumani and Ntoumanis, 2006] Thøgersen-Ntoumani, C. and Ntoumanis, N. (2006). The role of self-determined motivation in the understanding of exercise-related behaviours, cognitions and physical self-evaluations. *Journal of sports sciences*, 24(4):393–404.

[Thompson et al., 2014] Thompson, D., Cantu, D., Bhatt, R., Baranowski, T., Rodgers, W., Jago, R., Anderson, B., Liu, Y., Mendoza, J. A., Tapia, R., and Buday, R. (2014).

Texting to Increase Physical Activity Among Teenagers (TXT Me!): Rationale, Design, and Methods Proposal. *JMIR Research Protocols*, 3(1):e14.

[Tudor-Locke and Bassett, 2004] Tudor-Locke, C. and Bassett, D. R. (2004). How many steps/day are enough? *Sports medicine*, 34(1):1–8.

[WayBetter, 2019] WayBetter (2019). DietBet.

[Wilson and Rodgers, 2004] Wilson, P. M. and Rodgers, W. M. (2004). The relationship between perceived autonomy support, exercise regulations and behavioral intentions in women. *Psychology of Sport and Exercise*, 5(3):229–242.

[Wood, 2017] Wood, S. (2017). *Generalized Additive Models: An Introduction with R, Second Edition*. Chapman & Hall/CRC Texts in Statistical Science. CRC Press, Boca Raton, Florida, US.

[Xu et al., 2017] Xu, T., Guo, A., and Ma, J. (2017). Analysis of Temporal Features in Data Streams from Multiple Wearable Devices. In *2017 3rd IEEE International Conference on Cybernetics (CYBCONF)*, pages 1–6, IEEE Operations Center 445 Hoes Lane Piscataway, NJ 08854-4141 USA. IEEE.

Chapter 5

General Thesis Discussion

This chapter discusses five main ideas that distill across the results of the four studies presented in this thesis. These ideas revolve around app feature embodiments, the SDT Basic Psychological Needs, and how our app features supported these. Also, we examine why this support did not work well. Later, we introduce the generalization of the thesis findings, we discuss around behavior change and society, present the limitations of this thesis work and conclude by framing future directions for research and design.

5.1 Thesis Research Questions

With this thesis work, we answered the research questions presented in the Introduction. In the following lines, we synthesize the answer to each question.

RQ1: Can the Self-Determination Theory suggest characteristics that behavior change apps would need to possess to support behavior change interventions? Do behavior change apps on the market possess these?

Behavior change apps have features that the SDT Basic Psychological Needs constructs can inform. These constructs specify the importance of autonomous actions, activities that exercise competence, and the importance of developing a sense of belonging and relation with others. These characteristics were mapped to the APPLE APP STORE app features. Therefore, we conclude that the SDT can suggest characteristics that behavior change apps should have to support interventions. Furthermore, our app review concluded that only 25% of our sample included features supporting the three BPNs. The work presented in chapter 1 answers this research question. It contributes with an *artifact* represented by a design tool to evaluate how the app features and combination of these may support the motivational process of individuals towards achieving their goals. The work presented in chapter 1 guided us on the design of behavior change features. To study the behavioral effects of these features, we implemented and tested an app, and then we proposed the following research question.

RQ2: What are the individuals' perceptions about the hypothesized app features that aim to support autonomy, competence, and relatedness when it comes to improving their physical activity?

The app we designed included hypothesized features that aim to support the BPNs. During four weeks, participants used the app. This time allowed enough exposure for them to develop perceptions of the app features. We captured these through interviews and diary entries. We noticed that participants perceived support to their BPNs related to the specific design of our app. The work presented in chapter 2 answers this research question. It contributes with *empirical* evidence that the features of the research app correctly mapped to the BPNs. Having a correct map between features and perceptions is a solid first step towards designing behavior change features, but it is not enough. It is essential to test if these perceptions also lead to a change in the individual's behavior. Therefore, in the subsequent study, we posed the following research question.

RQ3: What are the BPNs feature combinations that produce the most positive effects on the individuals' behavior when it comes to improving their physical activity?

We deployed seven combinations of BPNs supportive app features. During four months of experiment intervention, we captured objective data (number of steps) and subjective data (Intrinsic Motivation), which answered our research question. The work presented in chapter 3 answers this research question. It contributes with *empirical* evidence that demonstrates the positive effect of the *competence* and *relatedness* BPNs feature combination on the physical activity level, intrinsic motivation, and engagement of our participants.

5.2 Behavior Change Feature Embodiments

When designing behavior change app features, it is pertinent to understand the role of different feature embodiments or implementations. The fact that two different features support the same BPNs does not guarantee that they do it at the same level or quality. For instance, two leader boards might support *relatedness*. However, when comparing the two embodiments, one leader board lists user pseudonyms, and the other lists the users' real names. These different embodiments might be perceived differently and produce different behaviors. Therefore, having multiple embodiments of an app feature can cause two scenarios: (1) the feature can go unnoticed, meaning that the individual does not perceive it; (2) the feature is perceived, but the level of support is not enough to provoke a behavior change. Therefore it is not enough to have features that support the BPNs to cause behavior change; app creators should guarantee that app features are noticed and support the behavior.

5.3 Towards Autonomy Supportive Design

The work presented in chapter 1 helped us distinguish app features that might support the Basic Psychological Needs. For example, an app function that enables users to set up their goals might provide support to the *autonomy* need. Therefore, in chapter 2 we considered this finding to design and test a weekly goal-setting feature that provided information to users about their physical activity and enabled them to choose a step-goal. The results of this chapter demonstrated that informed and personalized choices supported *autonomy*. We remark that not all goal-setting feature embodiments will support the *autonomy* need. The specific design we tested in chapter 2 suggested having achieved it. Further, because the intervention was designed to support autonomous motivation as opposed to controlled motivation, we expected the sources of motivation were *somewhat internal* or *internal* [Ryan and Deci, 2000]. The findings presented in chapter 2 showed that participants expressed having a sense of autonomy while selecting their weekly goal, and they did not express emotions of punishments or compliance, which relates to controlled motivation. These findings support employing our specific feature design to promote autonomous motivation. Furthermore, participants accepting or denying a goal increase based on their past week's performance might have also supported their need for *competence*. When they felt effective in achieving their weekly goal, they aimed to increase it for the following week. Thus in chapter 3, we introduced minimal modifications to this goal-setting feature aiming to extend the support to *autonomy*. Specifically, we provided three options of goal increments (i.e., 0%, 5%, or 10%). The weekly goal-setting feature introduced in chapter 2 allowed participants to agree or disagree with the goal suggested by the app. Therefore, in this new design version, we expected that participants would experience more freedom in their goal-setting, which would result in a higher impact on motivation and performance. However, we realized that the support provided by this particular design did not provoke observable changes in terms of physical activity levels. These specific design nuances should be considered when aiming to replicate our results. Testing other variations of these designs might yield different results (e.g., allowing users to set specific goal increments instead of having three options to choose from).

Now we discuss several reasons that might cause this lack of observable changes on our participants. First, in the design of our research app, participants were exposed daily to the features supporting *competence* and *relatedness*. However, the exposure to the *autonomy* feature was for a short period (i.e., < 5 minutes) and once a week. This short exposure time window left our design unbalanced concerning how much support each Basic Psychological Need received. Participants might indeed have experienced a sense of autonomy while selecting their weekly step-goal. However, we also believe that this fulfillment occurs in participants already naturally interested in performing a walking activity. However, on the contrary, if the participant finds more appealing another physical activity (e.g., cycling, basketball), they might not be interested in setting goals for an activity they are not excited about.

Another reason for the lack of observable changes is that the current design implies a continuous growth model in which participants should constantly increase their performance week-on-week. However, given personal time constraints (e.g., school work, sickness, running errands), performance might quickly reach a plateau, and therefore the feature might expose the users to the situation of repeatedly being unable to make progress with their physical activity. This situation might have an overall detrimental effect on performance. Furthermore, due to this incremental goal-setting model (where the goal is set based on the past week's performance), their weekly steps averages significantly increased when participants were involved in unusual sports events (e.g., sports competitions, weekend hikes). Because the goal-setting feature did not have an option to set a smaller goal than the weekly average, participants were forced to set unrealistic goals for the following week, which might negatively affect their motivation and performance. This negative effect can be explained considering the source of motivation (or regulation) shifted from autonomous to controlled. For instance, participants might have experienced interest and enjoyment (autonomous motivation) in the walking activity during the first weeks of selecting goals. However, because of the *continuous growth model*, they were forced to comply with increasing their weekly goal, causing them to act out of controlled motivation. Furthermore, this continuous growth model might direct participants to an excessive practice of physical activity, which might lead to harmful outcomes and lead to abrupt dropout [Zhu et al., 2019]. Therefore, future studies implementing this model should consider users' contextual and personal restrictions to avoid unnecessary pressure to perform physical exercise. Also, another variation of this model might suggest a regulated amount of exercise sessions to the users to help them avoid falling into excesses of physical activity.

Therefore, while this series of experiments indicate that providing choices to users indeed supports *autonomy*, the design tested in chapter 3, demonstrates that we are still unable to support autonomy over long periods properly. These findings call for more design explorations to support *autonomy*. For instance, a future design modification for this feature could allow people to choose the type of activity they want to perform, and then setting the goals according to that activity would yield more robust autonomy support. Another potential design might consider grabbing ideas from the *motivational messages* taxonomy category and aim at showing the *value* of the targeted activity to the users, for example, by displaying information capsules showing the benefits of the intended activity (See Figure 5.1 a)). A different design approach could detect when the previous week's average increased dramatically and display to the user a goal-setting option without considering their past week's performance and instead of selecting a specific goal themselves. In simple words, it is giving the user a choice to select when and which goal they want to achieve (See Figure 5.1 b) and c)).

5.4 Combined Effect of Supporting Multiple BPNs

The preliminary results presented in chapter 3 cautions about providing multiple support for the Basic Psychological Needs. The message is: more is not always better. As we could observe, if the support is not correct (e.g., weekly goal-setting feature), the detrimental effect on one feature might completely overshadow the positive effect

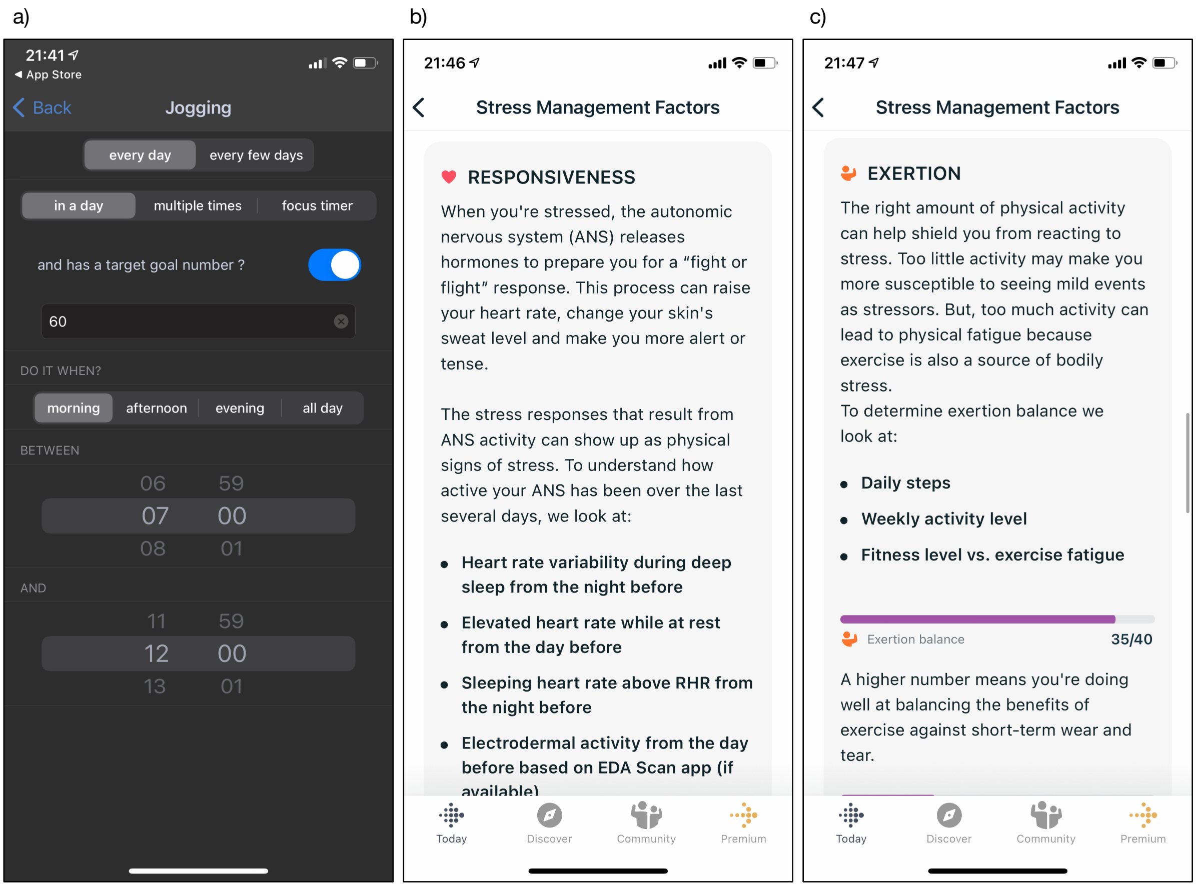

Figure 5.1

These figures serve as inspiration for exploring future *autonomy* supportive designs. Figure a) Shows a goal-setting feature that allows customization of the type of goal, frequency, execution time-frame, and target goal number. Figure b) and c) show information about stress factors that relate to the importance of physical activity.

of the other features. This is observed on the *competence* and *relatedness* support experimental conditions versus the group with full support (i.e., *ACR*).

Going forward, we advise designers to thoroughly integrate support for the Basic Psychological Needs in their apps. Designers should test features individually and then combine them with other features to reveal possible detrimental effects.

These results call for modular design of behavior-change apps where designers can employ a design through theory approach similar to the one we followed in chapter 2. This approach will help create multiple feature combinations (e.g., different apps supporting autonomy with different features like goal-setting and reminders); and multiple support (e.g., an app that supports relatedness by incorporating leader boards and performance-sharing simultaneously). A modular design would allow designers to easily add or remove features and explore the effects of having various features supporting the same BPNs. Furthermore, a modular design would facilitate iterative testing, which aids the refinement of the feature support for each need. Furthermore, iterative testing allows app creators to detect design errors and adjust accordingly promptly.

5.5 Supporting Relatedness Might Require Rich Interactions

The results from chapter 2, showed that the research app only partially afforded support to the *relatedness* Basic Psychological Need. This can be explained by looking at the specific design characteristics of the peer comparison feature. Notably, the embodiment of this feature in chapter 2 was limited to displaying a list of names and their steps. However, there was no further information that could allow individuals to feel connected to each other. Also, there was no way to communicate between themselves, which could also aid in creating a relation with the other participants. Further, because of the way this specific feature was embodied in this study, it might have also supported the BPNs of *competence* by allowing participants to compare their performance to that of other participants. Therefore, in the following version of this app presented in chapter 3, we modified the design to better support this need. Specifically, we extended the peers-comparison feature by implementing an interaction that when a user taps on the other participant's nickname, the app displays the participant's information card, including their avatar, nickname, age, and hobbies. The goal was for participants to have more information about the others and not limit themselves just to see their nicknames. We also implemented team competitions, and through the course of the study, participants remained on the same team and competed against different teams every week, all this with the purpose that participants develop a sense of belonging.

However, from our preliminary results, we noticed that supporting *relatedness* alone does not provoke observable changes in terms of physical activity levels. We do observe a difference when this feature is combined with support to *competence* need. The addition that the *competence* feature brings to the design is that participants receive feedback in the form of trophies for every weekly competition won; this success confirmation was absent on the *relatedness* alone design.

Likewise, we highlight that these particular preliminary results distill from the specific peer comparison feature we tested. Going forward, designers might explore alternative designs to increase relatedness support that might not require feedback—for example, giving social support by implementing ways to interact with each other by chatting or receiving and sending kudos (e.g., thumbs-up, congratulations message). These new functionalities might increase the support for relatedness and make this feature useful as a stand-alone.

5.6 Behavior Change and Society

In this thesis, we frame *behavior change* in a positive scope, meaning that it always aims for an improvement in the individual's life. Therefore, it is imperative to consider each person's particularities in how behavior change interventions are framed to seek this improvement. For instance, individuals with special cardiac conditions should not exceed a determined physical exercise intensity. Shortly, behavior change interventions should also be designed to be more inclusive and equitable by allowing individuals with specific physical and health conditions to improve their behavior. For instance, designing for individuals with reduced mobility or physical impairment. Another aspect that should be considered when designing behavior change interventions is maintaining a balance on the number of activity sessions. For instance, the individual does not incur extreme efforts that can have detrimental effects instead of improving their health.

5.7 Limitations of this Work

The work presented in chapter 1 is limited in the sample of apps that we analyzed, given that we explored *free* apps from the APPLE APP STORE only. The features analysis was also time limited—the apps were tested for several hours across multiple weeks, which provided us with a longitudinal exposure to the features. Therefore our analysis does not take into account tailoring strategies on long-term interventions. Finally, we did not examine wearables that work in parallel with the behavior change apps. The results from our work in chapter 2 can not be generalized to other domains due to the nature qualitative nature of the study. We capture participants' perceptions about the app features. For future research, a longitudinal study might be needed to capture objective data and more details concerning the efficacy of the proposed design. Also, our app is limited to one particular implementation of BPNs supportive features. It would be necessary for future research to design and compare alternative designs. Because the participants involved in this study are university students between 20 and 30 years old, our findings can not be generalized to older users. The work presented in chapter 3 has similar limitations to its previous chapter. In addition, the step-count measurements captured in this study are limited to the fact that participants must carry their phones with them; therefore, this might cause our steps-data to lack precision. Furthermore, the behavior changes proposed in this thesis are framed to produce a positive impact on the individual's life. Also, in all our interventions, individuals are aware they are part of an intervention, and we did not implement any subliminal cue. However, we acknowledge that these thesis findings can be employed to harm users or produce adverse effects on them (e.g., supporting addictions, fostering unhealthy competitions, disclosing personal data). For example, a collecting rewards feature can be used by app designers to encourage customers to increase tobacco consumption.

5.8 Future Directions for Research and Design

The findings of this thesis can be generalized to different domains that support other positive behavior changes. For instance, goal-setting features can be employed in the yoga practice, allowing individuals to set weekly objectives for meditation sessions and set the frequency and duration. Similarly, in the healthy eating domain, rewards can support individuals in improving the food intakes' quality, for instance, by collecting a certain number of points that grant them a coupon for lunch at a healthy restaurant. Furthermore, these thesis findings are strictly related to the specific embodiment of each of the app features, and different embodiments likely yield different behavioral results. Therefore, in the following paragraphs, we present some design opportunities that distill from the findings of this thesis.

- *Goal-Setting*: Concerning the goal-setting feature, a new design could allow more flexibility in the specification of the goal so that it does not force a constant goal increase but also allows people to set specific objectives that adapt to their necessities. Moreover, individuals might benefit from a design that allows them to set their goals using different metrics like time and distance. Having this possibility may benefit individuals with reduced mobility who want to exercise but for which the metric of steps is not applicable. Another design variation that may benefit the working population is choosing the days when they want to perform physical activities to align with their work demands and allow them to choose resting days.

- *Rewards*: Concerning rewards, a new design approach may incorporate *leveling* by adding categories of rewards (e.g., having gold, silver, and bronze stars). This design may counter losing interest in earning rewards (stars) because it is monotonous and dull. Moreover, another design opportunity may look at the use of *universal design* to consider individuals with visual impairments and select the best color combination rewards leveling should have. Further, another design opportunity is to incorporate real-time feedback when the individual achieves a goal. This design might yield a positive reinforcement caused by receiving a reward as the user might relate it to the action they are performing at a specific time. Also, an additional design could incorporate a record or history of rewards, allowing individuals to track the achieved goals visually.
- *Peers comparison* Concerning peers comparison, several options can be explored, for instance, enhancing the present design's personalization by including relevant data that individuals like to share (e.g., achievements, purpose to perform physical exercise, performance). This new design could also incorporate privacy configurations so that individuals can control the information they share. Further, another design approach that could be explored includes adding communication functionalities between participants (e.g., chat) or mechanisms to support each other (e.g., sending kudos or thumbs-up).

Furthermore, looking at other future research opportunities, we acknowledge that we explored the effects of SDT supporting features in the physical activity domain in this thesis. Future research calls for exploration in other areas such as education, working environments, and disease prevention.

Also, in this thesis, we tested specific features, future research could support the exact needs with other app designs, for instance, testing how *reminders* and *pre-commitments* might support *autonomy*; and how *self-monitoring* would support *competence*.

Also, other venues that can be explored in the future include understanding how the different personality types affect the perceptions and support of BPNs supportive features. For instance, some people might feel more comfortable having self-imposed goals, while others prefer an app to define their activity objectives for them.

Another area that would be interesting to explore is designing not only for the satisfaction of the BPNs but also for engagement—considering apps that are not static through time but that present variation support users' interest at different points and can re-engage them.

5.9 Conclusion

The results presented in this thesis pave the path towards designing behavior change apps. We achieved this by testing specific feature designs and capturing the individuals' perceptions and objective behaviors. Other researchers and app designers can leverage our methodology in translating psychological theory into app design to create and test new feature embodiments. Shortly, researchers should test multiple feature designs and combinations to continue expanding the knowledge of behavior change technology.