Designing, Developing, and Implementing a Personalized Gamified Goal-setting Mechanism for a Sleep-tracking Mobile Application

by

Yannick Zapfe

Thesis of 30 ECTS credits submitted to the School of Technology,
Department of Computer Science
at Reykjavík University and
the Department of Information Technology,
Faculty of Science and Engineering
at Åbo Akademi University in partial fulfillment
of the requirements for the degree of
Master of Science (M.Sc.) in Intelligent Software Systems

May 2023

Examining Committee:

Anna Sigríður Islind, Supervisor
Associate Professor, Reykjavík University, Iceland

Lisa Schmitz, Co-supervisor
PhD Student, Reykjavík University, Iceland

Erna Sif Arnardóttir, Examiner
Associate Professor, Reykjavík University, Iceland

Dragos Truscan, Examiner
Senior University Lecturer, Åbo Akademi University, Finland

Designing, Developing, and Implementing a Personalized Gamified Goal-setting Mechanism for a Sleep-tracking Mobile Application

Yannick Zapfe

May 2023

Abstract

Goal setting is a commonly employed game element and an implicit effect of gamification. Research in goal-setting theory has found specific and difficult goals to increase the effectiveness of mobile health interventions. Previous work has indicated a lack of research about the impact of different types of goals, especially those with time constraints. This work introduces a goal-setting mechanism with two types of goals, namely continuous and time-bound, to the Sleep Revolution app, a mobile application for sleep healthcare that enables users to set goals regarding several activities they track in the app's sleep diary. The implementation follows concepts and recommendations from previous work. To assess and compare the effects of the different goals on user compliance with health recommendations, a four-week randomized controlled trial was conducted. In a second step, the feasibility of extending the mechanism with personalized goal recommendations based on the data collected in the trial is explored. Different machine learning algorithms are compared regarding their applicability to this problem and their potential effectiveness.

Keywords: Goal-setting, Sleep, Sleep Revolution, mHealth, Gamification, Personalization, Machine Learning, Goal Recommendations

Designing, Developing, and Implementing a Personalized Gamified Goal-setting Mechanism for a Sleep-tracking Mobile Application

Yannick Zapfe

maí 2023

Útdráttur

Í leikjavæðingu (e. Gamification) er algengt að notendur get sett sér ákveðin markmið. Rannsóknir sýna að notkun ákveðinna og erfiðra markmiða í heilsutengdum snjallsímaforritum (e. mobile health applications) hafi hvetjandi áhrif á notendann. Hinsvegar er lítið um rannsóknum á áhrif mismunandi tegunda markmiða, og þá sérstaklega markmiða sem hafa tímamörk. Í þessum pappír er kynnt innleiðingu tveggja markmiða, áframhaldandi (e. Continuous) markmið og tímabundið (e. Time-bound) markmið fyrir snjallsímaforritið „Sleep Revolution“. Innleiðingin gerir notenda kleyft að setja markmið á alla þá þætti sem notandi skráir niður, svo sem fyrir svefn eða koffíninntöku. Innleiðingin er byggð á niðurstöðum og ráðlegginga fyrri rannsókna. Til að meta áhrif hvorrar tegunda markmiða fyrir snjallsímaappið var framkvæmd fjagra vikna slembivalin og stýrð rannsókn (e. Controlled study). Þar að auki er skoðað raunhæfi þess að hafa persónubundin markmið sem eru búin til út frá þeim gögnum sem hefur verið safnað. Að lokum er borið saman mismunandi vélnámareiknirit (e. Machine learning algorithms), hversu árangursríkir og hæfilegir þeir eru.

Efnisorð: Markmiðasetning, Svefn, Sleep Revolution, mHealth, Gamification, Sérstillingar, Vélanám, Ráðleggingar um markmið

Designing, Developing, and Implementing a Personalized Gamified Goal-setting Mechanism for a Sleep-tracking Mobile Application

Yannick Zapfe

Thesis of 30 ECTS credits submitted to the School of Technology,
Department of Computer Science
at Reykjavík University and
the Department of Information Technology,
Faculty of Science and Engineering
at Åbo Akademi University in partial fulfillment
of the requirements for the degree of
Master of Science (M.Sc.) in Intelligent Software Systems

May 2023

Student:

Y. Zapfe

Yannick Zapfe

Examining Committee:

Anna Sigríður

Anna Sigríður Islind

L. Schmitz

Lisa Schmitz

Erna Sif Arnardóttir

Erna Sif Arnardóttir

Dragos Truscan

29.5.23

date

Yannick Zapfe
Master of Science

I dedicate this thesis to my Mum and Dad who have supported me throughout my studies, no matter how far it took me from home.

Acknowledgements

This research is a part of the Sleep Revolution project, with funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 965417.

I want to thank my family and friends who, close or far, have always supported me in so many ways. I would not be anywhere close to where I am now - not just geographically - and I will be forever grateful for this chapter in my life that is coming to an end now with this thesis. Thank you for all your love and support!

Contents

List of Figures

List of Tables

List of Abbreviations

ANOVA Analysis of Variance

API Application Programming Interface

BAA Behavioral Analytics Algorithm

EU European Union

LSTM Long Short-Term Memory

MAB Multi-Armed Bandit

MIBLP Mixed-Integer Bi-Level Program

MILP Mixed-Integer Linear Programming

ML Machine Learning

PA Physical Activity

RCT Randomized Controlled Trial

RNN Recurrent Neural Network

RU Reykjavik University

SDB sleep-disordered breathing

UCB Upper Confidence Bound

UI user interface

UN United Nations

WHO World Health Organization

Chapter 1

Introduction

In the past decades, mobile technology has been increasingly employed to support health care services, resulting in the emergence of mHealth, or mobile health. The World Health Organization (WHO) and the United Nations (UN) have both acknowledged the potential of mHealth to transform the way health services are delivered across the globe [1]. The Sleep Revolution mobile application is among the many mHealth applications in use today. The Sleep Revolution project is a multidisciplinary cooperation of 39 active partners that is funded by the European Union (EU) [2]. It aims to fundamentally change clinical sleep medicine with the introduction of a new diagnostic and digital management paradigm. The digital Sleep Revolution platform [3] is designed to bridge the gap between patients, healthcare professionals, and researchers. The app is a part of this platform and allows patients to track and monitor their sleep and sleep-related data. The sleep data is collected through a sleep diary in the app that the users are asked to fill out daily in the morning and the evening. Examples of the collected data are sleep times, time spent in bed, caffeine and alcohol intake, screen usage, and exercise [4].

In previous user studies and surveys, users of the Sleep Revolution app have frequently requested the introduction of gamification to the application. Gamification is defined as the "use of game elements in non-game contexts" [5, p. 10]. It aims to improve the user experience and increase user engagement and the overall value creation provided by the gamified application. Like mobile technology, gamification has shown to be a promising addition to health and well-being applications. This is particularly attributed to its ability to foster motivation which, in turn, is the main driver of behavior change, one of the main goals of most mHealth applications. [6] Goal setting is both commonly used as a game design element itself and is facilitated by other elements, such as progress bars, in gamified applications. Goal-setting theory has first been formalized by Locke & Latham with the core finding that goals should be specific and difficult [7], [8]. In addition to their other conceptualizations, the scientific community seems to be reaching a consensus that optimal goals are specific, measurable, attainable, realistic, and time-bound [9]. Goal-setting theory has also been applied to mHealth interventions and, when following several recommendations, has been shown to improve health outcomes [10].

This work introduces a goal-setting mechanism to the Sleep Revolution mobile application. It allows users to set goals regarding several activities they track in the app's sleep diary. Previous research suggests that the effects of different goal types, particularly time-bound goals, have received little attention [11]. Therefore, two types of goals, continuous and time-bound, are developed and evaluated in this work. Continuous goals refer to daily goals. In contrast, time-bound goals have a fixed duration and a target number of days to achieve the goal within this time period. The effects of the goal-setting mechanism on user compliance with health recommendations are investigated and the two goal types are compared in a four-week user study. The study population is divided into three groups: one that does not have access to goals, one that can use continuous goals, and one that uses time-bound goals. The goal-setting mechanism is assessed in this work performing a theoretical evaluation and a quantitative analysis of the preliminary user study results.

Personalization has been found to be an effective means to improve the outcomes of mHealth interventions and to promote behavioral change [12]. It also has the potential to increase the effectiveness of gamified systems [13]. However, gamified mHealth applications are labor-intensive to implement as it is. Therefore, Machine Learning (ML) has been proposed to automate the personalization process [14]. To improve the outcomes of the goal-setting mechanism developed in this work, three different ML algorithms are explored to generate personalized goal recommendations based on the user's history and current situation. They are theoretically evaluated regarding their applicability to the problem. The behavioral analytics algorithm by Mintz et al. [15] is a promising solution especially with smaller amounts of data. As data grows, Recurrent Neural Networks (RNNs) can potentially be applied effectively to the problem too.

In summary, this work aims to accomplish the following:

1. Develop and evaluate a gamified goal-setting mechanism for the Sleep Revolution mobile application following theoretical concepts and recommendations from previous work in the fields of goal-setting theory, gamification, and mHealth.

2. Explore different ML algorithms to generate personalized goal recommendations that can be integrated into the implemented goal-setting feature.

The remainder of this thesis is organized as follows: chapter 2 presents the theoretical foundations of this work, including mHealth, goal-setting theory, gamification, and ML, in more detail. In chapter 3, the methodology followed to implement the goal-setting mechanism, to explore and evaluate the different ML models, and to conduct the user study are presented. The implementation of the goal-setting mechanism and the theoretical application of the ML algorithms to generate personalized goal recommendations are explained in chapter 4. Following that, the results of the user study are presented chapter 5 and discussed in chapter 6. In chapter 7, several potential ethical issues of gamified goal-setting implementations are addressed. chapter 8 contains limitations and suggestions for future work, and the thesis is concluded in chapter 9.

Chapter 2

Theoretical Background and Related Work

In this chapter, the theoretical and technological foundations for this work are presented and they are referenced extensively in chapter 4 as the basis for the implementation. These theories and technologies include mHealth, the Sleep Revolution project and mobile application, goal-setting theory, gamification, personalization, and ML. The following sections also outline the relationships and intersections of the aforementioned fields.

2.1 mHealth

The Sleep Revolution platform (see section 2.2) falls in the category of eHealth (electronic health) products. More specifically, the mobile application that is part of that platform and the primary target of this work belongs to the category of mHealth (mobile health) which is introduced in this section.

The United Nations' Sustainable Development Goals[1] include both "Good Health and Well-Being" and "Reduced Inequalities". They particularly promote a better future for all people through, among other things, improved access to health services using information and communication technology. [16] The growth in smartphone availability and usage and in mobile cellular network coverage has reduced the global digital inequality and given rise to new opportunities for the implementation and integration of mHealth services. [1][17] Both patients and healthcare providers have, ever since, shown an increasing interest in adopting mobile technologies to create, communicate, and consume healthcare information. [17][18]

mHealth is generally considered to be a component of eHealth and is defined as "medical and public health practice supported by mobile devices" [1, p. 6] by the Global Observatory for eHealth[2]. Additionally, in a literature review of mHealth definitions, Hallberg & Salimi gather the following aspects of mobile health:

- Dissemination of health information and education for patients and healthcare providers
- Personal health management based on statistical analyses of the patient's data

[1] https://sdgs.un.org/goals

[2] https://www.who.int/observatories/global-observatory-for-ehealth

- Monitoring of patients outside of hospital conditions
- Healthcare services in a direct, low-cost, and engaging manner
- Support and integration with traditional healthcare services

They also acknowledge that mHealth, like eHealth, does not replace human actors entirely, as both concepts rely on social and material interactions. Therefore, the "native relics" must not be compromised when developing mHealth services. [16]

There is a wide variety of categories and applications within mHealth. Most mobile health services fall into at least one of the following categories that were compiled in [1], [17], and [18]:

- Treatment compliance, i.e. drug adherence and appointment reminders
- Remote dissemination of information and educational material
- Data collection and monitoring
- Diagnostic treatment and support, including digital therapeutics

The disrupting effects of mHealth can potentially transform the delivery, access, and management of health services across the globe to make them available to wider populations. This increased accessibility also enables more personalized and patient-centered healthcare [1]. To promote and improve general health and well-being, mHealth fosters communication and the relationship between patients and healthcare professionals [16].

Despite its potential and benefits, mHealth also comes with several challenges that must be addressed when developing mobile health applications. The one major concern in this field is data security and privacy when storing and transmitting sensitive health data. [1] Further challenges include the lack of standardization in the field, legal regulations, a deficiency in communication between different healthcare institutions, and too little patient involvement in the design and development of mHealth solutions [16], [17].

The Sleep Revolution mobile application and, therefore, this work fall into the category of mHealth, which is transforming the way healthcare services are provided for more than a decade now, making these services more easily and widely accessible and improving communication and treatment personalization.

2.1.1 Notifications

Notifications are a commonly deployed mechanism in both mobile applications in general and mHealth apps. To use this feature effectively and purposefully in this work, previous research on mobile notifications, especially in health-related services, is presented in this section. It comprises the five primary components a notification is made of as well as a list of design recommendations to follow.

Previous studies have shown that - when used appropriately - digital notifications can enhance the effectiveness of health interventions. This potential can be attributed to the fact that notifications emphasize an internal or external goal in the user's natural environment at an appropriate time. They increase the user's attention to and engagement with the respective application and their compliance with the promoted health objectives. Despite their potential, the evidence of the effectiveness of digital triggers has been mixed, and they come with several challenges. Besides privacy and security concerns that are addressed in the previous section (see section 2.1), notifications can hinder habit formation by creating dependencies on technology, and too many notifications can lead to alert fatigue. In the latter case, users are likely to ignore notifications. [19][20]

In [19], Muench & Baumel separate digital triggers, i.e. notifications, into five primary components that can be used not only to deconstruct and analyze existing notifications but also to design new ones:

1. **Who** (sender)
   People attribute human characteristics, including credibility, to computers and programs. Therefore, the user's perception of the sender of the message, that is the notification, influences their reaction.

2. **How** (stimulus type, delivery medium, heterogeneity)
   The stimulus type refers to how the notification is presented to the user, e.g. using words, images, or sounds, and to the complexity of the perception and the desired magnitude of the reaction. The delivery medium is the hardware, e.g. a smartphone, and the software mechanism, e.g. emails or push notifications, used to deliver the notification. It has been shown to have a relatively small effect on the user's engagement following the notification. Trigger heterogeneity means the diversity of stimulus types and delivery mediums, times, and sources.

3. **When** (delivered)
   The timing and perceived disruption of a notification highly affect the impact it has on the user's engagement with the application following the trigger. Ideally, a notification is sent when the recipient has the necessary time and energy for the intended interaction. This depends on the complexity of that interaction and the task the user is currently engaged in.

4. **How much** (frequency)
   The frequency with which notifications are sent has a significant impact on the user's reaction to the triggers. Most importantly, more notifications are not necessarily better and can result in alert fatigue. Instead, the frequency should be determined by the user's perceived importance of the application.

5. **What** (structure, target, narrative)
   The trigger's structure simply means the arrangement of its content. The target of the trigger is the short-term or long-term goal of sending this particular notification. The short-term goal refers to the immediate reaction to the trigger; the long-term goal to the user's sustained engagement with the application. The latter can also be supported by a larger story that is told by the notifications as a whole.

Additionally, several previous works provide a number of design recommendations for notifications:

- Too many notifications, e.g. about actions the user initiated themselves or system events, are not appreciated. [19]–[21]

- The content of notifications should vary to keep users engaged. [20]

- Tailoring, i.e. personalizing, notifications - and applications - to the individual increases the user's engagement. [19], [20]
- Interactive elements increase user engagement. [20]
- Including a sender increases credibility and user engagement. [19]
- Notifications about people and events are perceived as more important. [21]
- Notifications can be assigned priorities, and users can decide to receive notifications that surpass a certain importance only. [20], [21]

Notifications and their design can greatly improve the effectiveness of a health intervention, i.e. an mHealth application, and, therefore, the aforementioned guidelines and recommendations were taken into account when designing and implementing the goal-setting feature in this work (see section 4.1).

2.2 Sleep Revolution

This work is embedded into the Sleep Revolution project which is funded by the European Union's Horizon 2020 Research and Innovation Programme and ultimately aims at a paradigm shift in clinical sleep medicine for sleep-disordered breathing (SDB) towards a new diagnostic and digital management approach. The concept and objectives of the project are presented in [2] and summarized in this section.

For an estimate of almost one billion people that have obstructive sleep apnea [2], the overall goal of the Sleep Revolution project is to decrease costs and, therefore, improve the availability of sleep studies as well as the general quality of life. To achieve this, Arnardottir et al. define four objectives of the project:

1. "transform current diagnostic methods for SDB"
2. "bring advanced sleep diagnostics from the hospital into the patient's home"
3. "promote participatory healthcare with technological solutions"
4. "develop different personalized treatment options for patients with SDB"

To meet the first objective, the current diagnostics process, an overly simplistic index that disregards most information in the data, is replaced with novel ML algorithms. These both achieve a higher degree of accuracy and improve the knowledge about which signals in the data are most relevant and how to analyze them. The second objective refers to moving sleep studies from artificial laboratory environments to the home setting. It also includes extending the studies to multiple nights with diagnostic equipment and replacing the time-consuming process of manual scoring with automated tools. Objective 3 is to implement participatory patient involvement as part of the 4P medicine that is predictive, preventive, personalized, and participatory medicine. This approach aims to bridge the gap between patients, healthcare professionals, and researchers by combining medicine, digital technologies, and consumer-driven healthcare. The co-designed Sleep Revolution platform that also comes with a mobile application represents that bridge and allows the collection of, standardized calculations based on, and access for all stakeholders to relevant data. The fourth and final objective targets the personalization aspect of 4P medicine and addresses the urgent need for more personalized and preventive treatment approaches for patients with SDB.

Overall, the goal of the multidisciplinary Sleep Revolution project is to improve both the availability and the precision of sleep studies and diagnostics by introducing cutting-edge technology and increasing patient involvement in sleep medicine and research.

2.3 Goal-Setting Theory

Goal-setting theory, initially developed by Locke & Latham in [7] and updated in [8], among other related work, is a "theory of motivation that aims to explain the causes of people's performance in work-related tasks" [11, p. 1] and lays the psychological foundation of the goal-setting feature developed in this work. A goal is defined as the objective of an action that usually is to be achieved within a certain time limit. Therefore, a goal is a standard for judging satisfaction. Empirical research has shown that goals, like plans, intentions, and tasks, influence human behavior immensely. In the remainder of this section, the core findings of goal-setting theory as well as goal mechanisms and moderators and further related definitions and concepts are presented.

At the core of the theory, Locke & Latham found that goals lead to optimal performance when they are specific and difficult, i.e. when the objective is clear rather than abstract and requires a considerable but realistic effort. They discovered a linear relationship between goal difficulty and performance unless the difficulty exceeds the individual's ability. Goals then affect performance through the following mechanisms:

- **Direction**
  Attention and effort are directed toward goal-oriented activities.

- **Effort**
  Individuals mobilize effort in proportion to goal difficulty.

- **Persistence**
  The time spent on the task also increases for specific and difficult rather than vague or easy goals.

- **Knowledge or task strategy**
  People draw from their whole repertoire of knowledge and skills to achieve their goals.

- **Self-efficacy**
  The faith in their abilities and the attribution of the achievement as well as positive affect positively influence self-efficacy and, therefore, the variety and difficulty of goals an individual will take on.

In addition to these mediators, three main moderators affect the relationship between goal and performance:

- **Commitment**
  Commitment to the goal increases performance and can usually be traced back to the importance of the goal to the individual and the individual's self-efficacy.

- **Feedback**
  Feedback enhances the goal-performance relationship by enabling people to adjust their strategies and effort to the current progress.

- **Task complexity**
  With increasing task complexity, the goal effects exceedingly depend on the individual's ability to choose and carry out appropriate task strategies.

Further related concepts in Locke's & Latham's goal-setting theory include the differentiation between outcome goals that refer to accomplishing a specific result, performance goals, i.e. doing well on a task by one's own standards, and process or learning goals that have the acquirement of new skills and knowledge as an objective. Stretch goals are difficult or even impossible to reach - as opposed to the idea of setting difficult yet realistic goals. They serve as a guideline and provide direction rather than an actually desired outcome. Goals can also be divided into proximal and distal goals, where the former often facilitates the attainment of the latter, e.g. by breaking larger, distal goals into smaller, proximal goals. Finally, conflicting goals undermine performance if they motivate incompatible actions.

A more recent discovery that seems to be reaching a consensus within goal-setting research is that optimal goals are specific, measurable, attainable, realistic, and time-bound (SMART). [9] These properties can, for the most part, be derived from the aforementioned findings.

2.3.1 Goal-Setting in mHealth

Goal-setting theory has also been applied to mHealth interventions, especially when promoting Physical Activity (PA). Due to the lack of research on goal-setting in sleep-related applications, although this work does not focus on PA, the findings from these applications are summarized in the following section and taken into consideration when designing the goal-setting mechanism in this work (see section 4.1).

In accordance with the previously presented goal components, Baretta et al. [10] identified a total of six components that moderate the effectiveness of goals and that can - in goal-setting interventions to promote PA - serve as meaningful quality measures. They found that applications that took all of the following components into account are generally effective in promoting PA behavior:

- **Specificity**
  Refers to specific goals as explained in the previous section, i.e. well-defined and measurable outcomes.

- **Difficulty**
  Again, as mentioned before, this means setting realistic yet challenging goals.

- **Timing/timeframe**
  A combination of short- and long-term goals is most effective. If an intervention supports only one of these two, it should be short-term goals.

- **Action planning**
  The application should support action planning, i.e. provide or help to create a plan on when, where, and how to act.

- **Goal evaluation**
  Goal progress should be constantly evaluated, and feedback be provided to the user.

- **Goal re-evaluation**
  Goal-setting needs to be implemented as an iterative process that integrates updating and setting new goals as users progress toward and beyond their current goals.

In addition to their function as quality measures, these components can also be used as design guidelines when creating mHealth interventions. Several further recommendations are listed in [10] and [22]. Generally, goals work best when they are important to the individual, and an application is most effective for users that agree with the overall aim of the intervention. Another way to ensure that users value their goals is to allow them to set their goals themselves or participatively with an expert. Rather than setting goals for the individual, an intervention could provide guidelines or recommendations for new or updated goals. Besides neutral feedback, i.e. progress tracking, an application could deliberately provide incentives to the users as they progress towards their goals. [22] Finally, tailoring goals along with other app content to each individual user increases the effectiveness of the intervention. This can be achieved through the aforementioned goal components, especially difficulty and goal re-evaluation. [10]

To conclude this section, goal-setting theory, in general and when applied to the field of mHealth, offers a variety of findings and frameworks that can guide the design process of the goal-setting feature developed in this work.

2.4 Gamification

Both the Sleep Revolution mobile application and the goal-setting feature developed for the app as a part of this thesis work incorporate game elements. Therefore, in this section, the concept of gamification in general and in health and well-being contexts in particular is introduced, and its relation to goal-setting theory (see section 2.3) is elaborated and explained.

In [5], Deterding et al. define gamification as well as related concepts, such as gamefulness. The following paragraphs summarize their findings that have laid the foundation for further research in that area of interest. First of all, they generally define gamification as the "use of game design elements in non-game contexts" [5, p. 10]. In a service-marketing context, it also refers to an extra software service layer of reward systems with typical game elements, such as points and levels, that enhances the core service. The aim of gamification is to improve the user experience and, therefore, to motivate and increase user activity and engagement and to enhance the users' overall value creation. The application domains are manifold and include finance, education, sustainability, and, most importantly for this work, health. The only domain that is explicitly excluded herein is actual games.

User interface design has always been heavily influenced by other design practices, and gamification has arisen from a number of interacting trends and traditions in interaction design, including serious games and pervasive games. Serious games are any games that have been designed to incorporate features that go beyond entertainment, e.g. promoting health or providing education. On the other hand, pervasive games extend or blend into the real world by expanding play in a spatial, temporal, or social way [23]. Another related trend is the *ludification of culture* which denotes the transformation of video games into a cultural medium on par with literature or movies. More closely related to gamification is gamefulness, the experiential and behavioral quality of gaming. Gameful interactions are artifacts that afford that quality, and gameful design refers to designing for gamefulness, predominantly by using game design elements, i.e. gamification. To separate gamification from full-fledged games, which could be serious or pervasive games, it is important to note again that gamified applications merely incorporate elements that are characteristic of games, among other more functional modes.

The authors of [5] summarize their findings, i.e. their definition of gamification, as:

- "the use (rather than the extension) of
- design (rather than game-based technology or other game-related practices)
- elements (rather than full-fledged games)
- characteristic for games (rather than play or playfulness)
- in non-game contexts (regardless of specific usage intentions, contexts, or media of implementation)" [5, p. 13].

To facilitate analysis and guide the design of gameful applications and features, Chou identified eight - and a "hidden" ninth - core drives of gamification in [24]:

1. **Epic Meaning & Calling**
   A person believes that they were chosen to contribute to something greater than their individual benefits.

2. **Development & Accomplishment**
   The internal drive to make progress regarding individual knowledge and skills as well as outcomes

3. **Empowerment of Creativity & Feedback**
   The iterative process of creating or adjusting something according to the feedback about previous progress

4. **Ownership & Possession**
   A person feels responsible for and in control of their actions and outcomes.

5. **Social Influence & Relatedness**
   Social feedback and comparison influence and motivate people.

6. **Scarcity & Impatience**
   People tend to have a desire for rare and exclusive things.

7. **Unpredictability**
   Not knowing what happens next creates excitement.

8. **Loss & Avoidance**
   Avoiding a negative outcome, in contrast to simply working towards a positive outcome, can also facilitate and increase motivation.

9. **Sensation**
   Deals with physical sensations and pleasure through touch, hearing, sight, smell, and taste

The design of a gameful application or feature, like the goal-setting mechanism in this work, should account for these core drives to be effective. As the author provides a very broad framework, further details are omitted here but can be found in [24].

2.4.1 Gamification for Health and Well-Being

Gamification has, as previously mentioned, also been applied to health and well-being applications. In [6], Johnson et al. provide a literature review of health gamification that is summarized in the following section.

The drivers of health gamification and the sources of its potential overlap with those of mHealth (see section 2.1). They include its broad accessibility through mobile technology and applicability to many major health risks, the appeal to wide audiences that play games, and its cost-benefit efficiency as it can be used to either enhance existing applications or to design new ones. The most important promise of gamification for health and well-being is its ability to motivate through the engagement and enjoyment that are the intended effects of game elements. Individual behaviors affect all modern lifestyle health risks. Therefore, individual behavior change can significantly improve a person's overall health and well-being. The main driver of behavior change is motivation and, in the context of health, especially intrinsic motivation that can be facilitated through well-designed games and gamification.

The authors situate health gamification "at the intersection of persuasive technology, serious games, and personal informatics." Its sensor-based and data-driven health interfaces usually revolve around tracking individual behaviors enhanced by a goal-setting and progress feedback mechanism. Then, standard behavioral reinforcement techniques and reward systems are employed to further motivate and direct behavior change. In most cases, the delivery platform is a mobile application or website. The effects of gamified health interventions are reported to be mostly positive (59%) or mixed (41%), however, usually with moderate or lower quality of evidence. None of the reviewed interventions found direct negative impacts, and the benefits of health gamification even extend to users who had no pre-existing motivations to change the targeted behavior. The most commonly employed game design elements are rewards, i.e. points and achievements or badges, followed by leaderboards, avatars, and social interaction that have all been associated with improvements in different sub-domains. Physical health and especially physical activity have shown the highest popularity and success rates while most of the aforementioned game elements were not as beneficial for applications promoting mental health and mindfulness. However, Johnson et al. also identified several challenges that are common to gamified health interventions. Those challenges include the high cost and complexity of the design and development of these interventions, people's varied access to technology, and the poor usability of gamified applications in parts even resulting from the poor use of game elements. In general, the low quality of evidence in gamification research undermines the expressiveness of the results.

2.4.2 Gamification Principles Through Goal-Setting Theory

After delineating the concept of gamification in general and in health and well-being contexts in the previous sections, the following paragraphs provide an outline of Tondello et al.'s theory of gamification principles through goal-setting theory [11]. Those principles connect gamification with the other essential theoretical foundation of this work. In their paper, the authors conducted a literature review to develop a framework that relates important concepts from goal-setting theory to ideas and design elements in gamification research. They also investigated gamified implementations of goal-setting recommendations and the effects on gameful design.

Goals themselves are commonly employed game elements and are used to motivate and analyze specific elements in gamification design. Goal-setting theory often serves as a theoretical base for gamification systems. Most gameful design methods include goal-setting mechanisms by either providing the users with clear goals or allowing them to set their own goals. It is important to note that the mere existence of goals usually does not suffice to change the user's behavior. They need to regularly track their performance and receive feedback on their progress to commit to the set goals.

Tondello et al. identified a variety of game design elements that are used for goal-setting in gamification, with badges and leaderboards being the most popular ones, followed by rules, goals, challenges, and progress bars. While badges and levels are promising mechanisms for setting goals, progress bars can be used to facilitate feedback. Leaderboards had a similar effect as difficult or stretch goals and, like other social interactions, foster goal commitment. As mentioned above, goals can be given to or set by the user. In the latter case, the user typically sets explicit goals, whereas goals given by the system can be more implicitly presented, e.g. in the form of badges and achievements. The following game design elements were also mentioned in [11] and can be used to implement goal-setting in gamification: conflict, points, rewards, boss battles, certificates, collections, exploratory tasks, learning, quests, unlockable or rare content, unlockable access to advances features.

The framework resulting from their research maps all important notions from goal-setting theory (see section 2.3) to gamification concepts and gameful design elements. To begin with, setting clear goals, e.g. in the form of badges, and encouraging users to pursue them with the help of rewards directs their attention and efforts. Gamification can create a safe space to fail and try again to achieve a goal and, therefore, promote persistence. By adjusting the difficulty to the user's abilities, gamification can keep users engaged and help them improve existing or learn new skills. Here, it is also important to design the system around the already existing challenges rather than creating artificial ones. To reduce the difficulty of potentially too hard goals, they can be broken down into smaller challenges, e.g. tasks or quests, or the system could provide free cues or power-ups. Feedback mechanisms, such as progress bars, levels, rewards, and surrounding narratives, can help users feel part of something larger than themselves and experience positive emotions, thus fostering self-efficacy and goal commitment as well as progress feedback. Other types of goals can also be implemented using gamification. Outcome goals can come in the form of challenges, quests, and exploratory tasks, whereas performance goals are often presented to the user as badges, leaderboards, points, and levels. Process or learning goals can be implemented as onboardings or tutorials. However, learning often occurs more implicitly in gamification. A few ideas for stretch goals could be completing 100% of the side quests or an increased difficulty mode. The concept of SMART goals is, for the most part, often naturally implemented in gamification systems. Goals are typically clear and often even provide the specific steps to take. By implementation it is always possible to measure the progress towards the goal, and both goal-setting theory and gamification in practice promote difficult yet realistic goals. However, it is often hard to evaluate the user's skill level and, therefore, to adjust the difficulty correctly. In addition, gamification systems usually do not provide a time limit for their goals. An extensive table with goal-setting principles and their corresponding gameful design guidelines and elements can be found in the original paper [11].

This mapping of goal-setting principles to concrete game design elements is intended to guide future implementations of goal-setting mechanisms in gamification systems and, therefore, was considered for the implementation in this work.

2.5 Personalization

The second overarching goal of this work, in addition to developing and evaluating a goal-setting mechanism for the Sleep Revolution app, is to generate personalized goal recommendations. Therefore, this section presents an introduction to the field of personalization and its intersection with gamification.

Generally, personalization can be defined as the opposite of a uniform policy for the entire user population [25]. It is the process of tailoring information and content to each user based on their preferences and situation [26], [27]. In a business context, it refers to the concept of one-to-one marketing. Here, a business tailors its marketing to different (groups of) individual customers by understanding their needs, attitudes, and preferences. The business aims to increase its revenue and retention while satisfying the customers' objective to receive useful and interesting information in a timely manner [26]. Contextualization extends personalization in the sense that it, in addition, takes the context into account to provide personalized content in the right situation and at the right time [27].

The need for personalization in the digital world emerged, among other factors, from the information overload as a result of public access to the internet. Personalization, however, has been a business practice long before the digital age, e.g., when a local store owner knows their clients and informs them about new products every time they visit the store [26]. Individual-level personalization at scale, however, was only enabled by advances in computing power and data storage as well as developments in the field of ML that is presented in the next section [25].

The granularity of personalization strategies varies greatly. It ranges from very broad user or customer segments to individual treatment of each customer, usually based on a complex ML model. The ability to personalize highly depends on the differentiation between individuals. The latter is often challenging as it requires large amounts of data [25]. The current context adds to the complexity of the problem as the users' preferences can depend on it [27].

Personalization has been applied in many different domains, including health care and mHealth. It has been found to be an effective means to improve the outcomes of mHealth interventions and to promote behavioral change [12]. Again, the context of the user must be considered when choosing the right intervention [27].

2.5.1 Personalized Gamification

Personalization is an upcoming trend in gamification research as well. As discussed in section 2.4, studies of gamified applications have yielded mixed results. Personal and contextual differences have been pointed to when explaining some of these results [14]. In previous studies, it has also been shown that a gamified system evokes varying interpretations and uses in different users and that each user has a personalized sense of fun and motivation [14]. Therefore, personalization has the potential to increase the effectiveness of gamified systems compared to uniform approaches [13]. Personalized gamification differs from adaptive gamification in that it not only adapts to different situations but also takes the characteristics of different users in these situations into account [14]. One approach is to map a user's personality traits to different design elements [13]. However, gamification requires a lot of effort as it is, and personalizing the gameful system to each user only adds to the complexity of the task. Therefore, ML has been proposed to automate part of the personalization process. Knutas et al. have also extended an existing design process for gamified systems to incorporate the automation of developing a personalization strategy using ML algorithms [14].

To summarize this section, personalization has the potential to increase the effectiveness of gamified and mHealth applications. Due to the complexity and work intensity of the task, automated approaches using ML seem to be the most promising strategy.

2.6 Machine Learning

ML has been one of the most active research fields within Computer Science. It became popular due to advances in processor speed and memory size. As mentioned in the previous section, it can also be used to personalize (gamified) interactive applications and offers different solutions for personalized goal recommendations that will be presented in this section. To begin with, the following paragraphs draw from [28] to provide a broad overview of the ML field.

ML can be defined as the use of computers to simulate human learning, i.e. learning from experience and reasoning rather than with algorithms. It allows computers to acquire knowledge from the real world and use this knowledge to improve their performance on certain tasks [28]. Michalski et al. define machine learning formally as learning from experience E regarding a class of tasks T and according to a performance measure P, if a computer program improves its performance measured by P on tasks T improves with experience E [29]. ML algorithms have been applied to a multitude of problems in many application domains and can be classified as follows [30]:

- **Supervised learning**
  In this approach, the learning algorithm is provided with labeled training data, i.e. problem instances and the correct answers. The algorithm learns based on that data with the goal to generalize the knowledge to previously unseen samples in the real world.

- **Unsupervised learning**
  In unsupervised learning, the training data is not labeled. The algorithm is presented with the data and has to learn, i.e. usually to find hidden patterns in the data, on its own.

- **Semi-supervised learning**
  Semi-supervised learning algorithms deal with a small portion of labeled training data and, therefore, with missing information.

- **Reinforcement learning**
  Reinforcement learning algorithms follow an approach similar to teaching dogs. Learning is guided by external feedback, i.e. either rewards or penalties, for the decisions the algorithm makes.

2.6.1 Neural Networks

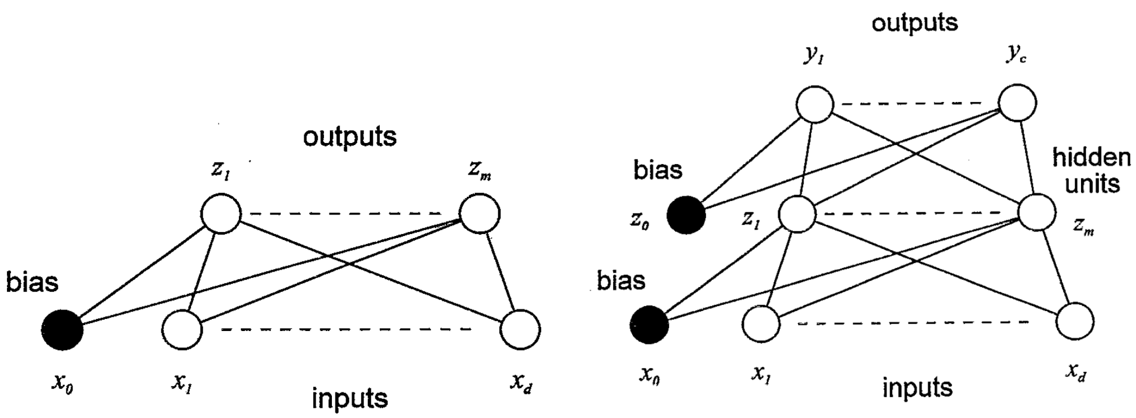
Neural networks have received a lot of research interest within ML and, until today, are among the most popular models in the field. This section provides a brief overview of the two most common types of neural networks that are both based on the multi-layer perceptron: feed-forward and recurrent neural networks. The following paragraphs introduce feed-forward neural networks based on the summary provided by Bishop in [31].

To begin with, artificial neural networks' architectures and their building blocks, artificial neurons, mimic biological neural networks. An artificial neuron is a simple mathematical model of a biological neuron. Essentially, it is a non-linear function that maps a set of input variables into one output variable. First, all input variables x_i with $i \in \{1, ..., d\}$ are multiplied by one parameter each, a so-called weight w_i. Then, all the weighted inputs and an offset parameter w_0, the bias, are added up. The output z of the unit is then computed using a non-linear activation function g, originally a threshold function, nowadays usually sigmoidal functions. This results in the following function that is modeled by the neuron:

$$z = g(\sum_{i=1}^{d}(w_i x_i) + w_0)$$

By linking together these artificial neurons, i.e. feeding the output of one neuron into another neuron as an input, neural networks can be constructed. The simplest architecture of such a network is known as a single-layer neural network and is depicted in Figure 2.1a. It is a set of m such units with common inputs, providing m outputs. The computational capabilities of this kind of network are very limited. A more powerful model is obtained when adding successive layers of processing units, i.e. artificial neurons, and when using non-linear, e.g. sigmoidal, activation functions to allow techniques of differential calculus. This enables the neural network to model any continuous mapping over a finite range of input variables to arbitrary accuracy. It is called a multi-layer neural network and is illustrated in Figure 2.1b. There, two layers of processing units are depicted: one hidden layer and the output layer. The inputs are fed into the units of the hidden layer. Then, the outputs of these units are passed on to the units in the output layer to calculate the outputs of the model.

Learning in neural networks refers to the adaptation of the weight values w_i for $i \in \{0, ..., d\}$ according to a training algorithm in response to the training data provided to the model. One of these training algorithms, namely error backpropagation, led to the breakthrough of neural networks in the 1980s. As indicated by the name, a network computes the output for a given input, compares it to the correct answer according to a loss function, and propagates the error back through the network to update the weights in each layer.



(a) Single-layer neural network (b) Multi-layer neural network

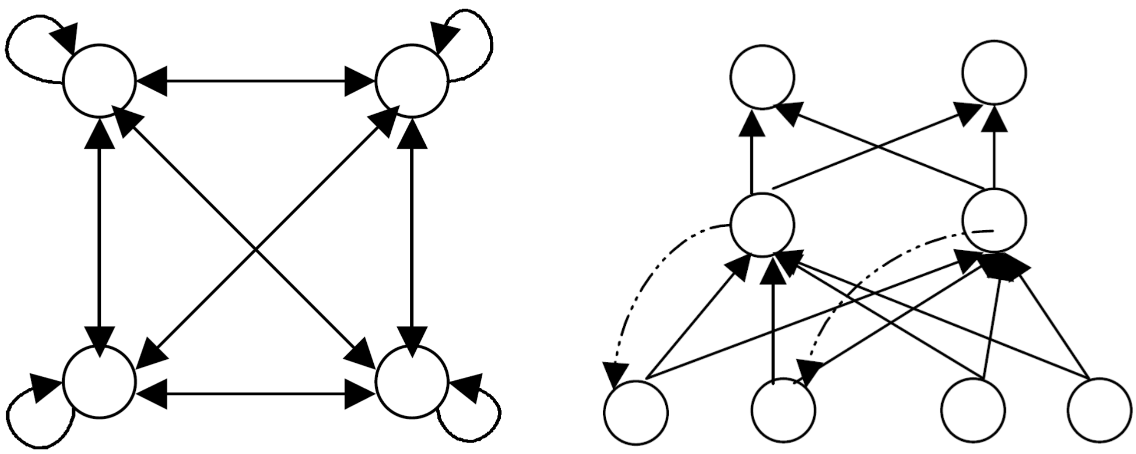
Figure 2.1: Neural network architectures (original images from [31])

The neural networks in this section are called feed-forward neural networks as they feed the outputs of one layer forward into the processing units of the next.

Recurrent Neural Networks

An extension to the previously presented model of neural networks is Recurrent Neural Networks (RNNs). They add feedback connections to the model and are specifically designed to learn sequential or time-varying patterns. Medsker & Jain have provided an overview of these networks in [32]. The following paragraphs summarize their account.

The class of Recurrent Neural Networks (RNNs) includes feed-forward neural networks and ranges from partially connected to fully interconnected nets. Feeding the output of a unit back into the unit is also possible. Figure 2.2a depicts a fully interconnected network, meaning that the output of each unit is fed into every unit as an input. In Figure 2.2b, the network is organized in layers with some outputs of the middle layer being fed back into the first layer. Both graphics omit the input layer.



(a) Fully connected RNN (b) Partially connected RNN

Figure 2.2: RNN architectures (original images from [32])

If an output of a unit is fed back into another unit that has been used to process the current already, it is only used when the next input is provided to the network. The backpropagation training algorithm is adjusted accordingly to take feedback connections into account, resulting in the backpropagation-through-time approach.

To conclude this section, neural networks form the basis for many ML models today and, in turn, are based on a simplified model of biological neurons. Especially recurrent networks are a promising model to generate goal recommendations based on the sequential diary data in the Sleep Revolution app in this work (see section 4.2).

2.6.2 Recommendation Systems

Recommendation systems, or recommender systems, often make use of ML methods to generate personalized item or service recommendations for users. Some of the most well-known applications to employ such systems are Netflix, Spotify, and YouTube. The following paragraphs summarize Portugal et al.'s overview [28] of the field and its usage of ML techniques.

Recommendation systems can be divided into three main categories:

- **Collaborative filtering**
  This approach considers the user data to find other users that share the same preferences to recommend items or services that these users have used. Collaborative filtering methods either rely on a similarity network of users or a predictive model trained on the stored preferences.

- **Content-based filtering**
  Content-based filtering methods, in contrast, base their recommendations on the item data and recommend items and services similar to the ones the user has used in the past. Again, either networks or classifiers are used to produce the recommendations.

- **Hybrid filtering**
  This approach combines collaborative and content-based filtering methods resulting in recommendations based on both the user and the item data.

One major challenge in recommendation systems is that they usually rely on large amounts of data to generate appropriate recommendations.

Beutel et al. augment this overview of recommendation systems with the consideration of context in these systems. While context is often modeled as an additional feature in feed-forward recommender systems, they suggest more complex models to fully appreciate the complexity that the context might add. According to them, Recurrent Neural Networks (RNNs) (see section 2.6.1) are at the cutting edge of recommendation systems because they model temporal patterns and by definition provide a second-order neural model.

2.6.3 Bandit Algorithms

Bandit algorithms, or Multi-Armed Bandits (MABs), are commonly employed in recommendation systems. Therefore, they are presented in this section and evaluated for the use in this work in section 6.2.

The term "multi-armed bandits" originates from the scenario in which a user plays multiple slot machines, also known as one-armed bandits, at the same time and sequentially selects slot machines that yield different payoffs to maximize their total reward [27], [33]. In an algorithmic sense, MABs provide a powerful approach to making decisions over time under uncertainty. In particular, bandit algorithms have different actions, or arms, to choose from in a number of rounds. In each round, the algorithm chooses and performs an action and collects a reward for it. The reward is drawn from some fixed distribution which is not known to the algorithm and, therefore, must be learned [33].

mHealth is among the various application domains of bandit algorithms. They provide a natural framework to learn which intervention, in the case of this work which goal, is the most sensible for a given user in a given situation [27]. The fundamental problem to solve in bandit problems is the exploitation-exploration trade-off [27], [33]. Exploitation refers to choosing actions that have yielded high rewards in the past, whereas exploration means trying previously less-used or unused actions to find, i.e. explore, new and possibly better choices and strategies. Exploration typically yields a lower short-term payoff. However, focusing on exploitation only can potentially omit better actions or strategies as the rewards for each action are initially unknown to the algorithm [27].

A bandit approach that has been shown to be effective in previous studies is the Upper Confidence Bound (UCB) technique which can be summarized as "optimism in the face of uncertainty". The UCB algorithm always chooses the action with the highest UCB. That is either an action that has been selected before and yielded a high reward, or an action with a very wide confidence interval and a high upper limit indicating great uncertainty about the reward it will yield. In the latter case, the action may yield a low reward but, in that case, also reduces the uncertainty about it. Therefore, the likelihood that it will be chosen again in a similar situation in the future decreases. The algorithm starts with an exploration phase to compute preliminary estimates of the reward an action may yield. Based on that, the UCB is calculated for each action and updated with each decision the algorithm makes according to the reward it returns [27].

Instead of developing a static strategy that would select the same action regardless of the current situation, bandit algorithms should take the context into account. The context includes the properties and preferences of the current user, allowing for personalized actions, as well as other attributes, such as the day of the week [33]. In this extended scenario, the algorithm aims to find an optimal mapping of contexts to actions. Due to the number of contexts in real-world scenarios, mapping each context to an optimal action is not a feasible approach. Therefore, contexts, e.g., users, are often grouped to result in manageable but sufficiently granular context categories [27].

In conclusion, bandit algorithms can be used to find good solutions from a range of possible options for different users in different situations. Therefore, they are considered for the task of generating goal recommendations in this work.

2.6.4 Behavioral Analytics Algorithm

Another promising method for goal recommendations in this work is the Behavioral Analytics Algorithm (BAA) introduced by Mintz et al. in [15]. This algorithm has also been applied in a similar setting to generate personalized daily step goals [34]. This section first introduces the BAA in general and, then, the results of its application in a step-counting mobile application.

The problem the algorithm tries to solve consists of a large pool of myopic, i.e. shortsighted, agents each with a set of utility function parameters. The agents are independent and do not interact with each other strategically. Each agent makes a decision at time t based on its motivational and system state to maximize its utility function. A single coordinator makes noisy observations of the agents' decisions and system states and assigns behavioral, or financial, incentives from a possibly limited pool. The coordinator's problem is to select the incentives for each agent to minimize its loss function and to improve the agents' motivational and system states.

The algorithm comprises three steps:

1. Develop a behavioral model to describe the decision-making process of an agent, i.e. its utility function. This model combines motivational and system, i.e., health, states. It should also fit the optimization method used by the algorithm (see below) mathematically.

2. Use past data to estimate the behavioral model parameters for each agent and use the model to predict future agent decisions.

3. Use the estimated model parameters to optimize the incentives provided by the coordinator.

The behavioral model and, therefore, the incentives are continuously optimized by repeating steps 2 and 3 as new observations are made.

The algorithm is built around mathematical formulations of the behavioral model that are solvable in a straightforward fashion using commercial solvers. I.e., Mixed-Integer Linear Programming (MILP) is used to estimate the parameters of the agents' utility functions. The incentive design problem can also be solved with MILP and consists of Mixed-Integer Bi-Level Program (MIBLP) in the case of a limited pool of incentives for the whole population of agents.

The authors have shown the algorithm to be asymptotically optimal w.r.t. the loss function that models the coordinator's objective. In a simulation of a weight-loss program, BAA maintained the efficacy of the program while reducing its costs by up to 60% compared to adaptive heuristics. It also improves the interpretability of the results compared to, e.g., reinforcement learning approaches.

In [34], the algorithm is evaluated in a Randomized Controlled Trial (RCT) in which it is used to generate personalized adaptive daily step goals. The setting resembles this work in that it considers sequential health data to generate personalized goals. Therefore, the setup and results of the study are presented in the following paragraphs.

In the 10-week RCT, the intervention group received personalized daily step goals generated by the BAA, while the control group received constant step goals of 10,000 steps per day. The main outcome measure was the relative change in daily steps between the run-in period of seven days and the ten weeks of the trial period. Secondary outcome measures included the step goal attainment, i.e. the fraction of step goals achieved by each participant. The difference in daily steps between the groups after ten weeks, which is the primary outcome measure, was 960 steps. In addition, participants in the intervention group achieved 30-40% of their goals, 15% more than the control group members. The study shows the short-term efficacy of the intervention and confirms the applicability and benefits of the BAA. The BAA combines the advantages of highly personalized, usually self-set, goals and assigned goals that do not require constant input by the participants.

In the study, BAA operated on the complete history of each participant individually. That included both the past steps and goals. The algorithm used the run-in data of all participants to construct an adequate behavioral model. It was, then, applied every week to compute the upcoming seven step goals with the goal of maximizing the participants' actual number of steps.

BAA constructs and uses a behavioral model specific to the problem at hand to optimize incentives, e.g. goals, that are provided to agents. The setting and success of the field study presented above are the reasons why BAA is further assessed for the goal-recommendation mechanism in this work in section 4.2.

Chapter 3

Methodology

This chapter contains the methodology followed in this work to implement and evaluate the goal-setting mechanism, including the user study design to effectively measure goals' effects on user compliance. In addition, the process of finding and theoretically evaluating different approaches to generate personalized goal recommendations is outlined.

3.1 Implementation

The implementation of the goal-setting mechanism is split in two: first, the feature itself is implemented in this work, and, then, the generation of personalized goal recommendations is explored in this work and potentially implemented in the future. The following paragraphs describe the implementation process and which concepts were used to guide it. The implementation itself is presented in chapter 4.

Goal-Setting Mechanism

The implementation of the goal-setting mechanism extends the current functionality of the Sleep Revolution mobile application. Several methodological frameworks and recommendations from the previous chapter were used to guide the implementation including, in particular, the following:

- Goals should be specific and difficult and are mediated by several mechanisms that must be considered [7].
- Goals should be specific, measurable, attainable, realistic, and time-bound [9].
- Implementations of goal-setting mechanisms in physical activity apps, and, therefore, potentially in mHealth interventions in general, should include six components: specificity, difficulty, timing/timeframe, action planning, goal evaluation, and goal re-evaluation [10].
- The feature should satisfy the definition of gamification as "the use of design elements characteristic for games in non-game contexts" [5, p. 10].
- The gamified feature should implement the Chou's core drives of gamification [24].
- The feature should include game elements that foster goal-setting [11].

- The five components of digital triggers [19] serve as a template for the notification design.
- The notification design should consider the several recommendations by [19]–[21] listed in subsection 2.1.1.

The conformance of the implemented mechanism with these concepts and recommendations is assessed qualitatively in section 5.1.

Machine Learning Models

To integrate personalized goal recommendations into the goal-setting mechanism, a machine learning approach was chosen due to the promising potential outlined in section 2.5 and section 2.6. The following paragraphs describe the process of selecting suitable machine learning algorithms from the multitude of available solutions.

To begin with, as presented in section 2.6, existing machine learning algorithms were explored to find algorithms with properties particularly well-suited for the problem at hand and models that have been used in similar settings. This resulted in the presented list of theoretically or practically suitable approaches out of which three similarly promising but sufficiently different ones were chosen for further assessment. These three algorithms and the necessary adjustments to the goal recommendation problem in this work were described to get a grasp of how they can be used in a future implementation and what are the main benefits and challenges of each approach. These advantages and disadvantages were then discussed to identify the most promising techniques in different situations. The implementation of the models and a quantitative evaluation is left for future work due to the lack of training data from the user study.

3.2 User Study

To quantitatively assess the implementation of the goal-setting mechanism and its effects on user behavior, a user study was designed and conducted to compare the different effects of continuous and time-bound goals. This section describes the setup of the study in detail; the results are presented in section 5.2.

Study Design

The user study was a four-week RCT with three groups:

1. The first intervention group had access to the continuous goal-setting feature in the Sleep Revolution app.
2. The second intervention group had access to the time-bound goal-setting feature in the app.
3. The control did not have access to the goal-setting feature.

All participants had to fill a consent form and an initial screening questionnaire before being granted access to the application. The study started in April 2023, i.e. the first participants started their four-week trial on April 28, 2023. As, even for the first participants, the study exceeds the deadline of this thesis, the number of participants and the end date of the trial for the last participant can not be determined yet. The data collection described in this thesis was granted ethical approval by the National Bioethics Committee of Iceland (under the reference number: VSN-23-056) and each participant of the study has signed an informed consent.

Participant Recruitment

In total, 6 participants were recruited as of May 11, 2023. Recruitment started in April 2023 and was ongoing in May 2023. The user study was promoted to members and followers of Reykjavik University (RU) word-of-mouth by the researchers involved in this work, via physical and electronic posters at the university facilities, and on the university's Instagram page. The inclusion criteria for the study were:

- An Icelandic social security number
- A mobile device with Android or iOS
- Intention to track and possibly improve their sleeping habits
- Willingness to install and use the Sleep Revolution app
- Consent to data processing according to the consent form

Being a member of RU or being connected to the institution in some other way was not an inclusion criterion. However, the study was promoted to people that have some kind of connection to RU and, therefore, it was very likely that the majority of participants would be members of RU. There were no exclusion criteria other than not fulfilling the inclusion criteria. To sign up for the study, participants had to fill the online consent form that checked for the inclusion criteria.

Study Procedure

After filling the consent form, each participant received an email with a link to the online screening questionnaire and installation and setup instructions for the Sleep Revolution app. Once they had installed the app, created a user account, and joined this particular study, the participants could use the app independently for four weeks. The email contained instructions to fill the app's sleep diary every day and, if possible, i.e. if a participant is part of one of the intervention groups, use the goal-setting feature. The implementation of the goal-setting mechanism in the app and how to use it is detailed in subsection 4.1.3.

Randomization

The participants were assigned one of the three groups when signing up for the study in the Sleep Revolution app. As this sign-up happens individually, each participant was assigned the group with the least number of members at the time of the sign-up. This approach, in contrast to assigning the participants to one of the groups with equal probability, ensures that the group sizes do not differ by more than one member, especially when considering the expected small total number of participants.

Control

The control group only had access to the base functionality of the Sleep Revolution app not including the goal-setting mechanism. It does include the following features:

- Sleep diary
- Statistics based on the diary
- Morning and evening notifications to remind the user of filling the diary
- Brain games, i.e. a set of games to assess the user's fatigue, reaction, memory, and perception.

In this work, the diary entries of the users are of interest as a baseline for sleep behavior without goals. The results of the brain games are not considered here.

Intervention

The intervention groups both had access to the base functionality described in the previous paragraph. In addition, they had an extra icon in the bottom navigation to use the goal-setting mechanism, either with continuous or time-bound goals (see subsection 4.1.3). The feature allows users to view their current goals and progress, to set new ones, and to update or remove existing goals. They also received notifications reminding them about their goals (see subsection 4.1.4).

Outcome Measures

The primary outcome measures concern the different goal schemas that are presented in detail in subsection 4.1.2 individually. For the caffeine and alcohol goal schemas, the daily mean consumption each week and over the whole course of the study was considered; for the screen time goal schema the daily mean screen time was used in the same way. For the bed and wake-up times we considered the standard deviation and the deviation from the goal, if one is set, per week. That was because there is no universally optimal bed or wake-up time but it is beneficial to keep a regular bed time routine. In all cases, zero is the best possible value. The primary outcome measure is the mean, or standard deviation, over the whole course of the study. The secondary outcome measure was the difference between the values of the different weeks, especially the first and the last week to see whether the different interventions influence the participants differently over time. A possible third outcome measure that would only be applied to the two intervention groups - the primary and secondary measures are applied to all groups - measures the goal attainment as the percentage of achieved goals. However, due to the different types of goals in this work, no meaningful measure could be found that would not naturally favor one of the two types. E.g., using the percentage of days on which a goal was achieved would benefit continuous goals as in time-bound goals participants can decide to aim at less than 100% of the days; measuring goal attainment on a per goal basis would neglect the different time frames of continuous and time-bound goals as well as within time-bound goals. Therefore, the goals are only compared according to their impact on the participants' sleep habits rather than the actual goal-setting behavior. The primary and secondary outcome measures relied on the participants' diary entries. All data was stored in a central database on the server when it was entered in the mobile application. It was accessible for the researchers to process it in the context of this work.

Statistical Analysis

To analyze the possibly different effects of the interventions on the three groups, as explained in the previous section, the mean and standard deviation values for different sequential data , i.e. the timelines of certain diary entry types of each participant, were used. To check potential differences for statistical significance, the Analysis of Variance (ANOVA) and the Kruskal-Wallis test, two statistical hypothesis tests, were applied. As in each case, only one factor, i.e. a certain diary entry type, was analyzed for multiple treatments, i.e. the two interventions and the control group setup, these tests were applicable. In section 5.2, the additional assumptions of the more powerful ANOVA test, especially for small datasets, were checked and, if the check is successful, the test was applied. Otherwise, the Kruskal-Wallis test was used.

Chapter 4

Implementation

This chapter, firstly, describes the implementation of the goal-setting mechanism in the Sleep Revolution app and, secondly, theorizes about the integration of personalized goal recommendations using different ML algorithms.

4.1 The Goal-Setting Mechanism

The development of the goal-setting mechanism ranges from designing the continuous and time-bound goals' structures based on previous work to implementing the feature in the Sleep Revolution mobile application. I.e., the Django server of the Sleep Revolution platform and its GraphQL Application Programming Interface (API) were extended by models for goal schemas (see subsection 4.1.2) and goals and API endpoints for users to create, read, update, and delete goals. A new navigation tab with several screens was added to the React Native mobile application of the Sleep Revolution platform and they are described in detail in subsection 4.1.3.

4.1.1 Goal Types

To begin with, this subsection covers what a goal comprises in the developed goal-setting mechanism and how the two types of goals, namely continuous and time-bound, differ. Previous work suggests that there is a lack of studies that compare different types of goals and their effects on performance. In addition, time-bound goals, among other concepts in goal-setting theory, have been found to be underrepresented in research as well. [11] Those are the reasons why, in this work, two different types of goals were developed rather than just comparing the effect of goals in general with the users' performance in the absence of goals. The data structures used in this implementation, as well as the goals' presentation to the user, are described in subsection 4.1.3. The following paragraphs present the concept of - continuous and time-bound - goals in this work.

The properties that define a goal and that continuous and time-bound goals have in common are:

- The **user** the goal belongs to
- The chosen goal category, here called ***goal schema***, that is the real-life activity the goal targets and that is tracked by the user (see subsection 4.1.2)

- The **limit** for this activity set by the user according to the goal schema's measure, i.e., a number, a time duration, or a time of day
- The **start date**, i.e., the date when the goal was set by the user

Continuous and time-bound goals both have further type-specific properties that they do not share with the respective other goal type. Continuous goals, conceptually, do not add any properties to the ones listed above. They refer to goals that are set once by the user and that can be completed every day individually. To clarify the idea, the following example is used: a user sets the goal to drink at most one (the limit) caffeinated drink (the goal schema) per day starting from today (the start date). The user can achieve this goal or fail on a daily basis. The app tracks on which days the user has achieved the goal. The goal applies infinitely unless the user updates or removes it. A time-bound goal requires two additional properties: a **duration** between one and four full weeks, i.e., an end date, and a **target** number of days to achieve the goal within the given time period. In the example, the user could aim to achieve the goal on ten days (the target) in the next two weeks (the duration).

These concepts of continuous and time-bound goals were developed to comply with the findings and recommendations from goal-setting theory [7]–[9] presented in section 2.3. First of all, the goals described above are specific and difficult. They provide a clear limit that can easily be compared to the corresponding diary value for the day to determine whether the goal was achieved or not. To continuously adjust a goal's difficulty to the user's abilities, notifications (see subsection 4.1.4) remind the user to set lower or higher limits when a goal is too frequently achieved or failed respectively. The three moderators that influence the goal-performance relationship are implemented in the following ways:

- **Commitment**
  Goal commitment is usually facilitated by goal importance and self-efficacy. A description paragraph for each goal schema explains how the corresponding activity affects sleep and sleep quality and, therefore, raises awareness of the real-life relevance of the goal. Self-efficacy is increased by directly linking the goals to the sleep diary in the app. The activities the user performs and tracks determine whether a goal is achieved or failed on that day.
- **Feedback**
  Progress feedback is provided at several places and in various forms, e.g., check marks, progress bars, and marked calendars (see subsection 4.1.3).
- **Task complexity**
  Goal difficulty, as mentioned above, is encouraged to be adjusted through notifications once goals become too difficult or easy.

An essential difference between continuous and time-bound goals in this work is that continuous goals correspond to performance goals in goal-setting theory and time-bound to outcome goals. Continuous goals can be achieved or failed on a daily basis but do not provide a clear overall outcome to be accomplished. It is up to the user and their standard of what doing well means, whether they are successful in pursuing the goal or not. In contrast, the accomplishment of time-bound goals can be objectively determined once the end date is reached by comparing the target and the actual number of days the goal was achieved. For continuous goals, an ongoing and possibly infinite streak serves as a stretch goal that is unlikely to be achieved unless the goal is too easy. Both goal types provide proximal, i.e., daily, goals to break down distal goals, which are the improvement in the corresponding activity, the quality of sleep, and the overall quality of life. Additionally, time-bound goals provide intermediate goals for a time span of one to four weeks. Following the above reasoning, the goals in this work are specific in that they provide a clear limit, measurable as they can be computationally compared to their corresponding diary entries for each day, attainable, i.e., regarding daily activities, realistic because the notifications guide the continuous adjustment of their difficulty, and time-bound in the sense that they are meant to be achieved every day by the end of the day in the case of continuous goals and literally time-bound for time-bound goals.

4.1.2 Goal Schemas

In this section, the aforementioned concept of goal schemas is described in more detail and the goal schema instances available in this initial version of the goal-setting mechanism are listed.

Firstly, a goal schema consists of a **name** that refers to the associated real-life activity, a corresponding **diary entry** type, the unit of the goal limit, called the **measure**, that must comply with the diary entry type's value, and a **description** of the activity's impact on one's sleep. The available measures are **number**, i.e. an integer that is greater than or equal to zero, **time**, which is a duration in hours and minutes, and **datetime**, which refers to a time of day. Matching the measure with the diary entry type's value ensures the seamless computation of whether a goal is achieved. The goal schemas that are available in the current version of the goal-setting mechanism are listed in Table 4.1. The English descriptions for each goal can be found in section A.1.

Table 4.1: Initially available goal schemas

Name	Diary entry	Measure
Alcohol	alcohol.count	number
Bedtime	morning.wentToBedAt	datetime
Caffeine	caffeine.count	number
Screen Time	evening.totalScreenTime	time
Wake-Up	morning.gotUpAt	datetime

The initially available goal schemas were chosen to cover all three measures and were derived from existing diary entry types, especially tracked activities. This keeps the implementation simple and, at the same time, allows for easy extension to further existing and new diary entry types, such as datetime limits for caffeinated beverages, alcohol consumption, and screen time, and time limits for napping and snoozing.

4.1.3 Goals

This section provides a more detailed explanation of the underlying data structures for goals and the presentation to the user in the app. First, the database model for goals is presented, followed by a detailed description and rationale for the user interface screens and elements.

Table 4.2 list the properties for the database models of continuous and time-bound goals in compliance with the concept of a goal presented in subsection 4.1.1. Each goal is assigned exactly one of the three limits according to the measure of its goal schema. The start date is assigned automatically when creating the goal to be today whereas the end date is required to be chosen by the user for time-bound goals and used differently for continuous goals. For continuous goals, the end date is set when the goal is either removed or updated. In the latter case, the end date is set to yesterday and a new goal is created. This new goal links to the current version via the "Previous goal" field to keep track of changes made to the goal over time. When a time-bound goal is removed, the "Canceled at" field is set to today.

Table 4.2: Database models for continuous and time-bound goals with data types in parenthesis and required fields in bold font

Continuous goals	**Time-bound goals**
Goal schema (foreign key)	
User (foreign key)	
Number limit (integer)	
Time limit (integer)	
Datetime limit (time)	
Start date (date)	
End date (date)	**End date (date)**
	Target days (integer)
	Canceled at (date)
Previous goal (goal)	

Based on these data models, the following features can be computationally derived:

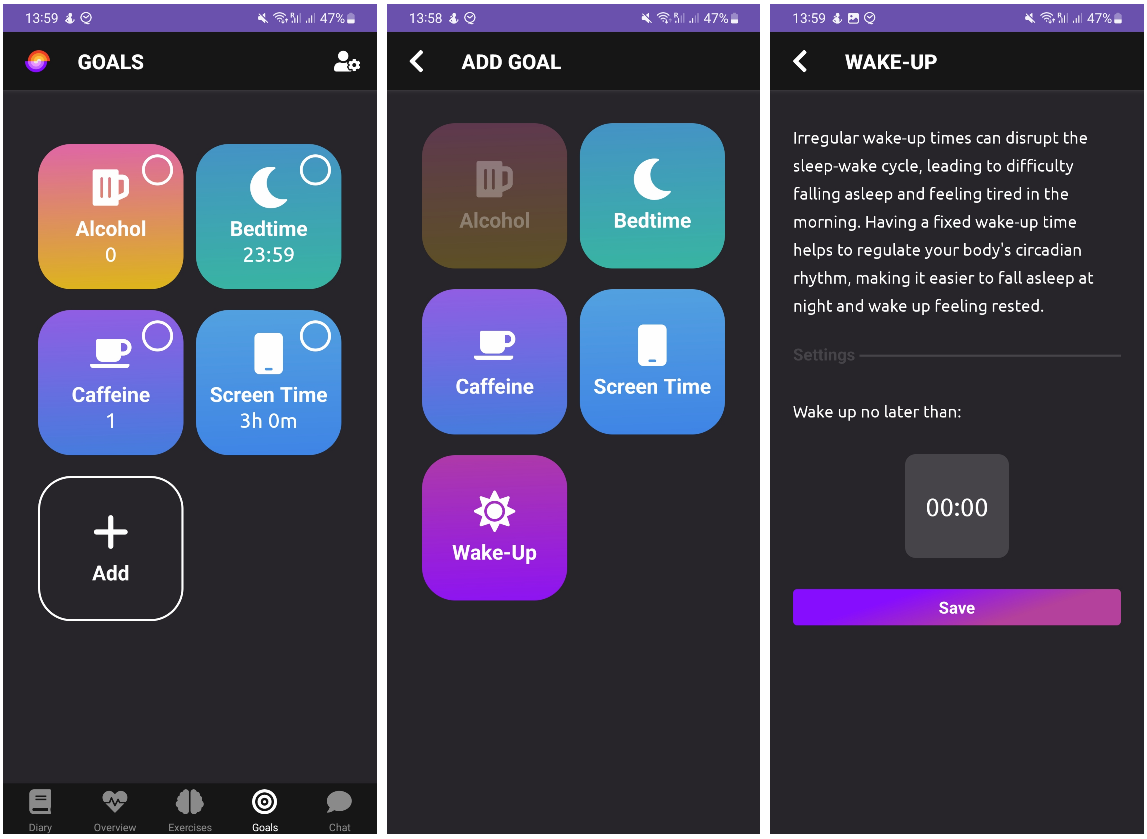
- Whether the goal is **active**, i.e. if the end date is either not set or not reached yet and the "Canceled at" field is not set
- A list of days on which the goal was **achieved** using, in addition, the diary entries of the user
- The length of the current goal **streak**, which is the number of consecutive days the goal was achieved that includes today
- Weekly and monthly **progress**, i.e. the percentage of days when the goal was achieved in the last seven and 30 days respectively

The goal screens that were designed for Sleep Revolution mobile application, like the data models above, are mostly similar for continuous and time-bound goals. They only differ where the distinct features of the two goal types make it necessary. This way, comparability is ensured and the influence of the different screens on the user is kept at a minimum.

Continuous Goals

To recap, a continuous goal can be achieved on a daily basis and continues indefinitely unless the user updates or removes it. Goal streaks and weekly and monthly completion rates act as implicit distal goals. The following paragraphs motivate and describe in detail the screens and elements for continuous goals in the Sleep Revolution app (see Figure 4.1 and Figure 4.2).

Goal Overview The goal overview screen in Figure 4.1a opens when the user chooses the "Goals" tab in the bottom navigation. It displays all currently set goals and an "Add" button to set new goals in a tile arrangement. Each goal tile contains the name of the goal schema and a corresponding icon, the set limit for the goal, and a checkbox that indicates whether the goal has been achieved for today. The latter implements goal feedback and evaluation for all goals at a glance. These two properties have been found to be critical for the effectiveness of goal-setting (see section 2.3). Depending on the current progress on the active goals, there can be an encouraging comment at the top of the screen highlighting an active goal streak or the overall completion percentage.



(a) Continuous goals overview (b) Add goal (c) Create continuous goal

Figure 4.1: Screenshots of the goal-setting feature in the Sleep Revolution mobile app for continuous goals

Add Goal Adding a new goal is split into two screens: the "Add goal" screen in Figure 4.1b and another screen specific to the chosen goal schema (see Figure 4.1c). The former lists all goal schemas as tiles with its name and the appropriate icon. Only the available goal schemas, i.e. those without a currently set goal, are clickable; the others are grayed out. The second screen consists of a description of the goal schema chosen in the previous screen, an input field to set the goal limit that corresponds to the goal schema's measure, and a "Save" button to finish the process. The description contains information on why setting this goal and improving in that area would be beneficial for the user. It is meant to foster goal importance (see section 2.3) by explaining the goal's value and impact. The goal is also linked to the overall goal of improving one's health and well-being through the description and it is ascribed a higher meaning that exceeds the daily achievement of this goal. This has also been shown to further motivate the user to pursue the goal. Letting the users set their goals themselves empowers their creativity and establishes a feeling of responsibility and ownership of their goals and, therefore, their actions. This covers another two core drives of gamification (see section 2.4).

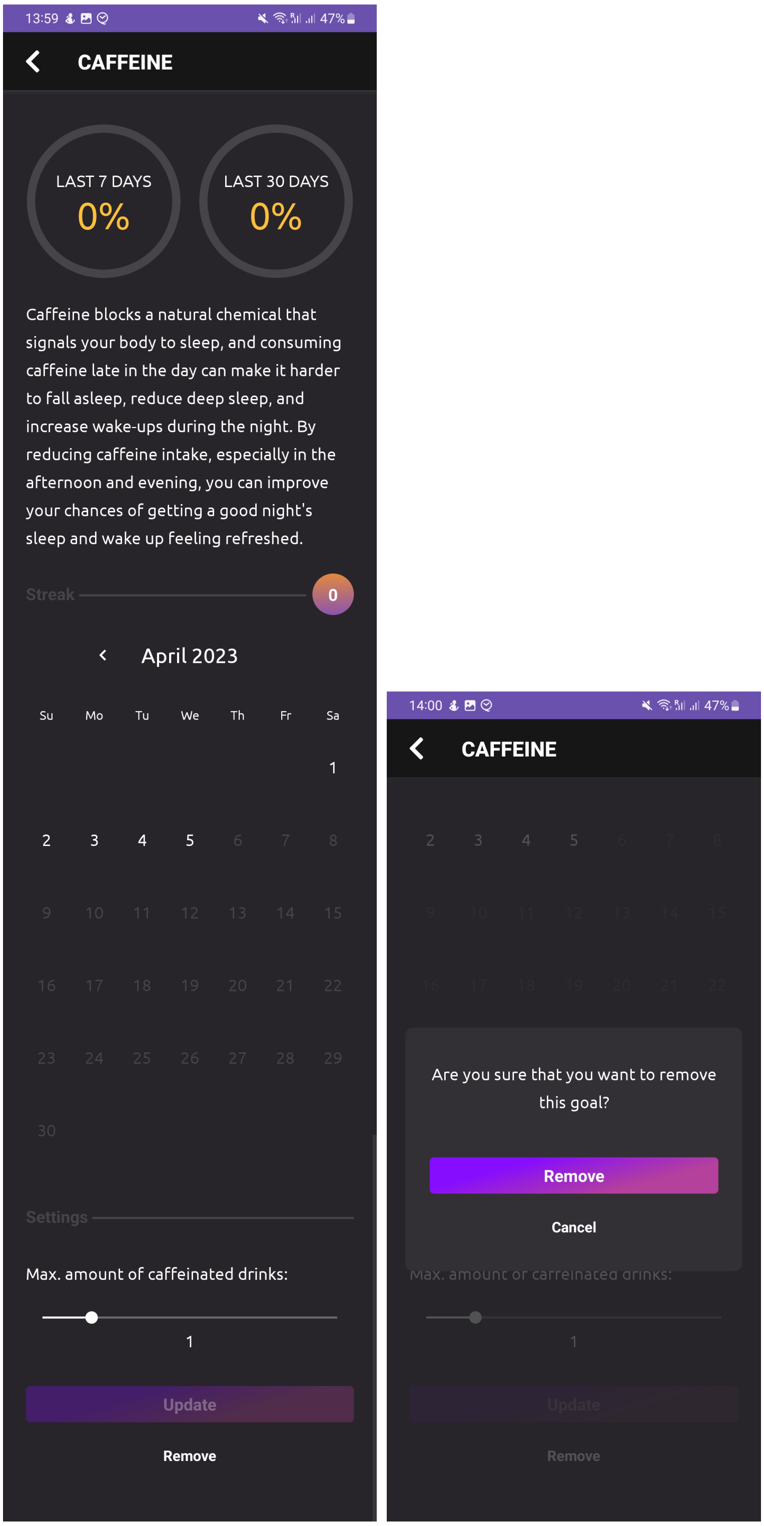
View Goal When a goal is added, a new tile appears in the goal overview that, on tap, navigates the user to the detail view of the goal (see Figure 4.2a). At the top of that screen, two circular progress bars illustrate the completion percentage of the goal for the last seven and 30 days respectively. Below, the goal description is displayed again (see previous paragraph), followed by the current goal streak, i.e. the number of days the goal was completed in a row. The streak is zero when the goal was not completed today, one when it was completed today but not yesterday, two when it was completed today and yesterday but not the day before that, etc. A calendar shows the user in detail on which days they achieved the goal. These days are highlighted in yellow, future days are grayed out. The user can use arrows to navigate between different months. All elements take previous limit values of this goal into account in case it was updated in the past. The progress bars, the goal streak, and the calendar all provide progress feedback and goal evaluation following the recommendations from subsection 2.3.1. An active streak acts as an implicit goal; perfect completion rates and a possibly infinite streak as stretch goals (see section 2.3). Keeping the streak, i.e. avoiding to break it, implement the eighth core drive of gamification (see section 2.4). The other elements in this screen are explained in the next two paragraphs.

Update Goal Below the previously described elements, the user can use the same limit field as in the "Create continuous goal" screen (see section 4.1.3) and an "Update"-button to update the goal. As explained before, updating a goal sets an end date for the current version and creates a new one with the new limit and a link to the previous version. Therefore, when the goal is updated, the old limit is applied to past days and the other elements in the detail view do not change.

Remove Goal A secondary button at the bottom of the detail view of the goal can be used to remove the goal. When that button is tapped, a modal dialogue opens to either confirm or cancel the removal of the goal, as this is a destructive action that cannot be undone by the user. In case of removal, the user does not have access to the goal anymore. Hence, removing a goal and setting a new one for the same goal schema is not the same as updating a goal. However, it is still stored in the database and available for analysis.

Time-Bound Goals

The concept and, therefore, the screens for time-bound goals deviate from the previously presented screens for continuous goals in various ways that are highlighted in the following paragraphs. As described in subsection 4.1.1, time-bound goals have a

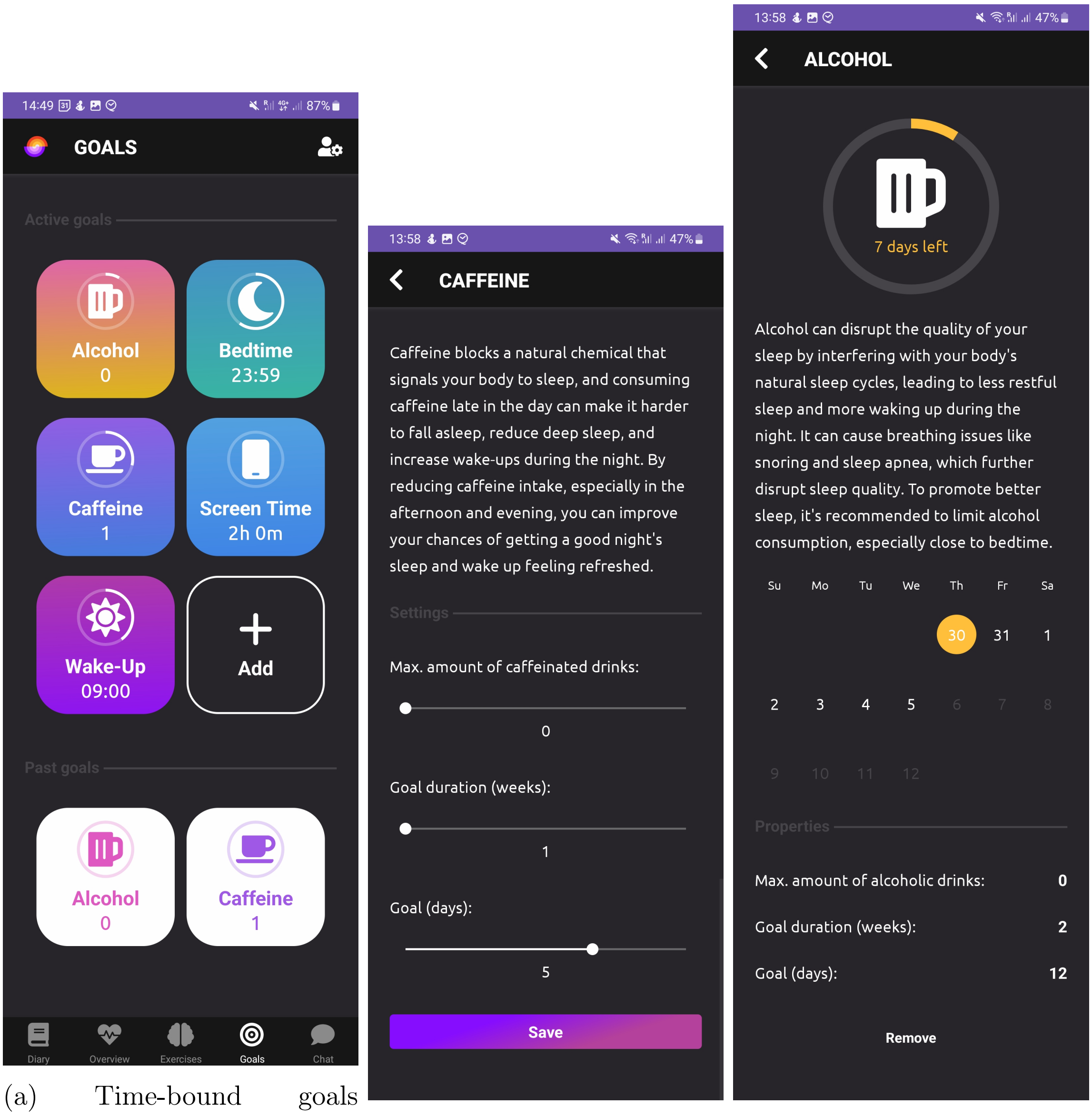


(a) View continuous goal (b) Remove continuous goal

Figure 4.2: Screenshots of the goal-setting feature in the Sleep Revolution mobile app for continuous goals

set duration and end date and a set target number of days for the goal to be achieved within the given time period.

Goal Overview The goal overview for time-bound goals is split into two sections: active and past goals. The former contains all active goals, i.e., those that have an end date in the future and that were not canceled by the user, and an "Add"-button to create new goals. The latter comprises all goals with an end date in the past and canceled goals. Each goal tile consists of the name of the goal schema and the corresponding icon, a circular progress bar around the icon, and the goal limit. The progress bar indicates the ratio of the number of days the goal was completed and the target set by the user. It replaces the check mark used for continuous goals and facilitates progress feedback. The past goal tiles are color-inverted to guide the focus of the user to the more colorful active goals. Like the continuous goal overview, this screen can contain an encouraging message about the current progress of the goals at the top to provide further feedback and positive reinforcement.



(a) Time-bound goals overview (b) Create time-bound goal (c) View time-bound goal

Figure 4.3: Screenshots of the goal-setting feature in the Sleep Revolution mobile app for time-bound goals

Add Goal Adding a time-bound goal follows the same process as creating a continuous goal (see section 4.1.3) except for two additional fields in the second screen that is depicted in Figure 4.3b. The first extra field is used to set the duration of the goal to one to four weeks; the second one is to set the target number of days to achieve the goal within the set duration. As reasoned before, these values implement the time-boundness of goals (see section 2.3) or, in other words, a time frame (see subsection 2.3.1) to the goal. The first screen after pressing the "Add"-button in the goal overview is exactly the same as the previously presented screen for continuous goals (see Figure 4.1b).

View Goal The detail view of time-bound goals deviates most significantly from continuous goals and can be found in Figure 4.3c. At the top, a circular progress bar illustrates the progress towards the goal, i.e., the ratio between the number of days the goal has been achieved to the target number, like in the goal overview (see Figure 4.3a). The goal description is the same and serves the same purpose as for continuous goals, that is to foster goal importance. The calendar that follows the description only includes the days between the start and end date of the goal with the days on which the goal was achieved highlighted in yellow and future days grayed out. Below the calendar, the goal settings are displayed but, unlike for continuous goals, not editable. A time-bound goal cannot be updated, it can only be replaced by a new goal.

Remove Goal In the user interface, removing a time-bound goal works the same way as removing a continuous goal (see section 4.1.3). In the backend, instead of setting the end date to today, like for continuous goals, the "Canceled at"-field is set to today. As a consequence, the goal will be displayed among the past goals in the goal overview.

4.1.4 Notifications

The notifications that are sent through the mobile application are another integral part of the implementation. subsection 2.1.1 explains the importance of notifications for user engagement, especially in mHealth apps. It also contains different frameworks and recommendations that guided the development of the notifications in this work. This process and its outcomes are described in this section.

Firstly, to design the notifications for the goal-setting mechanism, the questions about the five components of digital triggers by Muench & Baumel [19] were answered. While some of these components were the same for all resulting notification types, others had to be developed individually for each type. The three notification types are compliance reminders, update reminders, and daily highlights. They are explained in more detail later in this section. The following are the components that all three types have in common:

- **Sender**
  The notifications are sent to the users through the mobile application. Like the previously existing notifications in the app, they do not specify a real or artificial sender other than the app itself. They are, however, kept in plain language and address the user directly to create a sense of responsibility.

- **Stimulus type and delivery medium**
  The messages are delivered using the mobile device's native push notifications with the corresponding stimuli set by the user, such as vibration and sound.

- **Target**
  From the developer's perspective, the goal of the notifications is to assist and motivate users to set quality goals and, especially, to re-evaluate the set goals regularly. This also supports the users' goal of improving their sleep quality.

All notification types are embedded into an overarching narrative that is the continuous re-evaluation and adaptation of the goals. Depending on the current progress, the user receives feedback about their goals and, if necessary, recommendations for goal updates. This leaves three components of digital triggers to be developed for each notification type individually: when to deliver the notification, the frequency of sending it, and the structure. In addition to these components, sending conditions and concrete messages were developed for the three different notification types. They are outlined in the following subsections.

Compliance Reminders

A compliance reminder is sent when a user does not achieve a continuous goal for multiple consecutive days or when they are running the risk of failing a time-bound goal. Compliance reminders are sent in the morning, if set, together with the morning notifications, so the user can focus on achieving the respective goal on that day. This also motivates and supports action planning for the day as suggested in subsection 2.3.1. The reminders are sent as a batch and, therefore, at most once a day. For each goal that is neglected by the user, reminders are sent every other day until the user achieves it again or updates it. A compliance reminder comes as a push notification with a title and a description in plain text. Depending on the goal type, the text of the notification is the following:

- **Remember your [goal] goal**
  You last achieved this goal [days] days ago. Today is a good day to work on it again!

- **Remember your [goal] goal**
  To complete this goal, you only have [days] more days to achieve it [completions] more times.

goal refers to the name of the goal schema, *days* to the number of days that have passed since last achieving the goal for continuous goals and to the number of days until the end date for time-bound goals, and *completions* to the difference between the target number and the actual number of days the goal has been achieved so far.

Update Reminders

Update reminders are sent when a user either has no goals set at the moment, when a continuous goal has been achieved on more than five or less than four days in the past week and when a time-bound goal is either successfully completed or impossible to complete. These reminders facilitate goal (re-)evaluation and help to keep the goals realistic yet challenging (see section 2.3). They are sent in the evening, together with the evening notifications, if these are set. This way, the user can take some time to reflect on their goals and update them for the following day. Like compliance reminders, update reminders are sent as a batch of push notifications once per day, if necessary. They are repeated every day until the goal is adjusted or the sending condition does not hold anymore. The following messages are sent as update reminders in the corresponding scenarios:

- **Time to set new goals**
  You currently have no goals set. Set yourself challenging yet realistic goals!

- **Time to update your [goal] goal**
  You are doing great on this one! A good time to set a more ambitious goal.

- **Time to update your [goal] goal**
  This goal seems to be too difficult at the moment. You should set a more realistic goal for now.

- **Time to set a new [goal] goal**
  Congrats! You completed your [goal] goal! Time to set a more ambitious one.

- **Time to set a new [goal] goal**
  This goal is out of reach now. You should set a new, more realistic goal.

goal, again, refers to the name of the goal schema of the affected goal.

Daily highlights

Daily highlights, unlike compliance and update reminders, are not delivered as push notifications but appear at the top of the goal overview screen. They appear when a user's performance exceeds a certain threshold according to their own goals, i.e., when, for continuous goals, the overall completion rate is higher than 65% for the past week or one of the goals was achieved on more than five consecutive days, or, for time-bound goals more than two goals were completed in the last week. The message, if there is any, is updated when the diary is filled for the day. Positive reinforcement was not only requested by users of the Sleep Revolution app but also facilitates progress feedback and goal evaluation and fosters motivation among users (see section 2.3). The concrete messages used in the different scenarios are:

- Great job! You're on a [streak] day streak with your [goal] goal.

- Way to go! You completed [completed] % of your goals this week.

- Way to go! You completed [completed] goals this week.

streak refers to the length of the goal streak in days, *goal*, once again, to the name of the goal schema, and *completed* to the weekly completion rate of continuous goals or the number of completed time-bound goals respectively.

4.2 Machine Learning Models for Personalized Goal Recommendations

In this section, three machine-learning approaches from section 2.6, namely RNNs, bandit algorithms, and the BAA, are presented as potential solutions to generating goal recommendations. First, the problem is described in more detail before analyzing the applicability of the different approaches to the problem. Therefore, the necessary adjustments to the collected data and the ML algorithm are presented. The three approaches are evaluated and compared qualitatively in section 6.2.

The goal of this second part of the implementation is to generate personalized goal recommendations, both new goals and updates to the currently set ones, to further support the users in setting their goals and to minimize individual effort. The actual implementation of this mechanism - due to the lack of user data to train the ML models and the limited scope of this thesis - is left for future work. However, the following is a description of one way to present the goal recommendations to the users. In addition to the "Add"-button in the goal overview, a recommended goal can be presented as a transparent tile that resembles the tiles for set goals and, when tapped, opens a pre-filled goal creation form (see subsection 4.1.3). The sliders in the goal creation and update forms could display a static second dot representing the recommended value. The problem to solve can be specified by the inputs and the desired outputs of the mechanism to implement. The idea is to use available data of a user, i.e., diary entries that consist of the user, the date, the entry title, and the value for that day, and all goals the user has set in the past, to produce new goals or updates to recommend to the user. The model can either generate one goal at a time or multiple goals for a certain future time period at once; it can also generate the full goal or the goal settings based on a given goal schema. The selected outcome in the final implementation does, not least, depend on the chosen ML algorithm. The three aforementioned algorithms that are adjusted to the problem and discussed in this work are described in the following subsections.

Recurrent Neural Networks

RNNs (see section 2.6.1) are specifically designed to learn from time series data. They range from simple, minimal models to increasingly complex extensions such as Long Short-Term Memories (LSTMs) [35]. Like other neural networks, they require large amounts of labeled training data to learn from. The input data for the model would be a sequence of diary entry values and goals set by a user. There are two possible approaches to applying RNNs to the problem of generating personalized goal recommendations:

1. The model predicts the future diary entry values and, then, uses heuristics to generate challenging yet realistic goals, e.g., by setting the goal slightly lower than the predicted value, assuming less is better for the diary entry at hand. In this scenario, the different diary entries, and their respective goal schemas, are considered separately by multiple RNNs.

2. The model produces the goal recommendation directly based on the input data. Here, the time series of diary entries and goals can be either regarded separately according to the diary entry or in a joint fashion. In addition, this approach

requires goals as labels for the training data and, as not every goal set by a user in the past should be recommended, a quality measure for goals. That way, the RNN only learns "good" goals to recommend to the user.

The decision of whether to separate the different pairs of diary entries and goal schemas reoccurs for all three approaches in this section and is part of the discussion in section 6.2. In general, this model is trained on all available data from users of the goal-setting mechanism to learn about as many situations as possible. That way, in theory, it can recommend quality goals appropriate to the user's current situation.

Bandit Algorithms

Bandit algorithms, or MABs, are a class of reinforcement learning algorithms commonly employed in recommendation systems (see subsection 2.6.3). Implementations of MABs center around the exploitation-exploration trade-off. In the context of this work, the available actions of an MAB refer to the possible goal recommendations and the external reward can be the successful completion of the goal as well as the improvement according to the metric related to that goal, i.e. the respective diary entry. E.g., when the MAB recommends a goal for caffeine intake, the trend of caffeine consumption by the user after setting the goal indicates whether the goal impacted the behavior positively. An extension to the MAB is the contextual bandit that maps different contexts or situations to optimal actions, i.e. goal recommendations. This is critical to the problem in this work as the goal recommendations are supposed to be personalized according to the current situation of the user. The mapping is often achieved by training multiple MABs for different situations. However, this poses the main challenge in adopting a bandit algorithm: modeling the context. It should be specific enough to enable the algorithm to produce appropriate recommendations but not too specific to avoid overfitting and a too large amount of models. A relatively simple way to model the context in this work would be to consider a moving average of each diary entry over, e.g. seven days, and divide the value range into segments. For caffeine intake, these segments could be less than one, one to three, and more than three caffeinated drinks per day. Considering the decision between separating and combining the different diary entries, i.e., whether to train one MAB for all entries or one per entry, the way context is integrated here calls for a separation of diary entries to avoid combinatorial complexity. This approach, like RNNs, requires a sufficiently large amount of data either provided by a large user population or collected over a long time to allow for exploration.

Behavioral Analytics Algorithm

The BAA developed by Mintz et al. in [15] (see subsection 2.6.4) can be used to generate and optimize incentives in three steps: first, a behavioral model is developed based on agent behavior within a system; then, the model parameters are determined for each agent and future agent decisions are predicted; the incentives are optimized to maximize the agents' utility functions. This algorithm has been used in [34] to generate daily step goals and can be applied to the problem in this work too. The main effort to apply BAA to the problem at hand lies in developing the behavioral model. In [15], one such model is presented for a weight-loss program. However, it is not clearly explained how it was derived and whether it is possible to automate that process using past agent data. A request for clarification as well as the source code used in [34] has not been answered by the time this thesis was finished. To model the users in the Sleep Revolution system, the diary entries, e.g. a seven-day moving average, can be used to represent the system state of the agents. The diary entries on a certain day constitute an agent's decision. The behavioral incentives provided by the BAA refer to the goal recommendations. The most difficult part is to create a model of the temporal dynamics of the agents' system and motivational states. However, manually developing such a model would substitute the need for a large amount of training data, as the incentives can be optimized mathematically as long as the model satisfies a number of assumptions listed in the original paper [15]. To keep the model's complexity manageable, it is advisable to model the different diary entries, and respective goal schemas, separately. The more goal schemas are included, however, the higher might be the necessity to account for mutual influence and, therefore, a joint model.

To summarize the previous paragraphs on the applicability of the different ML techniques, all three have characteristics that make them promising approaches for the problem in this work. All of them allow for both per-goal-schema and global models, although some are more suitable for each model than others. An apparent trade-off is between the required amount of data and manual effort. The benefits and challenges of each approach are discussed in more detail in section 6.2.

Chapter 5

Results

This chapter presents the results of the theoretical and quantitative evaluation of the goal-setting mechanism implemented in this work. In the theoretical evaluation, quality measures and recommendations from previous literature that were listed in section 3.1 are applied to the implementation. The quantitative evaluation refers to the results of the user study described in section 3.2. These results are discussed in section 6.1.

5.1 Theoretical Evaluation

As described in subsection 4.1.3, the goal-setting feature for the Sleep Revolution app was implemented following recommendations from goal-setting and gamification literature. This subsection structures the rationales for each implementation decision given before to evaluate the implementation against these guidelines.

According to Latham & Locke [7], effective goals are specific and difficult. Both continuous and time-bound goals in this work provide the user with a clear limit that applies to exactly one of the diary entries in the application. The implementation ensures difficult, i.e. challenging yet realistic, goals as each goal aims at a behavior change towards healthier habits but can be achieved on a daily basis. In the case that a goal appears to be too difficult or easy, i.e. when the user does not achieve it or achieves it too easily or often, updating the goal is encouraged through notifications. Given that the user follows these recommendations, the goals should always be challenging, yet realistic.

The authors of [7] also identified three mechanisms that mediate the effectiveness of goals: commitment, feedback, and task complexity. The latter, here, is covered by keeping goals difficult as explained above. Goal commitment is fostered by goal importance and self-efficacy. As the general purpose of the Sleep Revolution app for the user is to improve their quality of sleep and the goals promote healthier sleep-related habits, goal importance is ensured. To highlight the connection of the behaviors related to the different goals and the overall aim to improve the user's health, descriptions are provided when creating and viewing a goal. These explain how each goal is important for the user. Self-efficacy, the second source of goal commitment, is facilitated by directly linking the goals to the user's sleep diary. The effect of the user's sleep-related habits on the goals is immediate and clear giving the user full responsibility for their goals. The second mediator of goal effectiveness, feedback, is established by a variety of user interface (UI) elements. The goal overview screen contains checkboxes for continuous goals and progress bars for time-bound goals to provide the user an overview of their progress at a glance. Additional progress bars are displayed in the detail view of each goal as well as a calendar where the days on which the goal was achieved are marked. For continuous goals, the detail view also contains the current streak for the goal. Notifications provide another kind of feedback and remind the user to stick to their goals and to update them regularly. This variety of feedback mechanisms and elements ensures that the user is aware of their goals and progress.

SMART goals are specific, measurable, attainable, realistic, and time-bound [9]. The goals in this work are specific and realistic as explained above. They are measurable too in the sense that each goal links to a diary entry and whether the goal was achieved is computed automatically by the application. Each diary is measurable in the sense that the participant can keep track of their daily actions. The goals are attainable as they all regard everyday activities that lie within the power of the user to change. They are time-bound in the sense that the user can achieve them on a daily basis. However, continuous goals, as indicated by the name, possibly continue indefinitely whereas time-bound goals have a clear start and end date. This is due to the fact that, in this thesis, the effect of exactly that property is investigated.

Baretta et al. identified six components in physical activity apps that facilitate goal-setting: specificity, difficulty, timing/time frame, action planning, goal evaluation, and goal re-evaluation [10]. The first three are covered in the previous paragraphs; goal evaluation refers to feedback (see above). Action planning is integrated into the goal-setting mechanism in the sense that each goal schema targets a specific activity. So, in order to improve their overall quality of sleep, the user is provided with a number of behaviors to change. However, no concrete actions or strategies are provided to, e.g., reduce the caffeine intake in a day. Goal re-evaluation is encouraged through notifications. These recommend to regularly update goals according to the user's progress.

The above recommendations all concern goal-setting. The following paragraphs regard guidelines for gamified applications. Deterding et al. define gamification as "the use of design elements characteristic for games in non-game contexts" [5, p. 10]. The goal-setting mechanism satisfies this definition, as it makes use of game design elements, such as goals, progress bars, and streaks, in a non-game, i.e. a health, context. It can not be considered a full-fledged game as it neither uses game technology - instead, standard web and mobile app frameworks are employed - nor is entertainment its primary aim. In [11], goals and progress bars are recommended game elements to facilitate goal-setting in gamified applications. These two are primarily employed to afford gamification in this work. However, other commonly used elements, like badges and leaderboards, are not used to keep the users' focus on their real-world goals rather than artificial ones. The potential ethical issues of manipulating users into adopting healthier habits through extensive gamification rather than promoting them explicitly is also discussed in chapter 7.

Several papers [19]–[21] provide a list of recommendations for notifications in mHealth applications (see subsection 2.1.1). As update and compliance reminders are sent at most twice per day and only if necessary, never by default, the number of notifications in this implementation is reasonable. The notifications are tailored to the user's current situation. However, the texts for each situation are simple and do not vary. Each notification consists of a title and a description and does not include interactive elements other than that tapping them opens the application. They do not include a sender except the app itself. Important notifications have been found to be about people or events. The notifications in this work concern the receiver and they are triggered by the user's diary entries. The notifications cannot be disabled unless the user disables all notifications for the Sleep Revolution app. That is because they are crucial to ensure effective, i.e., challenging yet realistic goals.

5.2 Quantitative evaluation

In this section, the preliminary results of the user study described in section 3.2 are presented. The statistical analysis of these results (see section 5.2) demonstrates the evaluation of the complete study data that is left for future work. The code used to process the data can be reused once the user study terminates after four weeks.

Start of the User Study

As presented in section 3.2, the user study was planned as a 4-week RCT, but the recruitment and the user study started later than originally planned in the end of April 2023. As of May 11, 2023, two weeks of results from a total of six active participants were collected. For the statistical analysis of the collected data, a minimum of 30 participants per group, i.e. 90 participants in total, would have been needed to ensure sufficient statistical significance. There were several options to evaluate the available preliminary data given these restrictions:

1. Conduct a qualitative evaluation of the user study using interviews to get insights into how the goal-setting mechanism is used and perceived by the users. Nevertheless, the users only used the app for two weeks by the time the interviews would have been scheduled, and it usually takes several days to get used to how the app works. In addition, there would have been only one week left to conduct and evaluate the interviews. Therefore, it was decided to leave a qualitative evaluation of the user study for future work.

2. Demonstrate the planned statistical analysis of the results using incomplete or fully generated data. The data processing pipeline could then be reused once the complete user study data is available and the results could be evaluated following this blueprint.

The second solution was chosen as it promised to be more useful for future work. The generation of partly random data is described in the following paragraphs. The statistical analysis of the data is presented in section 5.2. The RCT with the current number of participants serves as a proof-of-concept study to prepare a larger user study. This study is already ongoing and will be used to validate the preliminary results in this work.

Synthetic Data Generation

As motivated above, the statistical analysis was performed both on the incomplete real data from the user study as well as on a fully generated synthetic data set. In this subsection, the generation of the synthetic data is described in detail.

The fully generated data set includes data from 30 users per group, resulting in a total of 90 users. Group 1 has no access to the goal-setting mechanism; group 2 can set continuous goals; and group 3 can use time-bound goals. Since the goals themselves are not part of the statistical analysis, as explained in section 3.2, only diary entries are generated. Over the course of four weeks, i.e., the planned duration of the user study, for each user, goal-relevant diary entry type (alcohol.count, caffeine.count, evening.totalScreenTime, morning.gotUpAt, morning.wentToBedAt), and date, a value is generated (or not) with a probability of 50%. On average, a user in this data set fills their sleep diary every other day. If a value is generated for a date, it is computed according to the following logic:

1. Find the maximum possible value for the given diary entry type max.

2. Generate a delta value to subtract from the most recent known value of that user and entry type using a random value from a Gaussian distribution. In the case of the first three entry types, let the mean of that distribution be $0|0.2|0.4 * max$ for groups 1, 2, and 3 respectively, and let the standard deviation be $0.2 * max$. As the wake-up and bedtime are meant to be optimized regarding their standard deviation rather than their mean, let the mean of the Gaussian distribution be 0 and the standard deviation be $0.3|0.2|0.1*max$ for groups 1, 2, and 3 respectively.

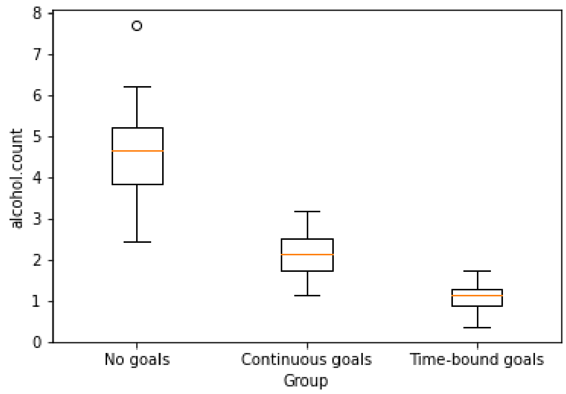
3. Subtract the delta from the most recent known value. Using the above distribution for the delta, groups 2 and 3, i.e. the groups that have access to goals, should have a higher chance of adopting healthier habits, i.e. consuming less caffeine and alcohol, spending less time looking at a screen, and getting up and going to bed more regularly, than group 1.

Statistical Analysis

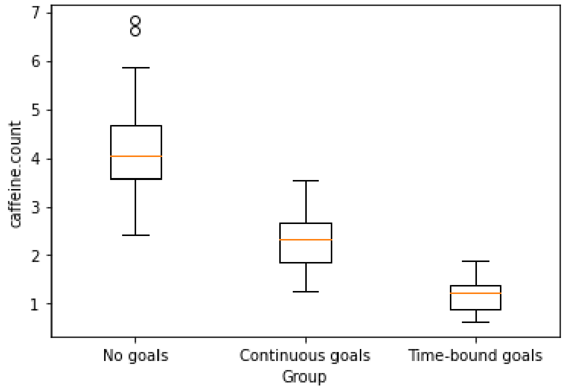
In the following paragraphs, the results of the statistical analysis according to the methods described in section 3.2 for both the incomplete data set collected in the ongoing user study and the fully generated data set are presented.

To begin with, the process is demonstrated using the fully generated data set. This data represents, to some extent, the expectations for the results, i.e. that using the goal-setting mechanism, especially with time-bound goals, leads to the adoption of healthier sleep-related habits. However, as described above, this data is completely artificial. The primary outcome measure is the difference in mean values per group and diary entry type. Figure 5.1 depicts the distributions of daily mean caffeine and alcohol consumption, and screen time as well as the standard deviation of daily wake-up and bedtime for each user over the whole course of the study as box plots. For each entry type, 0 is the optimal value and the lower values are better regarding each diary entry type. The box plots display the median and the first and third quartile of the values per entry type and group. As intended during data generation, the time-bound goals group performs best for each diary entry type, followed by the continuous goals group. The group without goal access performed worst for all diary entry types.

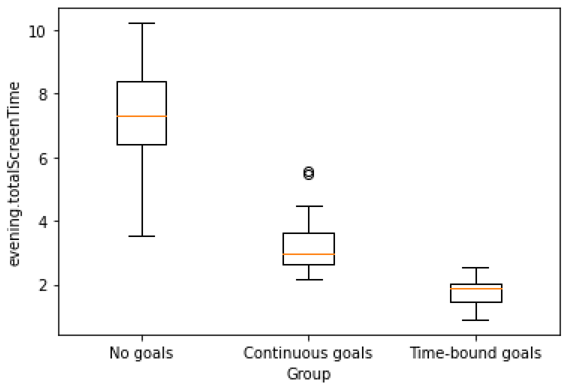
Table 5.1 contains the p-values for the differences between all pairs of two groups and for each diary entry according to the Kruskal-Wallis test [36]. In this work, a significance level of 1% is applied, meaning the difference between two groups is considered statistically significant if the p-value does not exceed 0.01. As introduced above, group 1 refers to the group without access to the goal-setting mechanism; group 2 has access to continuous goals; group 3 can use time-bound goals. For the generated data, all differences between groups are statistically significant except for the difference between groups 2 and 3 for the bedtime diary entry. This observation can also be made in the corresponding box plot (see Figure 5.1e) where the boxes show similar quartiles for these groups.



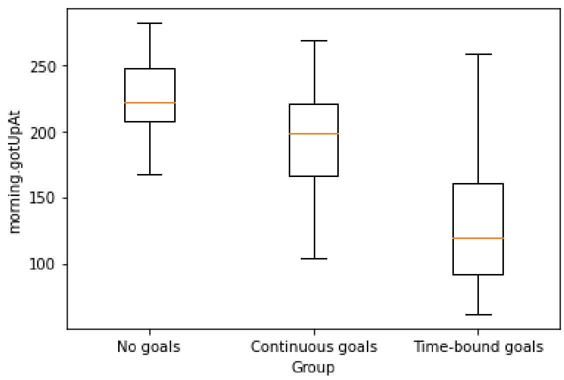
(a) Mean daily consumption of alcoholic beverages per study group



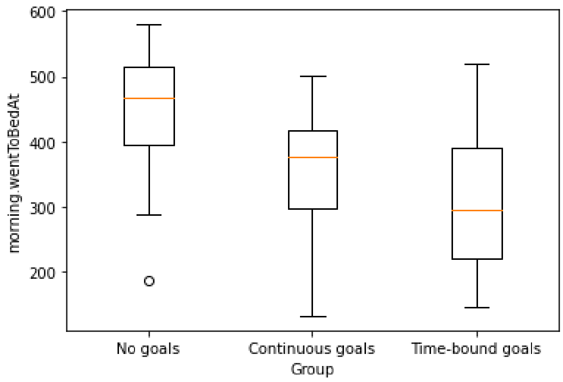
(b) Mean daily consumption of caffeinated beverages per study group



(c) Mean daily screen time per study group in hours



(d) Standard deviation of daily wake-up time per study group in minutes



(e) Standard deviation of daily bedtime per study group in minutes

Figure 5.1: Descriptive statistics per study group for generated data set as box plots

The secondary outcome measure regards the mean daily caffeine and alcohol consumption, and screen time, and the standard deviation of the wake-up and bedtime per group and week. The measure compares the values of the first and last week, expecting bigger differences, i.e. more progress, for the goal groups. The line charts

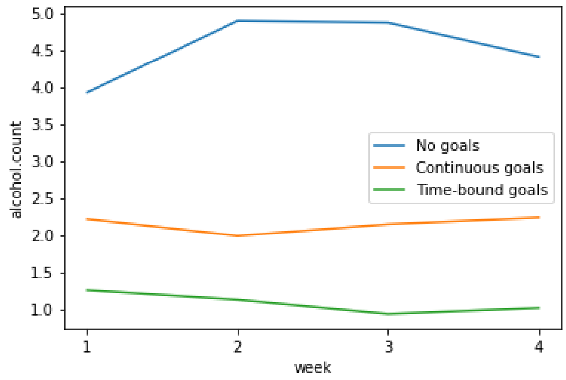
Table 5.1: p-values according to the Kruskal-Wallis test [36] for differences in the overall mean or standard deviation between user study groups per diary entry type using generated data set

Group A	Group B	Diary entry	p-value
1	2	alcohol.count	$p < 0.0001$
1	3	alcohol.count	$p < 0.0001$
2	3	alcohol.count	$p < 0.0001$
1	2	caffeine.count	$p < 0.0001$
1	3	caffeine.count	$p < 0.0001$
2	3	caffeine.count	$p < 0.0001$
1	2	evening.totalScreenTime	$p < 0.0001$
1	3	evening.totalScreenTime	$p < 0.0001$
2	3	evening.totalScreenTime	$p < 0.0001$
1	2	morning.gotUpAt	$p \sim 0.0010$
1	3	morning.gotUpAt	$p < 0.0001$
2	3	morning.gotUpAt	$p < 0.0001$
1	2	morning.wentToBedAt	$p \sim 0.0002$
1	3	morning.wentToBedAt	$p < 0.0001$
2	3	morning.wentToBedAt	$p \sim 0.0668$

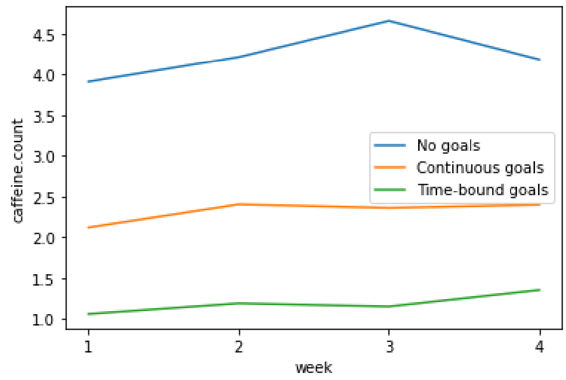
in Figure 5.2 display the values over the course of the study for each study group and diary entry type. The expected effect is most apparent in Figure 5.2c and Figure 5.2e for the screen time and bedtime diary entries. However, most lines show either no or no consistent progress over time. Consequently, the differences between the first and last week of the study do not differ significantly between different groups: the p-values exceed 0.01 clearly for every comparison and diary entry type, the only exception being groups 1 and 2 for screen time (see Table 5.2).

Table 5.2: p-values according to the Kruskal-Wallis test [36] for differences in the progress of mean or standard deviation between user study groups per diary entry type from week 1 to week 4 using generated data set

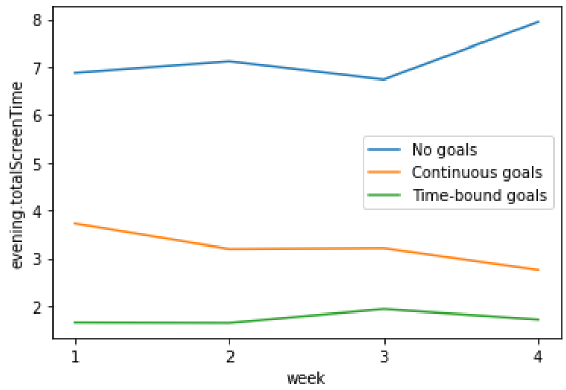
Group A	Group B	Diary entry	p-value
1	2	alcohol.count	0.24
1	3	alcohol.count	0.13
2	3	alcohol.count	0.29
1	2	caffeine.count	0.96
1	3	caffeine.count	0.98
2	3	caffeine.count	0.99
1	2	evening.totalScreenTime	0.12
1	3	evening.totalScreenTime	0.56
2	3	evening.totalScreenTime	0.007
1	2	morning.gotUpAt	0.54
1	3	morning.gotUpAt	0.78
2	3	morning.gotUpAt	0.62
1	2	morning.wentToBedAt	0.79
1	3	morning.wentToBedAt	0.39
2	3	morning.wentToBedAt	0.86



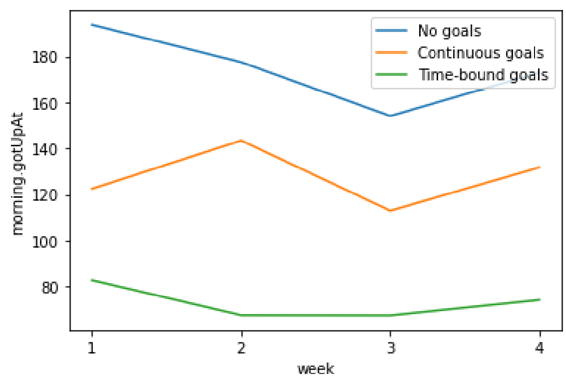
(a) Mean daily consumption of alcoholic beverages per study group over time



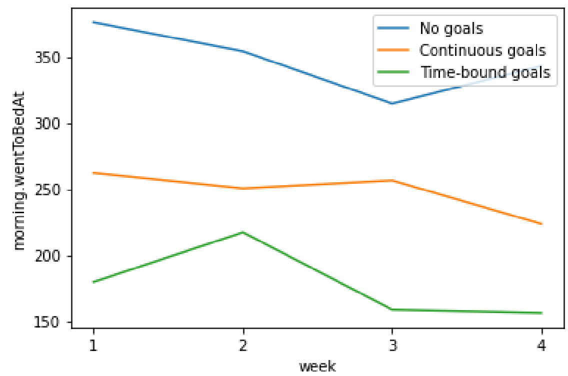
(b) Mean daily consumption of caffeinated beverages per study group over time



(c) Mean daily screen time per study group over time in hours



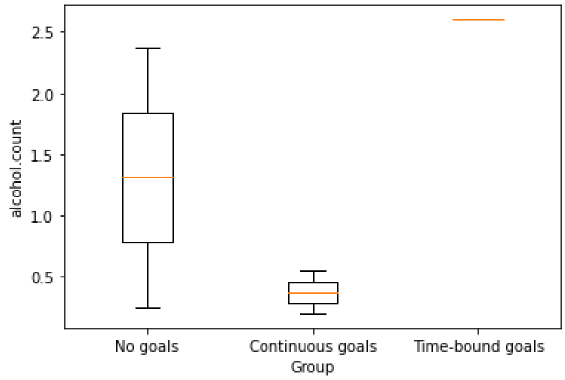
(d) Standard deviation of daily wake-up time per study group over time in minutes



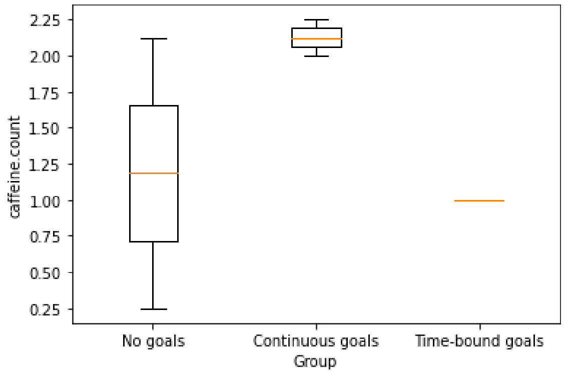
(e) Standard deviation of daily bedtime per study group over time in minutes

Figure 5.2: Descriptive statistics per study group over time for generated data set as line charts

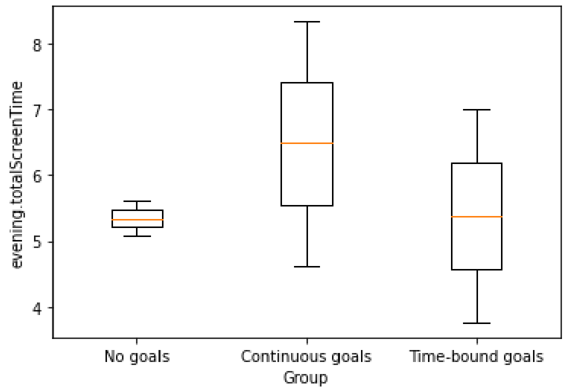
The same statistical analysis can be performed on the second data set that contains data for six users and the first two weeks of the user study and, in the future, on the complete data set once the user study terminates after four weeks and, possibly, after recruiting more participants to establish the statistical significance of the results. Figure 5.3 depicts the descriptive statistics for the different groups and diary entry types as box plots. These plots, again, can be used for demonstration purposes but do not provide a basis for conclusions as each group only contains two users. Consequently, the results do not seem to display any particular structure. In Figure 5.4, the different mean and standard deviation values are shown over time for the first two weeks that are available in the data set. Although the results, again, are far from statistically significant, tendencies in the lines resemble the expectation that groups with access to goals, especially time-bound, adopt healthier habits over time. In the charts, the goal groups improve their performance, i.e. show a decrease in value, for almost all diary entry types. Group 1, which does not have access to the goal-setting mechanism, only improves in performance considering their bedtime but deteriorates regarding their alcohol and caffeine consumption. These conclusions drawn from the graphs provide an example for the actual analysis of the complete user study data when that is available in the future.



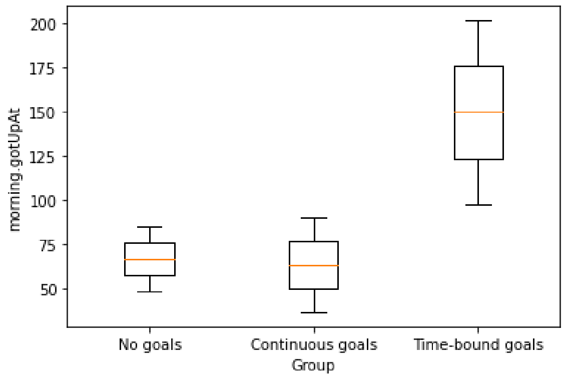
(a) Mean daily consumption of alcoholic beverages per study group



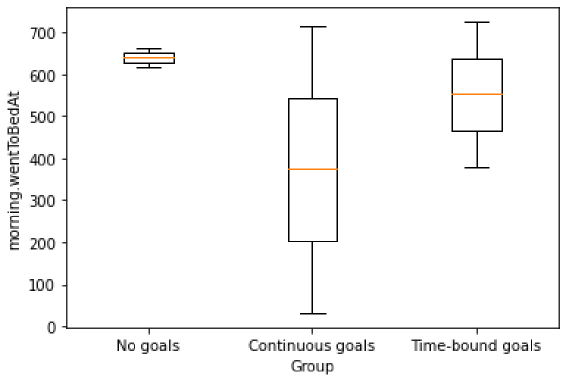
(b) Mean daily consumption of caffeinated beverages per study group



(c) Mean daily screen time per study group in hours

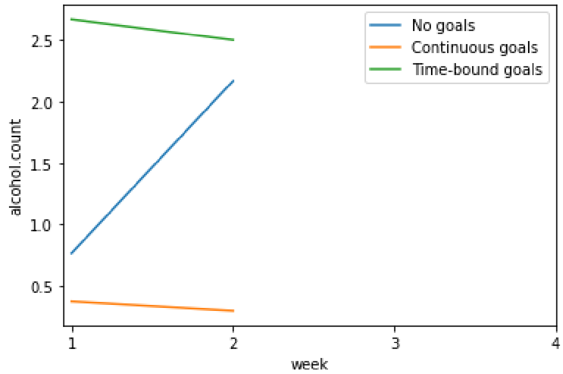


(d) Standard deviation of daily wake-up time per study group in minutes

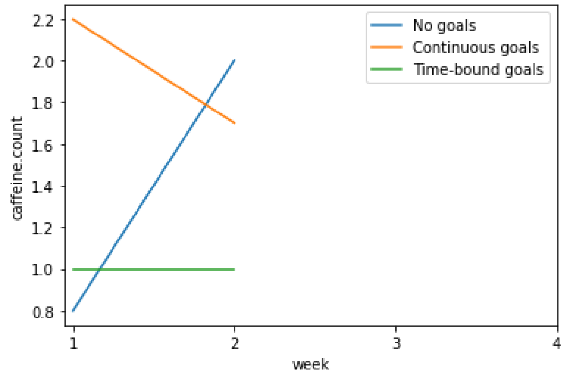


(e) Standard deviation of daily bedtime per study group in minutes

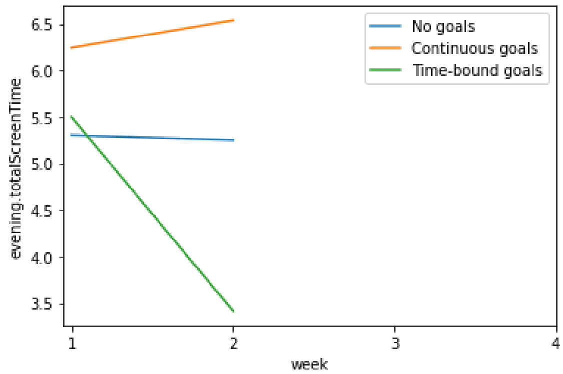
Figure 5.3: Descriptive statistics per study group for incomplete user study data set as box plots



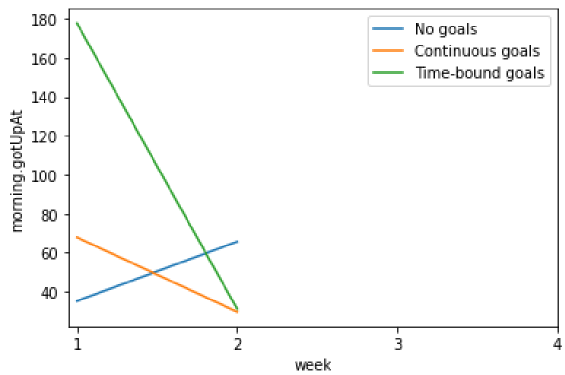
(a) Mean daily consumption of alcoholic beverages per study group over time



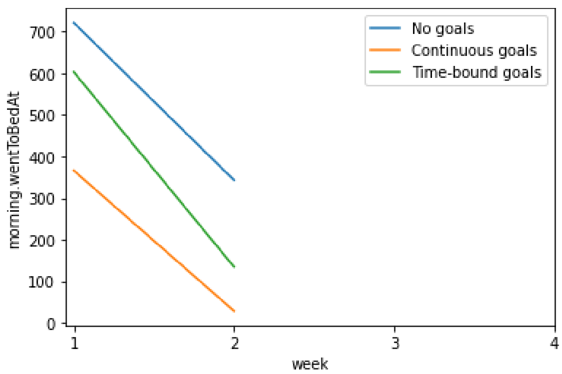
(b) Mean daily consumption of caffeinated beverages per study group over time



(c) Mean daily screen time per study group over time in hours



(d) Standard deviation of daily wake-up time per study group over time in minutes



(e) Standard deviation of daily bedtime per study group over time in minutes

Figure 5.4: Descriptive statistics per study group over time for incomplete user study data set as line charts

Chapter 6

Discussion

This chapter discusses the goal-setting mechanism based on the theoretical evaluation and the preliminary results of the user study from chapter 5 as well as the benefits and challenges of the different machine learning approaches presented in section 4.2. As the results are either theoretical conceptualizations or based on pilot and synthetic data, this discussion also provides several recommendations for directions in future research.

6.1 Goal-Setting Mechanism

The main takeaways from the analysis of the goal-setting mechanism in chapter 5 can be summarized as follows:

1. The implementation of the feature followed recommendations and considered insights from previous research in goal-setting theory, gamification, and mHealth. Theoretical concepts are applied stringently.

2. The goal-setting mechanism was preliminarily evaluated in a proof-of-concept user study. A quantitative evaluation was demonstrated using artificial and preliminary data. An extensive study to validate the promising results is ongoing and will be the subject of future work.

To discuss the results in more detail, this section starts with an overview of the implemented goal-setting mechanism and the theoretical evaluation from section 5.1. The latter discusses the implementation in comparison with recommendations from the literature already which is why this part is only summarized here.

To begin with, this work differentiates between two types of goals, continuous and time-bound, to investigate the effect of time-bound goals in more detail as suggested in [11]. Both goal types are implemented and share several features and screens, however, differ in the following way:

- Continuous goals are daily goals that, once set, continue indefinitely unless they are updated or removed by the user. They promote the daily achievement of each goal, e.g., through the display of the current goal streak and the weekly and monthly completion percentage. They are time-bound only in the sense that they can be achieved or not each day individually, meaning that they are measured at the end of the day.

- Time-bound goals have a fixed duration set by the user and a target number of days to achieve the goal within that time. They have a clear end date, and, therefore can be considered time-bound, as required by the SMART goals principle [9]. Due to the target number of days, in addition to the limit that applies every day, they are more specific than continuous goals. This follows the core finding of goal-setting theory that goals should be specific and difficult [7].

Each user in the user study (see section 3.2) has access to either continuous or time-bound goals - or to neither if they belong to the control group - to compare the effects of the different goal types on the user's behavior.

With each goal a user sets, they set a goal schema that links directly to a specific entry in the sleep diary in the Sleep Revolution app. The goal schemas that are available in the initial version of the goal-setting mechanism were chosen to cover a variety of sleep-related habits and can be extended. That allows for personalizing the set of goals to each user's current situation which has been found to elevate the effectiveness of gamified mHealth interventions [12], [13]. The goals can directly target certain habits that potentially impact sleep quality negatively. The user can choose the goals they consider most useful to them. The implementation of the goal-setting mechanism in the Sleep Revolution app gives the user complete control over their goals. Users can create, update, and remove goals and observe their progress towards their currently set goals constantly in several visualizations, including progress bars and calendars. At the same time, notifications guide the user to set challenging, yet realistic goals as suggested in previous literature [7], [9], [10].

Concepts from goal-setting theory, including specific and difficult goals, mediators of task performance [7], [8], and SMART goals [9], were implemented stringently and are resembled in almost every screen of the goal-setting feature. Personalized goal recommendations provide a promising addition to the feature to ensure that goals reflect the user's current progress and situation. Encouraging regular re-evaluation of the goals through notifications is important and a good compromise between implementation effort and accuracy of goal adjustments. The feature also satisfies Deterding et al.'s definition of gamification [5]. Only two game elements were used, namely goals and progress bars. Although there are many other game elements that have been recommended to facilitate goal-setting in gamified applications [11], there are good reasons to apply gamification sparingly: too many game elements might draw the focus of the user from the real-world goals, i.e. improving their quality of sleep, towards artificial goals such as badges or climbing in a leaderboard [37]. Therefore, the implementation finds a reasonable balance between reinforcing the user's goals and keeping the focus on exactly those. Previous work on notifications in mHealth interventions produced several recommendations [19]–[21]. However, the focus here is on goal-setting and both the implementation effort and the potential side effects were reasons to keep the notifications supporting the feature minimal. Notifications in the Sleep Revolution app are the topic of other research within the project.

The evaluation of the preliminary user study results presented in section 5.2 do not allow for an extensive discussion yet due to the size of the sample and the length of the study. However, the statistical analysis was demonstrated using the available data as well as another synthetic data set. Initial results are not statistically significant but meet the expectations regarding the effects of the goal-setting mechanism on user compliance. Most importantly, the users of the intervention groups actively engaged with the goals which is the foundation of meaningful results. These promising first findings will be validated in an ongoing user study with more participants. The evaluation of the complete user study remains a subject of future work. The study and recruitment is ongoing in order to find more participants that will use the goal-setting mechanism over a course of at least four weeks. In addition to the quantitative evaluation, interviews should be conducted to understand and optimize the participants' usage of the feature in the app. The implementation of the goal-setting mechanism with two types of goals, continuous and time-bound, and the ongoing user study pave the way to close the gap in the literature that is the effects of different goal types, and especially time-bound goals, on user behavior. This continuation of research, among others, is detailed further in chapter 8 that presents plans and opportunities for future work.

6.2 Goal Recommendations

This section discusses the three ML algorithms presented in section 4.2 regarding their applicability to the problem of generating personalized goal recommendations. Rather than implementing and evaluating the different models empirically, this work provides a starting point for further research by theorizing about their usage and extracting the potential benefits and challenges of each model.

To begin with, RNNs [32], according to Beutel et al. [38], are at the cutting edge of recommendation systems today. They naturally provide a higher-order model of feed-forward neural networks and are specifically designed to process sequential data, such as the diary and goal data in this work. They are also a standard ML model and, therefore, can be implemented rather easily using common libraries, e.g. *keras* in Python. As explained in section 4.2, they can be used in multiple ways to produce personalized goal recommendations. I.e., in a rather simple way, to predict the future user behavior that is then used to recommend goals according to a heuristic, and, in a more complex scenario, to generate the recommendations directly. That makes RNNs flexible toward potential challenges. As RNNs, especially when applied to the different goal schemas separately, can handle the unprocessed time series efficiently and recognize patterns in the data autonomously, little pre-processing or manual effort is needed. The model can be trained on existing data and training does not have to be integrated into the Sleep Revolution app. However, to yield adequate results, RNNs require large amounts of data. That is especially the case when the data becomes more complex, e.g. when the goal schemas are not separated but considered jointly and the RNN is expected to produce goal recommendations directly.

Bandit algorithms [33] are also commonly employed in recommendation systems [39] and, therefore, are a promising solution for this goal recommendation scenario. In contrast to RNNs, they provide a relatively simple way to deal with the complex shape of goals. RNNs usually produce simpler outputs which is another reason why they might be more suitable to predict the user's future performance rather than generating goals. As explained in section 4.2, the main challenge of using a bandit algorithm would be to model the context, i.e. the user's current situation, to find appropriate goals, especially when combining the different goal schemas in one model. Bandit algorithms also typically require a relatively large amount of data, however, less so than RNNs. Another key issue to consider when employing bandit algorithms is that they work best when integrated into the deployed system to explore different solutions in different situations. If there is enough data available, this process can be simulated.

The third model considered in this work is the BAA [15] which differs from the first two approaches in that it does not require a lot of data to produce accurate outputs. It has been applied to a similar problem in [34] to generate daily step goals. Therefore, it is very promising for this goal-setting mechanism, especially when applied to each goal schema separately. The main challenge is the development of a behavioral model that uses the user's past data to predict their future behavior and optimize the goals accordingly. This model gets increasingly complex when the goal schemas are modeled together in one model and when more goal schemas are added in the future. However, it enables modeling the mutual influence of different behaviors to be modeled precisely, so that existing knowledge about user behavior can be utilized. The original paper that introduced the algorithm [15] does not provide the source code used. Without that and further insight into how a behavioral model can be developed, the manual effort required to apply this algorithm increases even more. However, it is the most promising approach to deal with the lack of data at the moment.

The two most important issues to consider when choosing one of these models to use for personalized goal recommendations in the Sleep Revolution app are whether to generate the goals for each goal schema separately or jointly for all goal schemas, and the trade-off between the need for large amounts of data and the manual effort required to adjust the model to the problem at hand. Considering the different challenges each approach would introduce, it is advisable to start by modeling the different goal schemas separately and, for an empirical evaluation, to even reduce the number of goal schemas taken into account in the beginning. With the current lack of data, BAA is a reasonable choice to start with and explore, i.e. focusing on the manual effort to model user behavior. Once more data is available, RNNs could be easily trained and evaluated thanks to the minimal manual effort required and the availability of libraries to implement RNNs. Bandit algorithms, apart from the need for data, introduce several other challenges that make it the least promising option in this scenario.

Chapter 7

Ethical Considerations

Several works have identified potential ethical issues in both goal-setting and gamification practices. In this section, these issues are listed and discussed regarding their applicability to the gamified goal-setting mechanism in this work and their mitigation.

In [40], Ordoñez et al. find several "powerful and predictable side effects" [40, p. 3] of goals and recommend to prescribe goal setting selectively, to include warning labels, and to closely monitor the usage. They focus their research on goal setting in organizational contexts. However, the issues are formulated in a general way and can be applied to this goal-setting mechanism too. The following list contains the identified side effects in italic type and a discussion of the effect in this work in regular type:

- *Goals can focus attention so narrowly that people exclusively focus on that goal rather than other goals or aspects.*
  The goal schemas in this work are designed to provide a variety of goals and are meant to be extended in the future. However, each goal schema focuses on one diary entry only. E.g., the caffeine goal schema only considers the amount of caffeinated drinks in one day but ignores the time these drinks were consumed. This keeps the initial implementation in this work minimal but should be subject to change in the future, if the mechanism continues to become a permanent feature of the Sleep Revolution app.

- *The time horizon of a goal may be inappropriate for the user's current situation.*
  The time horizon of goals varies in this implementation and is one of the features that is explicitly explored and investigated in this work. Therefore, identifying inappropriate time horizons is an integral part of the conducted research.

- *Goals promote risky behavior and encourages users to adopt riskier strategies.*
  The goals promote behavior change towards healthier sleep-related habits. The goal descriptions provide rationales and explanation how a goal supports the quality of sleep to ensure a healthy adoption of the behavior. This behavior change does not imply any risks in most cases. However, in case of a caffeine or alcohol addiction, the adherence to too strict goals may lead to withdrawal symptoms. The adaptation of goals to the user's current situation should avoid that and, in addition, a severe caffeine or alcohol addiction is not a very common case.

- *Goals can promote unethical behavior, i.e. people may choose unethical methods to reach the goal or misrepresent their performance level, especially when they*

are close to reaching it.
In this context, as each goal only concerns the user's daily activities, there are barely any unethical ways to achieve the goals. As the sleep diary is filled manually, though, it is possible for users to represent their activities incorrectly. The goals do not reward the users in any other way than with the achievement of the goal itself. Therefore, the users do not have much reason to lie in their sleep diary. This potential effect should be taken into account, however, when analyzing the user data in this or future work.

- *Goal failure can impact the user's self-efficacy negatively and have other psychological consequence.*
  The continuous re-evaluation of goals helps to avoid setting unrealistic goals, and failing them in the case of time-bound goals. The low barriers to update and replace goals also offer ways to prevent users from failing too difficult goals. There are no consequences of failing a goal in the app either that would increase the impact of it.

- *Specific and challenging may inhibit learning by exploring different alternatives.*
  Again, the variety of available goal schemas should provide multiple ways to improve the user's quality of sleep and each goal can be reached using different strategies. In fact, adopting healthier habits according to the goals can inspire other changes in a person's daily activities.

- *Goals can create a culture of competition.*
  This is mostly true in contexts where users can view the progress of their peers and compare themselves to others, e.g., on leaderboards. The Sleep Revolution app provides no such mechanism and the users participate in studies independently, in contrast to professional settings where the people work together.

- *Goals can increase extrinsic motivation that is less sustainable and effective than intrinsic motivation.*
  The goals in this work are designed to facilitate the users' overall aim to improve their quality of sleep by changing related behavior. Therefore, the goals in the app represent exactly the real-world goals of the users and do not introduce artificial goals that would indeed shift the focus from the actual objective.

Kim et al. also found potential ethical issues in gamified applications [37] that are listed and addressed following this paragraph in the same way as above:

- *Gamification replaces real incentives with artificial ones.*
  This concern is shared with the issues in goal-setting practices and, as mentioned above, does not apply in this implementation as the goals in the app clearly represent real-world objectives that are set by a user exactly because they deem them relevant.

- *Gamification, by definition, aims to change its players' behavior and, therefore, manipulates the players to adopt new behaviors. Manipulation especially occurs when a player concludes, upon rational reflection, that their time would have been better spent otherwise.*
  Again, this goal-setting mechanism promotes goals in accordance with the overall aim to improve the user's quality of sleep. Therefore, the goals support the user

in what is the very reason why they chose to use the application. There is no need to manipulate the user.

- *Gamification can cause physical and psychological harm when players act in ways the designer did not anticipate or adopt negative character traits.*
  As discussed above, physical harm, if at all, might be inflicted by too strict goals in case the user has to make significant changes in their behavior, e.g., when curing an addiction. However, these cases are rare and too strict goals are meant to be prevented by adjusting them to the user's current situation. Negative character traits, as described in [37], often arise from interactions within gamified systems. The Sleep Revolution app does not include such a social aspect.

It is apparent that most of the ethical issues are addressed or do not apply to the goal-setting mechanism in this work. Those that potentially cause side-effects are mitigated by the implementation but should be continued to be considered in future work and, if necessary, investigated in more detail.

Chapter 8

Limitations and Future Work

This chapter addresses the limitations of this work and it derives and presents opportunities for future work that are mentioned in this thesis. Some future research has already been initiated and is ongoing within the Sleep Revolution project.

The small user study population and short duration only allowed for a proof-of-concept study that showed promising, yet no statistically significant results. These results need to be validated in a larger RCT ideally with at least 30 participants per intervention and control group, i.e. a total of 90 participants. The user study conducted in this work and its recruitment are currently ongoing to realize such a validation trial. This would allow for both a statistical analysis of the results as proposed in section 3.2 and a qualitative evaluation of the goal-setting mechanism, e.g. through interviews. That way the effects of continuous and time-bound goals on user compliance can be measured and the way the goal-setting mechanism is used by the participants can be understood and optimized. If the initial results are confirmed in the RCT and once the necessary adjustments are made to the goal-setting feature, it can be integrated to the Sleep Revolution app as a permanent feature. That would allow for further and even larger-scale analysis and further enhancements of the mechanism, e.g. by incorporating personalized goal recommendations.

Several ML algorithms were explored and theoretically evaluated regarding their applicability to generating these recommendations in section 4.2 and section 6.2. For the near future, BAA was found to be the most promising solution due to the size of the available data from the current user study. It is capable of producing meaningful results even with short histories of data per user. However, the algorithm requires the development of an adequate and mathematically sound behavioral model. This would be a reasonable first step in implementing a goal recommendation system. A request for further details about the creation of such a model in [34] as well as the source code is currently pending. Another starting point could be the implementation of a heuristic to generate personalized goals and integrating these into the implemented goal-setting feature in the Sleep Revolution app. section 4.2 briefly outlines a design idea for the recommendations: adding them as transparent tiles to the goal overview and pre-setting the values in the goal creation screen (see Figure 4.3b). Once a sufficient amount of data is available from further user studies, another promising approach to generate personalized goal recommendations are RNNs. They offer two different ways to accomplish the task: predicting the future diary entries of a user and computing the recommendations using heuristics, or producing the goal recommendations directly based on the users' past diary entries and goals. This allows for a trade-off between manual effort and data requirements. Apart from the development of an effective heuristic, the implementation of RNNs is comparatively simple due to the availability of a variety of libraries, such as keras in Python. An alternative to using machine learning or heuristics to generate personalized goal recommendations - however, expensive - is consulting experts. Another user study could investigate the effectiveness of such an approach compared to generated goals and potentially derive insights to integrate into a heuristic for automating the process.

Other possible extensions of the implemented goal-setting mechanism include:

- The addition of more goal schemas, e.g. a time limit to consume alcohol and caffeine and to use a screen, a maximum deviation from the target wake-up and bedtime, a minimum exercise duration per day, and maximum amounts and duration of naps and snoozing. Generally, goal schemas can potentially target any existing diary entry in the Sleep Revolution app using the current implementation.

- Experimenting with further gamification features and game design elements, e.g. in-app characters that motivate the user and send notifications as suggested in [19], badges, levels, and an avatar that are all commonly used game elements to facilitate goal setting [11].

- The integration of action planning towards currently set goals, such as providing strategies to reduce caffeine or alcohol consumption.

Two side-effects of relying on manual data entries in the sleep diary are that participants forget to enter their data or, due to currently set goals, intentionally misrepresent their daily activities as discussed in chapter 7. The currently available data from the user study suggests that participants do not consistently fill in the sleep diary every day. Evaluation of future studies should take potential side-effects into consideration, such as consistently forgetting to fill in the evening diary when going out to drink alcohol. That way, alcohol consumption would be systematically underrepresented in the user's sleep diary. However, goals can only be achieved when the corresponding diary entry is filled on that day. Interviews are a promising means to investigate the effect of goal setting on data entries and whether participants tend to misrepresent their daily activities to achieve their goals.

To conclude this chapter, there are a numerous opportunities for future work and, especially, further user studies to examine the expected effects on user compliance as well as side-effects of the current implementation and future extensions.

Chapter 9

Conclusion

In this thesis a goal-setting mechanism was developed according to theoretical concepts and recommendations from previous work and integrated into the Sleep Revolution mobile application. It was initially evaluated in a proof-of-concept user study and the promising results are subject of ongoing research. To extend the feature, three ML approaches were explored and compared regarding their applicability to generate personalized goal recommendations.

In chapter 2, the theoretical background and related work in the field of goal setting in mHealth interventions was presented. In particular, this theoretical foundation included an overview of the fields and intersections of mHealth, goal-setting theory, gamification, personalization, and ML. Based on the theoretical concepts and recommendations from these previous research activities, two types of goals, continuous and time-bound, were developed for the Sleep Revolution app. In [11], Tondello et al. suggested that different goal types, and, especially, time-bound goals, have received little interest which is the reason for the focus in this work. Continuous goals refer to daily goals that continue indefinitely unless the user updates or removes them. Time-bound goals, in contrast, have a fixed duration and a target number of days to achieve the goal within this time period. Each goal corresponds to one of the app's diary entries, with daily caffeine and alcohol consumption, screen time, wake-up, and bedtime available in the initial version of the implementation in this work. In the app, users have full control over their goals with the option to add, update, and remove goals. The user's currently set goals can be viewed at a glance in a comprehensive goal overview screen and progress towards the goals is visualized in the form of progress bars and calendars in this overview and each goal's detail view. This goal-setting mechanism implements theoretical concepts from goal-setting theory and stringently follows recommendations regarding goal setting, gamification, and notifications in mHealth interventions from previous work.

The goal-setting mechanism was initially evaluated in a proof-of-concept user study that was designed as a four-week RCT with three groups: two intervention groups with access to either continuous or time-bound goals, and one control group without access to the goal-setting feature in the app. After the first two weeks of the study and considering the theoretical evaluation, the preliminary results confirm the expectation that goals - time-bound even more so than continuous - increase the effectiveness of the intervention and foster behavior change towards healthier sleep-related habits. However, the results need to be validated in a larger study, that is already ongoing, according to the quantitative measures proposed in section 3.2. In addition, interviews should be conducted to understand and optimize the participants' usage of the goal-setting mechanism.

In a second step, three ML algorithms, namely RNNs, bandit algorithms, and BAA, were explored and theoretically evaluated to generate personalized goal recommendations. RNNs provide multiple ways to produce these recommendations based on the users' past diary entries and goals. Either, the future diary entries of a user could be predicted using the RNN and the goal recommendations would then be derived using a heuristic, or the RNN could generate the goals directly. In both cases, large amounts of data are required to train the model. MABs learn from trying different goals and evaluating them according to the user's behavior following setting these goals. They require appropriate context modeling as well as sufficiently large data. BAA is the only algorithm that could work well with the available data. However, it requires the manual development of a behavioral model that is used to predict a user's future performance and to optimize the goals accordingly. Nevertheless, it is advised to start implementing personalized goal recommendations using this algorithm to work with the already available data and investigate RNNs empirically, once sufficient data is collected in future user studies.

Bibliography

[1] *MHealth: New horizons for health through mobile technologies* (Global observatory for eHealth series). [Geneva]: World Health Organization, 2011, vol. v. 3, ISBN: 9789241564250.

[2] E. S. Arnardottir, A. S. Islind, M. Óskarsdóttir, *et al.*, "The sleep revolution project: The concept and objectives", *Journal of sleep research*, vol. 31, no. 4, e13630, 2022. DOI: 10.1111/jsr.13630.

[3] B. F. Sveinbjarnarson, L. Schmitz, E. S. Arnardottir, and A. S. Islind, "The sleep revolution platform: A dynamic data source pipeline and digital platform architecture for complex sleep data", *Current Sleep Medicine Reports*, 2023. DOI: 10.1007/s40675-023-00252-x.

[4] L. Schmitz, B. F. Sveinbjarnarson, G. N. Gunnarsson, and Dav\'\ia, "Towards a digital sleep diary standard", 2022.

[5] S. Deterding, D. Dixon, R. Khaled, and L. Nacke, "From game design elements to gamefulness: Defining" gamification", in *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*, 2011, pp. 9–15.

[6] D. Johnson, S. Deterding, K.-A. Kuhn, A. Staneva, S. Stoyanov, and L. Hides, "Gamification for health and wellbeing: A systematic review of the literature", *Internet interventions*, vol. 6, pp. 89–106, 2016. DOI: 10.1016/j.invent.2016.10.002.

[7] E. A. Locke and G. P. Latham, "Building a practically useful theory of goal setting and task motivation. a 35-year odyssey", *The American psychologist*, vol. 57, no. 9, pp. 705–717, 2002, ISSN: 0003-066X. DOI: 10.1037//0003-066x.57.9.705.

[8] G. P. Latham and E. A. Locke, "New developments in and directions for goal-setting research", *European Psychologist*, vol. 12, no. 4, pp. 290–300, 2007, ISSN: 1016-9040. DOI: 10.1027/1016-9040.12.4.290.

[9] G. B. Moskowitz and H. Grant, *The psychology of goals*. Guilford press, 2009.

[10] D. Baretta, P. Bondaronek, A. Direito, and P. Steca, "Implementation of the goal-setting components in popular physical activity apps: Review and content analysis", *Digital health*, vol. 5, p. 2 055 207 619 862 706, 2019, ISSN: 2055-2076. DOI: 10.1177/2055207619862706.

[11] G. Tondello, H. Premsukh, and L. Nacke, "A theory of gamification principles through goal-setting theory", in *Proceedings of the 51st Hawaii International Conference on System Sciences*, T. Bui, Ed., ser. Proceedings of the Annual Hawaii International Conference on System Sciences, Hawaii International Conference on System Sciences, 2018. DOI: 10.24251/HICSS.2018.140.

[12] L. Gosetto, F. Ehrler, and G. Falquet, "Personalization dimensions for mhealth to improve behavior change: A scoping review", *Studies in health technology and informatics*, vol. 275, pp. 77–81, 2020. DOI: 10.3233/SHTI200698.

[13] G. F. Tondello, R. R. Wehbe, L. Diamond, M. Busch, A. Marczewski, and L. E. Nacke, "The gamification user types hexad scale", in *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*, A. Cox, Z. O. Toups, R. L. Mandryk, and P. Cairns, Eds., New York, NY, USA: ACM, 2016, pp. 229–243, ISBN: 9781450344562. DOI: 10.1145/2967934.2968082.

[14] A. Knutas, R. van Roy, T. Hynninen, M. Granato, J. Kasurinen, and J. Ikonen, "A process for designing algorithm-based personalized gamification", *Multimedia Tools and Applications*, vol. 78, no. 10, pp. 13 593–13 612, 2019, ISSN: 1380-7501. DOI: 10.1007/s11042-018-6913-5.

[15] Y. Mintz, A. Aswani, P. Kaminsky, E. Flowers, and Y. Fukuoka, *Behavioral analytics for myopic agents*, 2017. DOI: 10.48550/ARXIV.1702.05496.

[16] D. Hallberg and N. Salimi, "Qualitative and quantitative analysis of definitions of e-health and m-health", *Healthcare informatics research*, vol. 26, no. 2, pp. 119–128, 2020, ISSN: 2093-3681. DOI: 10.4258/hir.2020.26.2.119.

[17] M. Nacinovich, "Defining mhealth", *Journal of Communication in Healthcare*, vol. 4, no. 1, pp. 1–3, 2011. DOI: 10.1179/175380611X12950033990296.

[18] S. P. Rowland, J. E. Fitzgerald, T. Holme, J. Powell, and A. McGregor, "What is the clinical value of mhealth for patients?", *NPJ digital medicine*, vol. 3, p. 4, 2020. DOI: 10.1038/s41746-019-0206-x.

[19] F. Muench and A. Baumel, "More than a text message: Dismantling digital triggers to curate behavior change in patient-centered health interventions", *Journal of medical Internet research*, vol. 19, no. 5, e147, 2017. DOI: 10.2196/jmir.7463.

[20] J. Woodward, Y.-P. Chen, K. Jurczyk, K. M. Ross, L. Anthony, and J. Ruiz, "A survey of notification designs in commercial mhealth apps", in *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021. DOI: 10.1145/3411763.3451639.

[21] A. Sahami Shirazi, N. Henze, T. Dingler, M. Pielot, D. Weber, and A. Schmidt, "Large-scale assessment of mobile notifications", in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2014, pp. 3055–3064. DOI: 10.1145/2556288.2557189.

[22] S. Consolvo, P. Klasnja, D. W. McDonald, and J. A. Landay, "Goal-setting considerations for persuasive technologies that encourage physical activity", in *Proceedings of the 4th international Conference on Persuasive Technology*, pp. 1–8. DOI: 10.1145/1541948.1541960.

[23] M. Montola, J. Stenros, and A. Waern, *Pervasive games: theory and design*. CRC Press, 2009.

[24] Y.-K. Chou, *Actionable gamification: Beyond points, badges, and leaderboards*. Packt Publishing Ltd, 2019.

[25] O. Rafieian and H. Yoganarasimhan, "Ai and personalization", *Artificial Intelligence in Marketing*, pp. 77–102, 2023.

[26] W. Kim, "Personalization: Definition, status, and challenges ahead", *Journal of object technology*, vol. 1, no. 1, pp. 29–40, 2002.

[27] M. Rabbi, P. Klasnja, T. Choudhury, A. Tewari, and S. Murphy, "Optimizing mhealth interventions with a bandit", in *Digital Phenotyping and Mobile Sensing*, ser. Studies in Neuroscience, Psychology and Behavioral Economics, H. Baumeister and C. Montag, Eds., Cham: Springer International Publishing, 2019, pp. 277–291, ISBN: 978-3-030-31619-8. DOI: 10.1007/978-3-030-31620-4{\textunderscore }18.

[28] I. Portugal, P. Alencar, and D. Cowan, "The use of machine learning algorithms in recommender systems: A systematic review", *Expert Systems with Applications*, vol. 97, pp. 205–227, 2018, ISSN: 09574174. DOI: 10.1016/j.eswa.2017.12.020.

[29] R. Michalski, J. Carbonell, and T. Mitchell, "Machine learning: An artificial intelligence approach", *Artificial Intelligence*, vol. 25, no. 2, pp. 236–238, 1985, ISSN: 00043702. DOI: 10.1016/0004-3702(85)90005-0.

[30] Z.-H. Zhou, *Machine learning*. Singapore: Springer, 2021, ISBN: 978-981-15-1966-6.

[31] C. M. Bishop, "Neural networks and their applications", *Review of scientific instruments*, vol. 65, no. 6, pp. 1803–1832, 1994.

[32] L. R. Medsker and L. C. Jain, "Recurrent neural networks", *Design and Applications*, vol. 5, pp. 64–67, 2001.

[33] A. Slivkins, "Introduction to multi-armed bandits", *Foundations and Trends® in Machine Learning*, vol. 12, no. 1-2, pp. 1–286, 2019, ISSN: 1935-8237. DOI: 10.1561/2200000068.

[34] M. Zhou, Y. Fukuoka, Y. Mintz, *et al.*, "Evaluating machine learning-based automated personalized daily step goals delivered through a mobile phone app: Randomized controlled trial", *JMIR mHealth and uHealth*, vol. 6, no. 1, e28, 2018, ISSN: 2291-5222. DOI: 10.2196/mhealth.9117.

[35] S. Hochreiter and J. Schmidhuber, "Long short-term memory", *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997, ISSN: 0899-7667. DOI: 10.1162/neco.1997.9.8.1735.

[36] P. E. McKight and J. Najab, "Kruskal–wallis test", in *The Corsini Encyclopedia of Psychology*, I. B. Weiner and W. E. Craighead, Eds., Wiley, 2010, p. 1, ISBN: 9780470170243. DOI: 10.1002/9780470479216.corpsy0491.

[37] T. W. Kim and K. Werbach, "More than just a game: Ethical issues in gamification", *Ethics and Information Technology*, vol. 18, no. 2, pp. 157–173, 2016, ISSN: 1388-1957. DOI: 10.1007/s10676-016-9401-5.

[38] A. Beutel, P. Covington, S. Jain, *et al.*, "Latent cross", in *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, 2018.

[39] D. Goldenberg, K. Kofman, J. Albert, S. Mizrachi, A. Horowitz, and I. Teinemaa, "Personalization in practice: Methods and applications", in pp. 1123–1126. DOI: 10.1145/3437963.3441657.

[40] L. D. Ordóñez, M. E. Schweitzer, A. D. Galinsky, and M. H. Bazerman, "Goals gone wild: The systematic side effects of overprescribing goal setting", *Academy of Management Perspectives*, vol. 23, no. 1, pp. 6–16, 2009, ISSN: 1558-9080. DOI: 10.5465/amp.2009.37007999.

Appendix A

Appendix

A.1 Goal Schema Descriptions

Alcohol

Alcohol can disrupt the quality of your sleep by interfering with your body's natural sleep cycles, leading to less restful sleep and more waking up during the night. It can cause breathing issues like snoring and sleep apnea, which further disrupt sleep quality. To promote better sleep, it's recommended to limit alcohol consumption, especially close to bedtime.

Bedtime

Irregular sleep schedules can disrupt the circadian rhythm, causing difficulty falling asleep and daytime fatigue. Having a fixed bedtime can benefit your sleep by regulating your body's internal clock, making it easier to fall asleep and wake up at consistent times, promoting better sleep quality and improving daytime alertness.

Caffeine

Caffeine blocks a natural chemical that signals your body to sleep, and consuming caffeine late in the day can make it harder to fall asleep, reduce deep sleep, and increase wake-ups during the night. By reducing caffeine intake, especially in the afternoon and evening, you can improve your chances of getting a good night's sleep and wake up feeling refreshed.

Screen Time

Exposure to bright screens before bedtime can suppress melatonin production, making it harder to fall asleep and reducing sleep quality. Additionally, engaging with stimulating content can further disrupt relaxation before bedtime. By reducing screen exposure, you can improve sleep quality, and enhance your ability to fall and stay asleep.

Wake-Up

Irregular wake-up times can disrupt the sleep-wake cycle, leading to difficulty falling asleep and feeling tired in the morning. Having a fixed wake-up time helps to regulate your body's circadian rhythm, making it easier to fall asleep at night and wake up feeling rested.