SAARLAND UNIVERSITY

Faculty of Mathematics and Computer Science
Department of Computer Science
Dissertation

Understanding Personal and Contextual Factors to Increase Motivation in Gamified Systems

Dissertation zur Erlangung des Grades des
Doktors der Ingenieurwissenschaften (Dr.-Ing.)
der Fakultät für Mathematik und Informatik
der Universität des Saarlandes

vorgelegt von
(M.Sc.) Maximilian Altmeyer
Saarbrücken
2021

Dean:	Professor Dr. Jürgen Steimle
Date of the colloquium:	May 3, 2022
Chairperson of the Examination board:	Professor Dr. Jürgen Steimle
Reporter:	Professor Dr. Antonio Krüger
	Professor Dr. Elisa Mekler
	Professor Dr. Lennart Nacke
Scientific Assistant:	Dr. Pascal Lessel

Notes on Style:
Throughout this thesis, the scientific plural "we" is used to reflect that the majority of the work that is presented in this thesis was done in collaboration with other students and researchers. References to web resources (e.g., links to articles published on the Internet) are provided as URLs together with the date of last access; longer URLs have been shortened to enhance the readability. In line with the guidelines established by the Association for Computing Machinery[1], charged terminology and gender pronouns are avoided throughout the thesis. Instead of mentioning gender, we used "they" as the neutral singular pronoun.

[1] ACM: *Words Matter - Alternatives for Charged Terminology in the Computing Profession*, `https://www.acm.org/diversity-inclusion/words-matter` (last accessed: 2021-12-01)

Acknowledgments

Before the thesis starts, I would like to thank those who made it possible by supporting me in many different ways.

I thank **Antonio Krüger** for his trust in this research topic being worth investigating, for supervising me and for providing valuable feedback on paper submissions and on this thesis in particular. Moreover, I am very thankful for the great work environment, three different kinds of coffee machines and for offering myriad possibilities to socialise with colleagues (pinball machines, game consoles, climbing walls...). Thank you!

I am also very grateful for the support of **Lennart Nacke**. Thank you for inviting me to your research group and for making my stay an unforgettable experience! I was very impressed by how involved you were in our collaborations and very thankful for the detailed and elaborated feedback you provided. Thank you in particular for reviewing this thesis.

Thank you **Elisa Mekler** for always being available to chat about Self-Determination Theory in games/gamification research and other topics! Thank you especially for agreeing to review this thesis. I really hope you enjoyed reading it!

I thank **Margaret De Lap** for proofreading this thesis and all papers contributing to it. I am sure your efforts convinced Reviewer 2 in many instances!

Most probably, I would not have started my PhD without **Pascal Lessel**. Thank you for sparking my interest in (gamification) research when advising me in my bachelor and master theses. Thank you also for all your advise, for our discussions (especially during long train rides), for your advise that led to discussions :) and for supporting (and challenging) new research ideas. A special thanks for providing feedback to this thesis!

The great work environment mentioned before would not be great without great colleagues. Luckily, I can say that this is the case. I always liked going to work, seeing each one of you and chatting about research, life and gossip. Although Covid-19 changed these habits recently, I hope it gets better soon! Thank you **Frederic Kerber** for sharing the office with me, providing help and feedback to on- and off-topic subjects (including DIY-advises) and for entertaining me and those around you with a huge fund of idioms. Thank you **Felix Kosmalla** for various funny (research?) projects, culinary highlights and Friday afternoon beer sessions (alcohol-free, of course). Thank you **Donald Degraen** for always offering help, fun conversations and for teaching me the most important words to get along in Belgium (Wa make, Palettekop?). I also would like to thank **Marc Schubhan** for being a great HiWi, an exemplary student and a great Mario Kart opponent. I am happy that you decided to start your PhD in the games (or gamification?) field! Thank you **Nico Herbig** and **André Zenner** for great times at the Software Campus seminars and beyond (Greek food and Whisky was definitely a highlight, although not necessarily in this order!). Thank you **Florian**

Daiber for providing wise suggestions and feedback and especially for being confused in a very entertaining way. Over the past years, there were also a lot of things to organize (refunds, ethical review board requests, travelling,...). I am very thankful for your help **Iris Lambrecht**, **Gundula Kleiner**, **Lisa-Marie Merziger** and the whole DFKI administration. Since the list of people who worked with me at DFKI is too long, I want to say thank you at this point to all of you!

Thanks also to all students who selected me as their thesis advisor. I hope you enjoyed writing your thesis with me. A special thanks goes to (in alphabetic order) **Kathrin Dernbecher, Vladislav Hnatovskiy, Subhashini Jantwal, Linda Muller, Tobias Sander, Marc Schubhan, Atiq Ur Rehman Waqar** and **Berina Zenuni**, as the results of our joint studies are now part of this dissertation. I would also like to thank **Patrick Schuck** for your help and support when working as a HiWi with us (and thereafter!). Thank you **Hanne Spelt** for sharing your expertise in analyzing psychophysiological data and for all your input and feedback. A big thanks also goes to **Gustavo Tondello** for providing a lot of helpful advice, feedback and for sharing your knowledge about personalization of gameful systems with me! I would also like to thank **Katja Rogers**, **Sukran Karaosmanoglu**, **Stuart Hallifax**, **Enrica Loria**, **Wilk Oliveira** and **Ana Cláudia Guimarães Santos** for involving me in your research – it was super helpful exchanging ideas with you.

Lastly, I want to say thank you to my family. I am so grateful for having the best **parents** I could imagine. Thank you for always supporting me, independent of what I was doing. I am also very thankful to have my sister **Paulina** in my life, who makes me laugh and is there for me whenever I need her. Finally, this thesis would not have been possible without my wife **Denise**. I am the luckiest person to have you on my side. Thank you for cheering me up when I feel bad, for supporting me and for encouraging me to do so many things in life that I would not have been brave enough to do alone.

Abstract

Gamification, the use of game elements in non-game contexts, has been shown to help people reaching their goals, affect people's behavior and enhance the users' experience within interactive systems. However, past research has shown that gamification is not always successful. In fact, literature reviews revealed that almost half of the interventions were only partially successful or even unsuccessful. Therefore, understanding the factors that have an influence on psychological measures and behavioral outcomes of gamified systems is much in need. In this thesis, we contribute to this by considering the context in which gamified systems are applied and by understanding personal factors of users interacting with the system. Guided by Self-Determination Theory, a major theory on human motivation, we investigate gamification and its effects on motivation and behavior in behavior change contexts, provide insights on contextual factors, contribute knowledge on the effect of personal factors on both the perception and effectiveness of gamification elements and lay out ways of utilizing this knowledge to implement personalized gamified systems. Our contribution is manifold: We show that gamification affects motivation through need satisfaction and by evoking positive affective experiences, ultimately leading to changes in people's behavior. Moreover, we show that age, the intention to change behavior, and Hexad user types play an important role in explaining interpersonal differences in the perception of gamification elements and that tailoring gamified systems based on these personal factors has beneficial effects on both psychological and behavioral outcomes. Lastly, we show that Hexad user types can be partially predicted by smartphone data and interaction behavior in gamified systems and that they can be assessed in a gameful way, allowing to utilize our findings in gamification practice. Finally, we propose a conceptual framework to increase motivation in gamified systems, which builds upon our findings and outlines the importance of considering both contextual and personal factors. Based on these contributions, this thesis advances the field of gamification by contributing knowledge to the open questions of how and why gamification works and which factors play a role in this regard.

Zusammenfassung

Gamification, die Nutzung von Spielelementen in spielfremden Kontexten, kann nachweislich Menschen helfen, ihre Ziele zu erreichen, das Verhalten von Menschen zu beeinflussen und die Erfahrung der User in interaktiven Systemen zu verbessern. Allerdings hat die bisherige Forschung gezeigt, dass Gamification nicht immer erfolgreich ist. So haben Literaturübersichten ergeben, dass fast die Hälfte der Interventionen nur teilweise erfolgreich oder sogar erfolglos waren. Daher besteht ein großer Bedarf, die Faktoren zu verstehen, die einen Einfluss auf psychologische Maße sowie auf das Verhalten von Usern in gamifizierten Systemen haben. In dieser Arbeit tragen wir dazu bei, indem wir den Kontext, in dem gamifizierte Systeme eingesetzt werden, betrachten und persönliche Faktoren von Usern, die mit dem System interagieren, verstehen. Geleitet von der Selbstbestimmungstheorie, einer der wichtigsten Theorien zur menschlichen Motivation, untersuchen wir Gamification und dessen Auswirkungen auf Motivation und Verhalten in Kontexten zur Verhaltensänderung. Wir liefern Erkenntnisse über kontextuelle Faktoren, tragen Wissen über den Einfluss persönlicher Faktoren auf die Wahrnehmung und Effektivität von Gamification-Elementen bei und bieten Möglichkeiten, dieses Wissen für die Implementierung personalisierter gamifizierter Systeme zu nutzen. Unser Beitrag ist mannigfaltig: Wir zeigen, dass Gamification die Motivation durch Bedürfnisbefriedigung und durch das Hervorrufen positiver affektiver Erfahrungen beeinflusst, was letztlich zu Verhaltensänderungen führen kann. Darüber hinaus zeigen wir, dass das Alter, die Absicht, das Verhalten zu ändern, und Hexad-Usertypen eine wichtige Rolle bei der Erklärung von interpersonellen Unterschieden in der Wahrnehmung von Gamification-Elementen spielen. Ebenso zeigen unsere Resultate dass die Anpassung von gamifizierten Systemen auf Basis dieser persönlichen Faktoren positive Auswirkungen auf psychologische und verhaltensbezogene Ergebnisse hat. Letztlich zeigen wir, dass Hexad-Usertypen teilweise durch Smartphone-Daten und Interaktionsverhalten in gamifizierten Systemen vorhergesagt werden können und dass sie auf spielerische Art und Weise erhoben werden können. Dies ermöglicht, unsere Erkenntnisse in der Gamification-Praxis zu nutzen. Auf Basis dieser Ergebnisse schlagen wir ein konzeptuelles Framework zur Steigerung der Motivation in gamifizierten Systemen vor, das die Wichtigkeit der Berücksichtigung sowohl kontextueller als auch persönlicher Faktoren hervorhebt. Diese Erkenntnisse bereichern das Forschungsfeld Gamification, indem sie Wissen zu den offenen Fragen, wie und warum Gamification funktioniert und welche Faktoren in diesem Zusammenhang eine Rolle spielen, beitragen.

List of Publications

Parts of this thesis, including text fragments, figures, and tables have already been published. The chapters of this thesis are partly based on these publications. These publications as well as the chapter in which they appear, are listed in the following:

Journal papers:

[11] Maximilian Altmeyer, Pascal Lessel, Subhashini Jantwal, Linda Muller, Florian Daiber, Antonio Krüger. 2021. Potential and Effects of Personalizing Gameful Fitness Applications Using Behavior Change Intentions and Hexad User Types. In *User Modeling and User-Adapted Interaction 31 (1)*. Springer, 675–712. (appears in Chapter 5)

Full conference papers:

[21] Maximilian Altmeyer, Berina Zenuni, Hanne Spelt, Pascal Lessel, Antonio Krüger. 2022. Do Hexad User Types Matter? Effects of (Non-) Personalized Gamification on Task Performance and User Experience in an Image Tagging Task. UNDER REVIEW at *Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '22)*. (appears in Chapter 5)

[16] Maximilian Altmeyer, Pascal Lessel, Atiq Ur Rehman Waqar, Antonio Krüger. 2021. Design Guidelines to Increase the Persuasiveness of Achievement Goals for Physical Activity. In *Proceedings of the International GamiFIN Conference (GamiFIN '21)*. CEUR-WS, 40–49. (appears in Chapter 4)

[17] Maximilian Altmeyer, Marc Schubhan, Antonio Krüger, Pascal Lessel. 2021. A Long-Term Investigation on the Effects of (Personalized) Gamification on Course Participation in a Gym. In *Proceedings of the International GamiFIN Conference (GamiFIN '21)*. CEUR-WS, 60–69. (appears in Chapter 5)

[20] Maximilian Altmeyer, Gustavo F. Tondello, Antonio Krüger, Lennart E. Nacke. 2020. HexArcade: Predicting Hexad User Types By Using Gameful Applications. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '20)*. ACM, 219–230. (appears in Chapter 6)

[8] Maximilian Altmeyer, Kathrin Dernbecher, Vladislav Hnatovskiy, Marc Schubhan, Pascal Lessel, Antonio Krüger. 2019. Gamified Ads: Bridging the Gap Between User Enjoyment and the Effectiveness of Online Ads. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, Paper 182, 1–12. (appears in Chapter 3)

[13] Maximilian Altmeyer, Pascal Lessel, Linda Muller, Antonio Krüger. 2019. Combining Behavior Change Intentions and User Types to Select Suitable Gamification Elements for Persuasive Fitness Systems. In *Proceedings of the International Conference on Persuasive Technology (PERSUASIVE '19)*. Springer, 337–349. (appears in Chapter 4)

[15] Maximilian Altmeyer, Pascal Lessel, Marc Schubhan, Vladislav Hnatovskiy, Antonio Krüger. 2019. Germ Destroyer – A Gamified System to Increase the Hand Washing Duration in Shared Bathrooms. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '19)*. ACM, 509–519. (appears in Chapter 3)

[14] Maximilian Altmeyer, Pascal Lessel, Tobias Sander, Antonio Krüger. 2018. Extending a Gamified Mobile App with a Public Display to Encourage Walking. In *Proceedings of the International Academic Mindtrek Conference (MindTrek '18)*. ACM, 20–29. (appears in Chapter 3)

Short conference papers:

[12] Maximilian Altmeyer, Pascal Lessel, Antonio Krüger. 2018. Investigating Gamification for Seniors Aged 75+. In *Proceedings of the Designing Interactive Systems Conference (DIS '18)*. ACM, 453–458. (appears in Chapter 4)

Posters:

[19] Maximilian Altmeyer, Marc Schubhan, Pascal Lessel, Linda Muller, Antonio Krüger. 2019. Using Hexad User Types to Select Suitable Gamification Elements to Encourage Healthy Eating. In *Extended Abstracts Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '20)*. ACM, 1–8. (appears in Chapter 4)

[18] Maximilian Altmeyer, Pascal Lessel, Marc Schubhan, Antonio Krüger. 2019. Towards Predicting Hexad User Types from Smartphone Data. In *Extended Abstracts Proceedings of the Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '19)*. ACM, 315–322. (appears in Chapter 6)

Workshops:

[9] Maximilian Altmeyer and Pascal Lessel. 2017. The Importance of Social Relations for Well-Being Change in Old Age - Do Game Preferences Change As Well?. In *Proceedings of the Positive Gaming: Workshop on Gamification and Games for Wellbeing co-located with the Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '17)*. CEUR-WS, 11–15 (appears in Chapter 4)

Moreover, at the time of submission, the author of this thesis supervised ten bachelor's (three ongoing) and nine master's (three ongoing) students. Implementations or user studies which were created or conducted as part of these thesis supervisions were in parts presented in the above mentioned publications. These theses are listed in the following, together with a reference to the corresponding paper(s) in which they were used:

- Atiq Ur Rehman Waqar. 2021. Enhancing the Persuasiveness – Behavior Change Intentions and the Perception of Achievement Goals for Physical Activity. *Master Thesis*. [16].
- Berina Zenuni. 2020. Investigating the Effects of Personalized Gamification on Task Performance, User Experience and Psychophysiological Reactions. *Master Thesis*. [21].
- Subhashini Jantwal. 2020. Endless Universe: A Gameful Fitness Application to Investigate the Effects of Behaviour Change Intentions and Hexad User Types on Task Performance and User Experience. *Master Thesis*. [11].
- Kathrin Dernbecher. 2019. Gamification in Online Advertising. *Master Thesis*. [8].
- Linda Muller. 2019. Personalizing Gamified, Persuasive Health Systems Using Behavior Change Intentions and Hexad User Types. *Bachelor Thesis*. [11,13,19].

- Marc Schubhan. 2018. Player Types 2 Go: Towards Predicting Hexad Player Types Using Smartphone Data. *Bachelor Thesis*. [18].
- Tobias Sander. 2018. Investigating the Effect of a Gamified Public Display to Encourage Physical Activity. *Bachelor Thesis*. [14].

Contents

Chapter 1
Introduction

This thesis contributes to the research field of gamification, "the use of game design elements in non-game contexts" ([105], p. 10). More specifically, we will contribute to the question of *why* and *how* gamification works, to better understand how to tailor it to the application context and the user, in order to enhance its outcomes. In this regard, we will shed light on how it affects motivation in behavior change contexts. We also contribute to the personalization of gamified systems by studying which factors mediate the perception of gamification, which effects personalized gamification has on psychological and behavioral outcomes, and how personalization can be facilitated.

To do so and to get an understanding of when gamification is considered to work, it is important to understand what gamification aims for. In this regard, gamification is described as "a developing approach for encouraging user motivation [..]" ([315], p. 16) or simply as "designing for motivation" ([104], p. 1). Thus, the ultimate goal of gamification is increasing motivation [104, 303, 315]. Consequently, we will investigate how and why gamification works through the lens of Self-Determination Theory [299], a major theory of human motivation.

We begin this chapter by highlighting the integral role of games and play for our culture and society. We highlight the great potential of video games in satisfying basic psychological needs and how they are consequently able to lead to pleasurable, motivational experiences. Afterwards, we focus on gamification, which aims to transfer these motivational experiences from games to a non-game context. We will also discuss gamification as a persuasive technology, i.e. the idea to utilize elements known from games to increase motivation for behavior change. After motivating the importance of investigating *contextual* and *personal* factors in order to realize gamified systems enhancing motivation, we establish the research questions and outline the structure of this thesis.

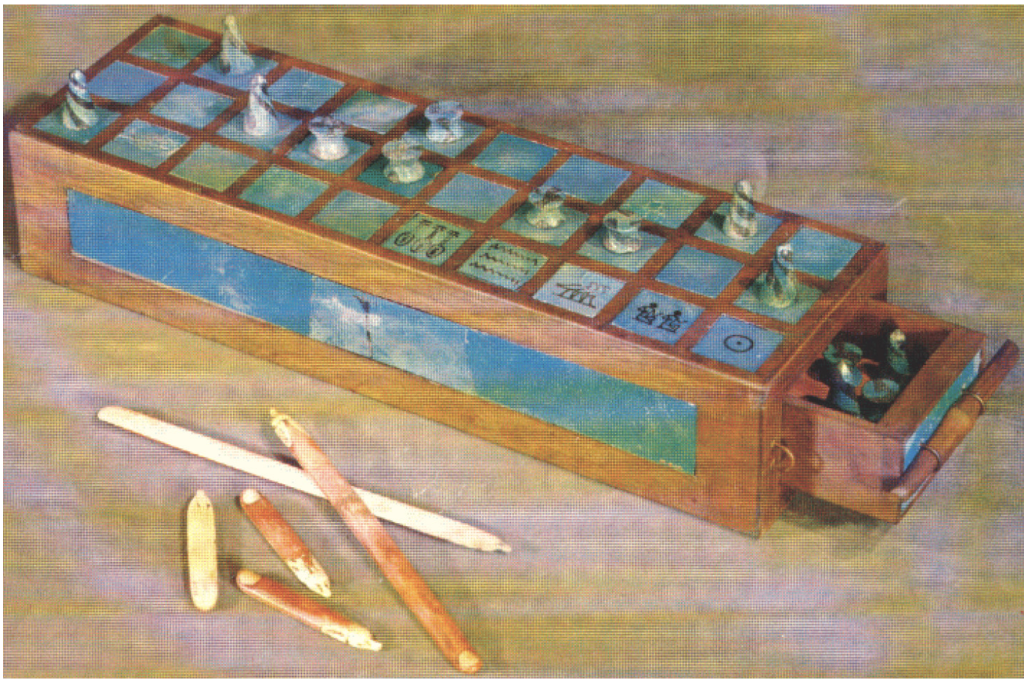


Figure 1.1: Ancient Senet board found in Abydos from about 1500 B.C. Taken from [270].

1.1 The Ubiquity of Games in Culture and Society

"In culture we find play as a given magnitude existing before culture itself existed, accompanying it and pervading it from the earliest beginnings right up to the phase of civilization we are now living in" (Huizinga [158], p. 4).

In their work, Huizinga describe Homo Ludens, "man the player", as a model according to which humans develop their personality and cultural abilities primarily through play. This model is based on the fact that all culture (e.g. religion, science, philosophy, art, politics) follows certain rituals containing elements of play. Moreover, Huizinga proclaims that humans discover their individual characteristics through the experiences they gain about themselves when playing. Thus, games and play are seen as the primary factors shaping both culture and personality.

In fact, the significance of games and play have a long tradition, since games were already played in early stone- and iron-age cultures, where toys were made from bones, ceramics, and iron [31]. Later, ancient cultures worldwide engaged in play and developed different types of games: Over 5,000 years ago, Egyptians invented a boardgame called Senet [270]. In this game, players had to strategically move game pieces across a board consisting of thirty squares, arranged in three rows of ten, to reach a superior position on the board. Figure 1.1 shows a Senet board from about 1,500 B.C. Also, games which are still popular today, such as Backgammon, were already played about 5,000 years ago in ancient Mesopotamia [57]. Besides board games, ancient cultures such as the Babylonians used talus bones as dice for games and amusement, more than 2,500 years ago [184]. Later, in the Middle Ages, people played many boardgames, such as Alquerque, a predecessor of Checkers, and engaged in a wide range of card, string and dice games [362].

Today, ongoing digitization and increasing global connectivity have transformed many aspects of our culture into the digital world. For instance, we are able to communicate and connect through social networks and real-time voice and video chats, or through completely new forms of synchronous or asynchronous communication. Ubiquitous computing technology is increasingly tracking our actions and behavior, and smartphones, allowing for interactive experiences and connectivity independent of place and time, are used by more than 85% of Americans[2]. These transformations of culture have also impacted and transformed how and what we play [300]. In video games, players can explore complex virtual worlds allowing for completely new forms of immersive experiences. They can perform a broad range of actions and behaviors, master challenges and interact with other players to compete or collaborate. Moreover, online video games are increasingly replacing traditional places to meet and interact with peers [249]. They have increasingly become popular over the past decades, making them – with an estimate of over 2.7 billion gamers worldwide and a revenue of $ 159.3 billion in 2020 [249] – the fastest-growing form of human recreation [249, 300]. However, elements of play are not only present in video games, but also in non-game contexts in digital culture: For instance, mobile fitness applications incorporate game elements such as points, badges or high score lists to motivate physical activity; websites borrow game elements to increase user participation; and online communities introduce levels or ranks to represent the reputation of their members. This practice of using elements known from games in non-game contexts has been defined as gamification [105]. The significant number of users worldwide who freely choose to spend a considerable amount of money, resources, and time indicates that there must be something about games and elements of games that captivates and motivates them. To better understand the underlying reasons, we will elaborate on the motivational power of games.

1.1.1 The Motivational Power of Video Games

In this thesis, we conceptualize motivation based on Self-Determination Theory ("SDT") [299], a leading[3] scientific theory of human motivation, which is widely used in the domain of games and gamification research [350]. SDT is a macro theory which consists of six mini-theories focusing on different aspects of motivation

[2] Pew Research Center: *Demographics of Mobile Device Ownership and Adoption in the United States*, `https://pewrsr.ch/3unCPmc` (last accessed: 2021-12-01)

[3] Self-Determination Theory: *An approach to human motivation*, `https://bit.ly/33wDCp2` (last accessed: 2021-12-01)

and well-being processes. A detailed description of SDT, the mini-theories that are relevant for this thesis, and a brief history of human motivation can be found in Section 2.1. At its core, SDT is based on three universal psychological needs which fuel motivation and well-being processes. The basic psychological need for *autonomy* refers to the experience of volition, having the feeling that actions are self-endorsed and authentic. The need for *competence* is satisfied when one feels in control of the outcome of an action, feeling capable and effective. Lastly, the need for *relatedness* describes the feeling of social belonging and reciprocal care. SDT further states that both *contextual* and *personal* factors play an important role in whether these basic needs are supported or thwarted (see Section 2.1.2 and Section 2.1.2). Thus, for stimuli such as game elements to be motivating, both these factors need to be considered and studied.

Contextual Factors

Context can be defined as "the circumstances that form the setting for an event, statement, or idea, and in terms of which it can be fully understood"[4]. This broad definition makes it hard to pinpoint single factors of a context. However, the context as a whole plays an integral part in SDT, as will be explained in more detail in Section 2.1.2. Broadly speaking, SDT states that the context may support or thwart the aforementioned basic psychological needs. As a consequence, this supports or undermines intrinsic motivation and has an effect on the extent to which behaviors are internalized. These processes are important for gamification, since they may have a considerable impact on the success of certain gamification elements and to what extent these affect motivation. That considering the context of a gamified system is important can be illustrated by an example from Disneyland in Anaheim, which introduced a leaderboard showing the speed of laundry workers on a public display. Although leaderboards have been shown to lead to positive outcomes in past research [146, 185], laundry workers at Disneyland were so afraid of the gamified system that they even skipped breaks[5]. Rather than leading to engagement, the system was perceived as an "electronic whip" and employees reported feeling "controlled even more". The fact that the users of the gamified system had an employment relationship is an important aspect in this context, as it may have shifted the type of motivation to extrinsic, because the system "made other employees worry that a reasonable pace won't be enough to keep the boss happy". Consequently, the feedback on one's performance might have been perceived as controlling rather than informative. In this thesis, we aim to enhance the understanding of how gamification affects motivation and related measures in different behavior change support contexts. Investigating this allows us to also reason more broadly about the role of contexts for the success of gamified systems.

[4] Lexico: *Definition of CONTEXT by Oxford Dictionary*, `https://bit.ly/2SIyc8n` (last accessed: 2021-12-01)
[5] Los Angeles Times: *Steve Lopez: Disneyland workers answer to 'electronic whip'*, `https://lat.ms/3vIdG6q` (last accessed: 2021-12-01)

Personal Factors

SDT states that motivational orientations differ across people (see Section 2.1.2), which means that people differ in how they interpret certain stimuli, such as gamification elements. As a consequence, whether these stimuli support or undermine intrinsic motivation is largely dependent on the person and their individual characteristics. In the field of gamification, past research has observed that such inter-personal differences are also prevalent in the perception of gamification elements [182]. While some users might embrace competition and challenges, other users might instead prefer exploratory gamification elements such as unlockable content or creativity tools [342]. These findings mark a turning point in gamification research – while most gamified systems relied on one-size-fits-all solutions, i.e. using the same set of gamification elements for all users, tailoring the gamified system to the user has gained more and more attention in research [186, 281]. To tailor gamified systems, several personal factors have been studied, including demographic factors such as gender or age, personality traits and player types [182]. In this thesis, we contribute to an enhanced understanding of such personal factors by considering age as a demographic factor, behavior change intentions as a dynamic factor related to the degree of internalization of behavior, and Hexad user types, representing the types of motivation that are particularly relevant for a user. We motivate the consideration of these factors based on SDT and provide insights on the relationships between the personal factors we have investigated and empirical findings from SDT research in Chapter 4.

Basic Psychological Needs in Video Games

When trying to understand why video games and elements of games are so attractive, we can observe that video games are incredibly powerful in increasing motivation by satisfying the aforementioned basic psychological needs introduced by SDT [299]. They provide a wide range of circumstances to fulfill the basic needs for autonomy, competence and relatedness [299], which are key to intrinsic motivation and the organismic integration processes surrounding extrinsic motivation (see Section 2.1.2). In simpler terms, Jesse Shell stated that "every game is a complex ecosystem of motivations" ([307], p. 153). Indeed, research has demonstrated that satisfactions of these basic needs lead to game enjoyment and even predict future play [300].

When we think of games such as *Animal Crossing* (see Figure 1.2 a), one of the most popular contemporary video game series where the current release ranks as the second-best-selling game on Nintendo's Switch console[6], the need for autonomy may be satisfied in myriad ways. First, the game allows one to create a human-like avatar representing the player in the virtual world, offering a substantial range of choices ranging from individual hairstyles to special clothes and accessories that

[6] Nintendo: *Top Selling Title Sales Units*,
`https://bit.ly/379x5CZ` (last accessed: 2021-12-01)

Figure 1.2: Games and need satisfaction. a) ***Animal Crossing: Pocket Camp*** **(by Nintendo, 2017; own screenshot). b)** ***Fruit Ninja 2*** **(by Halfbrick Studios, 2018; screenshot taken from press kit, `https://bit.ly/3by3S7I` (last accessed: 2021-12-01)). c)** ***World of Warcraft*** **(by Blizzard Entertainment Studios, 2004; screenshot taken from press kit, `https://bit.ly/3ibIF54` (last accessed: 2021-12-01)).**

can even be further customized to the players' preferences. Also, the core game mechanics are highly autonomous, as the player may freely decide to explore the open game world, talk to its inhabitants or collect and discover entities of the game world, such as new insects, fish or fossils, which can be traded in to obtain items that can be used to further customize the avatar and the virtual world. In fact, research has demonstrated that offering such choices in the game positively affects need satisfaction, and as a consequence leads to enhanced immersion and enjoyment: Birk et al. [42], for instance, allowed participants to create an avatar and adjust parameters related to the avatar's appearance, personality, and characteristics. They demonstrated that offering the choice to manipulate these attributes enhances the extent to which participants identified with the avatar. This, in turn, led to an increased satisfaction of autonomy, and thus fostered intrinsic motivation and game enjoyment. Besides the avatar, an open game world, allowing the player to travel around and freely explore and discover any part of it, inherently adds to the satisfaction of autonomy needs [299]. In popular video games such as *Grand Theft Auto V* or *World of Warcraft* this very element of exploration and discovery of the open world and the broad range of possible interactions with the environment is what many players enjoy [299]. While the Grand Theft Auto series has often been criticized because of its violence, research has shown that people primarily do not play the game because of the violence but because they can freely decide what they want to do [299]. Such a wide range of opportunities in games is what essentially makes them highly capable of supporting the need for autonomy [286, 299].

However, it is not only autonomy which can be supported by video games. When we think about what makes games so appealing, the rich, immediate feedback on our performance towards clearly communicated goals is an important aspect to consider, making games very effective in satisfying our innate need for competence [286, 299, 300]. A relevant example of competence support in games can be seen in casual games mostly played on mobile devices, such as *Fruit Ninja* (see Figure 1.2 b). In Fruit Ninja, a very popular game for mobile devices[7], the core game mechanic is very simple. Players have to slice fruits by performing a slicing gesture on the touch screen as fruits appear. The fact that the game is very easy to understand is part of its success, since we have to feel capable of eventually succeeding at a task to be motivated to engage in it (see Section 2.1.2). The game itself has various means to satisfy the need for competence – it gives instant performance feedback and provides stimulating challenges by altering the number of fruits shown at once or by introducing bombs that should not be sliced. For each fruit that is sliced, immediate positive feedback is given to the player in the form of points and praise, which is not controlling, but emphasizes the player's skill in an informative way. When slicing multiple fruits at once – called "combos" – the player receives extra points, adding to the feeling of mastery. But it is not only casual games that support the need for competence. In fact, almost every game has elements to it that are capable of satisfying the need for competence: In shooter games, killing enemies provides instant feedback and may evoke feelings of accomplishment; in racing games, it may be the the skill to master difficult race tracks or drive faster than the opponent and in platform games, solving puzzles to advance to the next level or combating enemies to receive power-ups may satisfy the need for competence very effectively [286]. In fact, besides game mechanics, it may even be the narrative of the game which adds positively to feelings of mastery, especially when the player takes on the role of a heroic character [286].

Lastly, games are also capable to fulfill relatedness needs: Referring back to the example of World of Warcraft (see Figure 1.2 c), a Massively Multiplayer Online Role-Playing Game ("MMORPG"), numerous gameplay elements are included, facilitating social interaction. The mere fact that players can simultaneously join a shared virtual space, see each other's avatar and communicate through various means ranging from gestures the avatar can perform to text messages and voice chats, may fulfill the need for social relatedness [299]. This need is further facilitated through features such as teaming up in guilds, cooperative play to accomplish tasks via teamwork, trading items among players or simply exploring the virtual game world together [299]. But relatedness is more than

[7] The Guardian: *Fruit Ninja: how the slice-and-dice game reached 1bn downloads*, `https://bit.ly/2ZcZt2W` (last accessed: 2021-12-01)

communication and belonging; it is equally important to have the feeling of being important to others (see Section 2.1.2). This aspect is inherently fulfilled in World of Warcraft (and similar games of this genre) by introducing several classes that players must select when creating their avatar, with choices such as priests or warriors. These classes have unique skills, rendering a certain importance to players of a particular class. Moreover, boss fights are often easier to complete in groups with a diverse set of characters, requiring users to value and appreciate each other's unique skills, which in turn might positively add to the feeling of reciprocal belonging. However, it is not only MMORPGs that support the need for relatedness – modern games incorporate leaderboards, shared goals and collaboration, or allow players to team up to play together in the virtual environment, offering a wide range of communication options. Therefore, it is not surprising that a cultural shift is being observed among younger people, with games increasingly replacing social networks as places to meet and interact with their peers [249].

Now that we have seen the potential of games for satisfying basic psychological needs and thus contributing to motivation and well-being, it seems intriguing to utilize these elements and experiences known from games in non-game contexts, to enhance motivation for various actions and desirable behaviors. This approach has been labeled *gamification* and will be introduced next.

1.1.2 Gamification: Game Experiences in Non-Game Contexts

When confronted with tasks or activities in which we find little or no interest or value, we may decide to either quit the activity, or seek aspects of the activity or its surrounding context which can be turned into strategies which foster more autonomous forms of motivation [305], e.g. by enhancing the perceived competence for a rather boring activity [286]. Sansone et al. [305] empirically demonstrated that when confronted with boring tasks, participants come up with a wide range of coping strategies, turning an uninteresting task into a more interesting one to self-regulate motivation and thus decouple part of it from external contingencies. This observation holds for many different contexts: Long distance truckers introduce certain game-like elements such as challenging themselves to spot deer near the road, whereas other workers cope with boredom by setting personal goals or pacing the tasks involved, e.g. by introducing time pressure [295].

In line with the strategies mentioned before, the concept of *gamification* builds upon a similar idea. The term itself was introduced roughly a decade ago, and many different definitions exist. For instance, Werbach [359] defines gamification as "the process of making activities more game-like" (p. 1), thus emphasizing the procedural aspect of gamification. Huotari and Hamari [161] follow this procedural approach. They define gamification as "a process of enhancing a service with affordances for gameful experiences in order to support [a] user's overall value creation" (p. 3). Although these definitions provide a good notion about what the concept of gamification means, they are less detailed about how processes can be made more "game-like" or how "gameful experiences" can be created.

In this thesis, we refer to gamification as "the use of game design elements in non-game contexts" ([105], p. 10), which is the most widespread definition [359], being cited more than 8900 times by December 2021, according to Google Scholar. The definition relies on "game elements", elements known from games, which can be used to create such gameful experiences. For the sake of this thesis, we consider game elements (also referred to as "gamification elements" when used in the context of gamification) as the "elements that are characteristic to games – elements that are found in most (but not necessarily all) games" ([105], p. 4), i.e. building blocks that are characteristic to games and respectively gamified systems [335], such as *points*, *badges*, *leaderboards*, or *virtual characters*. Although none of the definitions explicitly state that gamification is tied to digital computer systems, it is commonly approached and used within interactive systems [315]. Thus, although the examples given at the beginning of this section adhere to the definitions, it is arguable whether they would be considered gamification, since they do not incorporate any type of interactive system.

It is assumed that gamification has its roots in marketing and the digital media industry [105, 315], where e.g. loyalty programs have long since been using points and other collectibles. The advancement of technology and the substantial increase in digitization in the past years are considered as key enablers for gamification [239, 315]: A wide range of wearables and other tracking devices and systems are on the market nowadays, allowing users to seamlessly track their activities, while smartphones and other small-sized personal computers have become ubiquitous, allowing for a wide range of feedback and interaction channels [294]. As a consequence, gamification has become a thriving field in human-computer interaction research and has been investigated in a broad range of contexts, including education, health, crowdsourcing and commerce [146, 315]. Moreover, it has become an established practice in industry [239].

Before we provide some examples of how gamification is used in current products on the consumer market, we need to further differentiate various forms of game-like approaches, since not all of them are considered as gamification. *Serious games*, for instance, can be defined as "any form of interactive computer-based game software for one or multiple players to be used on any platform and that has been developed with the intention to be more than entertainment" ([289], p. 6). The term itself dates back to the work by Abt, which stated that "Games may be played seriously or casually. We are concerned with *Serious Games* in the sense that these games have an explicit and carefully thought-out educational purpose and are not intended to be played primarily for amusement. This does not mean that serious game are not, or should not be, entertaining" ([2], p. 9). Consequently, serious games – although serving purposes beyond mere entertainment – are considered as being a game, whereas gamification refers to adding game-like elements to a non-game context. Similarly, *games with a purpose* are regarded as

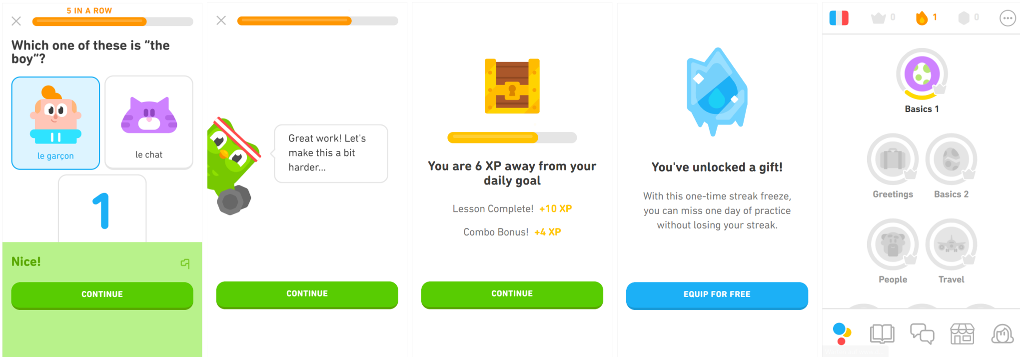


Figure 1.3: ***Duolingo*** **as an example of gamification, own screenshot**

games which are built to make players solve human computation tasks such as tagging images [354]. Thus, games with a purpose differ from serious games, as they are not concerned with learning. They, in line with serious games, also differ from gamification, as they are not adding game elements to a non-game context, but rather are games themselves.

There are numerous examples of the use of gamification in today's digital services or products. *Duolingo*, a web service to learn new languages, uses various game elements to add to the motivation of users and make learning a language more enjoyable (see Figure 1.3). Whereas users can set their own daily goals, Duolingo divides learning goals into smaller so-called exercises. In these exercises, progress bars are used, indicating how close users are to completing them. Also, encouraging sound effects provide positive feedback when tasks are solved correctly; this is also supported by virtual characters, providing further positive feedback from time to time. Completion of exercises is rewarded with experience points, power-ups or gems, and the user's skill tree is expanded, tracking their learning progress. In fact, the effectiveness of this gameful approach was also supported empirically, showing that Duolingo is very effective in teaching new languages [352].

Another example in the domain of physical activity is *Fitbit*. Fitbit is a company offering fitness trackers and corresponding mobile applications, allowing their users various insights into their physical activity (amongst others). In the mobile application, several game elements are employed such as daily step goals, progress feedback towards reaching these goals, badges that can be unlocked by being particularly active and leaderboards, through which users can compete with their friends. Thus, the mobile application falls under the definition of gamification, since various game elements are used to increase the motivation of users in a non-game context – physical activity in this case. Similar to Duolingo, research has demonstrated that Fitbit is well-accepted and effective in enhancing the motivation to walk, when compared against a standard pedometer [70].

There is also a huge number of other examples for gamification – websites such as *StackOverflow*, a Q&A platform mainly for programmers, use badges and points to represent users' the reputation; *PayPal*, a service for money transfers, uses progress feedback to motivate users to complete their profile information; *Starbucks*, a popular coffeehouse chain, introduced a loyalty program, where users can collect virtual stars by consuming products or interacting with the app. This unlocks levels depending on the degree of customer loyalty, and stars can be used to get certain products for free. Overall, it can be seen that gamification is almost ubiquitous in our daily lives, which is why research on this topic is not only relevant, but also essential to better understand how gamification affects motivation and how to prevent negative outcomes.

1.1.3 Gamification as Persuasive Technology

The term *persuasive technology* has been introduced by Fogg [126]. They define persuasion as "an attempt to change attitudes or behaviors or both (without using coercion or deception)" (p. 15) and thus refer to persuasive technology as "interactive computing systems designed to change people's attitudes or behaviors" (p. 1). In their book, they outline various benefits of computers as persuaders such as their interactivity, persistence, anonymity, capability to manage large amounts of data, scalability and ubiquity. Moreover, they see persuasive technology at the intersection of computers and persuasion, with persuasion including behavior change and motivation. Consequently, gamification, with its capability to motivate behaviors, can be seen as an instance of persuasive technology [106], with persuasive systems commonly using gamification strategies [7].

Indeed, gamification has been frequently investigated in the context of behavior change. Research investigated the capability of gamification to motivate people to do more sports [88], to eat more healthily [76], to drink enough water [82], to be more productive in the context of microtasks [224], to enhance learning outcomes [352] or to save energy [189], just to name a few. The effectiveness of gamification in such contexts has been demonstrated by literature reviews by Hamari, Koivisto and Sarsa [146] as well as Seaborn and Fels [315], concluding that gamification leads to positive behavioral and psychological outcomes.

1.2 Problem Statement and Contribution

Since its introduction to the academic field roughly a decade ago, there was soon substantial hype around gamification, with a rapidly growing literature on the topic [186]. This led to research on gamification being extended in various directions and domains, including health, commerce, education, work, crowdsourcing, sustainability, marketing, and others [146, 315]. In the first years of gamification research, the main research question was related to *whether* gamification works [239]. Since the literature quickly became massive, Hamari, Koivisto and Sarsa [146] published a literature review in 2014 summing up the question of whether gamification works by analyzing the results of articles published between 2010 and 2013. In 2015, Seaborn and Fels [315] followed with a similar literature review focused on the results and approaches in gamification research. The general answer to the question of whether gamification works or not was that, in most cases, gamification indeed can be considered as successful [146, 315]. Both literature reviews reported a broad range of both positive psychological and behavioral outcomes of gamification. However, they also found potential problems with gamification, and emphasized that the question of whether gamification works is multi-faceted, and that "gamification as a phenomenon is more manifold than the studies often assumed" ([146], p. 5). Indeed, it seems like the general attitude at this time took for granted that adding game elements to non-game contexts was inherently valuable [281]. In the midst of all the hype around gamification, many came to believe that "if you put points on something boring, add some badges, and provide a competitive leaderboard, that once boring product will automatically become exciting" ([84], p. 17 f.). As a consequence, this over-simplification of gamification and the mere focus on adding seemingly random game design elements to non-game activities led to estimations that as of 2014, 80% of current gamified applications would fail due to poor design [132]. Among the issues identified by the literature reviews in 2014 [146] and 2015 [315] was the disconnect between theoretical and applied work, such that theoretical work was not backed empirically, and applied work was not well-informed through theory [315]. This also includes the importance of understanding how certain game design elements affect motivation and which factors play a role in that respect [315]. Hamari, Koivisto and Sarsa [146] agree on this aspect, outlining that two main aspects – the role of the context being gamified and the quality of the users – should receive more attention. Thus, a major lesson learned from the initial phase of gamification research is that particular contexts and types of users should be studied more thoroughly and in a more theory-driven way.

In recent years, gamification research has evolved in several ways. For instance, research has come to the conclusion that one-size-fits-all approaches are inadequate to handle the huge amount of interpersonal differences in how gamification is perceived, leading to substantial efforts being spent in personalizing or customizing gamification approaches [182]. As a consequence, factors having a moderating influence on the perception of atomic gamification elements were extensively studied, including for example player or user typologies, gender, personality traits, age or culture, among others [182]. As part of this development, psychological theories were used increasingly in gamification, in particular Self-Determination Theory, informing research questions [281] and leading to a user typology specifically focused on gamification [342]. But even with the ongoing shift in gamification research from analyzing whether gamification works to better understanding *why* and *how* it works [239], Nacke et al. [239] emphasize that investigating "different individual and situational contexts is [...] very much in need today" (p. 5). In recent work, researchers highlight that research has not been able to provide definite answers to what constitutes or creates gameful experiences [186], that the mechanisms for how gamification affects motivation and behavior are mostly unclear and that the outcomes of gamification interventions are, although mostly positive, remarkably mixed [186]. In their literature review from 2019, Koivisto and Hamari [186] call for future gamification research to widen its thematic scope and study factors across a wider range of domains. They also emphasize considering the role of users and their individual attributes, as well as paying more attention to the usage context.

In this thesis, we contribute answers to the intertwined questions stemming from the effort to understand *why* and *how* gamification works. We investigate the effects of gamification on intrinsic motivation through the lens of SDT and across different behavior change contexts, allowing us to shed light on contextual factors and how they can support but also thwart need satisfaction and thus intrinsic motivation and organismic integration processes. Considering contexts is important, since both gamification researchers and SDT itself highlight the integral role that contextual factors play in the success of gamified interventions and human motivation, respectively. Moreover, besides contextual factors, we contribute insights on factors on the personal level that play a role in how motivating certain gamified interventions are. In this regard, we provide both theoretical knowledge about factors having an influence on the *perception* of certain gamification elements and empirical investigations on the *actual effects* of these factors – when using them to tailor gamified systems – on psychological and behavioral outcomes. Focusing on such differences on the personal level is equally important, since SDT, through Causality Orientations Theory (see Section 2.1.2), has demonstrated that motivation orientations can differ across individuals and since gamification research has demonstrated that accounting for these individual differences is essential to gamified systems. Moreover, to facilitate the utilization of these factors in practice, we contribute approaches to gain insights on individual preferences without disturbing the gameful experience. These contributions mainly advance the field of *gamification* within *human-computer interaction* ("HCI") and *persuasive technology*. More specifically, this thesis adds insights to ongoing efforts to understand how gamification affects *motivation* and to *personalization research* in the gamification domain. It also contributes domain-specific findings in the field of *public health*, *sports* and *advertising*, as well as adds to the body of knowledge regarding the field of *older adults in HCI*.

In sum, this thesis advances the understanding of how gamification affects motivation by investigating *contextual* and *personal* factors: We study to what extent basic needs in a given context can be satisfied by gamification; which personal factors have an influence on the perception of gamification elements; whether considering these factors in gamified systems has actual effects on motivation and behavior; and how we can facilitate the personalization of gamified systems without disturbing the gameful experience.

1.3 Research Questions

In this thesis, we mainly address the following fundamental research question:

> **How does gamification affect motivation in behavior change support contexts, what effect do personal factors have and how can we utilize this knowledge to increase motivation in gamified systems?**

This fundamental research question is centered around the goal to inform the design of gamified systems for behavior change in order to make these systems more enjoyable and motivating. We will investigate this fundamental research question by breaking it down into the following main research questions, which this thesis sets out to answer:

RQ1 How does gamification affect motivation and related behavioral as well as psychological measures in behavior change contexts?

RQ2 How do personal factors affect the perception of gamification elements?

RQ3 What effects does the personalization of gamified systems have on behavioral and psychological outcomes?

RQ4 How can we unobtrusively assess personal factors to personalize gamified systems without disturbing the gameful experience?

These research questions are based on theoretical assumptions of SDT as well as empirical results from the field: In Basic Psychological Needs theory ("BPNT", see Section 2.1.2) we learn that the three basic psychological needs are essential for intrinsic motivation and organismic integration. Thus, if gamification is capable of satisfying these needs, it should add positively to intrinsic motivation and should be able to support organismic integration by shifting external regulation to the more autonomous regulatory styles (identification, integration). From both BPNT and Organismic Integration Theory ("OIT", see Section 2.1.2), we know that the social-contextual environment affects need satisfaction and has a direct impact on internalization processes, which is why investigating gamification in specific contexts is important (**RQ1**). The need to research gamification in a broader range of contexts was also pointed out in a recent literature review by Koivisto and Hamari [186] and was supported by Nacke et al. [239].

Moreover, from Cognitive Evaluation Theory ("CET", see Section 2.1.2) we learn that intrinsic motivation can be undermined through external stimuli, such as gamification elements, when these external stimuli are perceived as controlling. Since the perception of such external stimuli is heavily dependent on the motivational orientation of an individual and thus differs across different users according to Causality Orientations Theory ("COT", see Section 2.1.2), investigating personal factors that may have an impact on the perception and effectiveness of gamification elements is equally important (**RQ2, RQ3**). While **RQ2** is more focused on finding theoretical evidence for the moderating role of certain factors in the *perception* of gamification elements, **RQ3** is centered around the *actual effects* of personalization of gamified systems based on these factors on both behavioral and psychological outcomes. **RQ2** is important, since it provides answers to what moderates the success of gamification, contributing directly toward answering the open question of what constitutes gameful experiences [186].

Although different factors have already been shown to have an impact on how certain gamification elements are perceived, the research on actual effects of personalized gamified systems is sparse, lacking empirical evidence of the actual effectiveness of personalized gamification [182], which underlines the relevance of **RQ3**. With **RQ2** and **RQ3**, we follow both the demand for gaining more theoretical evidence of which factors moderate how gamification is perceived and the demand to investigate these theoretical findings empirically in practice [315]. Since utilizing certain factors that are useful to personalize gamified systems require users to, for example, fill out extensive questionnaires [342], which may detrimentally affect the gameful experience, we also seek ways to assess such factors less obtrusively (**RQ4**). In addition, since the underlying idea of unobtrusively measuring factors that are relevant for personalization relies on the behavior of users, we expect that investigating **RQ4** also adds insights relevant to **RQ3**. While **RQ4** is less important to advance the theoretical knowledge in gamification research, is it essential to utilize the theoretically grounded factors in practice, which were empirically investigated as part of **RQ3**.

1.4 Thesis Outline

The remainder of this thesis is structured as follows. Chapter 2 will introduce important theories and models which have been used as a foundation for the research that was conducted. We will also present related work from the field and situate our research questions and findings within past research.

In Chapter 3, we investigate the effectiveness of gamification and its effects on need satisfaction and motivation in behavior change contexts. We demonstrate that gamification is capable of satisfying basic needs leading to higher enjoyment and intrinsic motivation. Moreover, we show that gamification affects the behavior of users and other psychological states related to a positive user experience. Besides showing these beneficial outcomes, we also shed light on the role of the context regarding our findings and to what extent context-inherent need frustrations may demand specific gamification concepts, contributing to **RQ1**.

In Chapter 4, we shift our focus from the role of the context towards personal factors that have an influence on how certain gamification elements are perceived. We investigate to what extent age is important to consider and show that preferences for gamification elements change in old age. Moreover, we show that behavioral intentions, i.e. to what extent behavior has already been internalized, affects how certain gamification elements are perceived. Next, we focus on the Hexad user typology and its impact on the perception of gamification elements. We demonstrate that Hexad user types indeed have a moderating effect on the perception of gamification elements and that combining Hexad user types and behavioral intentions is a promising concept to personalize gamified systems. Overall, the findings from Chapter 4 contribute insights relevant to **RQ2**.

In Chapter 5 we investigate whether using the theoretical knowledge obtained in Chapter 4 to build and implement personalized gamified systems actually has an effect on the behavior and motivation of users. We demonstrate that Hexad user types and behavioral intentions have a moderating influence on affective experiences and motivation in the context of physical activity. Moreover, as part of a controlled lab study, we show that a system which adapts its set of gamification elements dynamically to the users' Hexad type increases flow experiences and intrinsic motivation as well as positively valenced affective experiences. Lastly, in this chapter, we show that providing users suitable gamification elements according to their user type affects not only psychological outcomes, but also their behavior. In a longitudinal in-the-wild study over almost two years, our results suggest that offering personalized gamification elements has a more lasting impact on the behavior of users, compared to offering non-suitable ones. Overall, these insights contribute to **RQ3**.

Chapter 6 is focused on approaches that allow us to assess factors relevant to the personalization of gamified systems without disturbing the gameful experience of these systems. For instance, the Hexad user type is traditionally assessed by asking users to fill out a 24-item questionnaire, which might break the immersion of gamified systems and affect their gameful experience detrimentally. Therefore, we investigate whether Hexad user types can be predicted based on the user's personal smartphone data. We found first evidence that this approach might indeed allow user types to be assessed unobtrusively, to a certain extent. Moreover, we demonstrate that Hexad user types can be assessed in a gameful way and can be predicted based on the interaction patterns in gameful applications. This opens the door for the personalization of gamified systems in practice. While these insights contribute mainly to **RQ4**, we also obtain knowledge about the relationship between interaction with gamification elements and Hexad user types, which contributes toward **RQ3**.

Lastly, in Chapter 7, we provide a general conclusion of our findings and contributions. We discuss the fundamental research question in regard to our findings and, based on this, propose a conceptual framework to guide the successful use of gamification to enhance motivation. The consideration of the usage context and its inherent need frustrations, as well as accounting for individual differences in motivational orientations, are the core pillars of this framework. Finally, we outline opportunities for future work and ways to build upon this thesis.

Chapter 2
Background and Related Work

In this chapter, we will introduce important models, frameworks and theories playing a major role for our research questions, hypotheses, and study methodologies, and for interpreting our findings. Due to the integral role of SDT for this thesis, we will discuss the connection of other models and theories to SDT. Afterwards, related works from fields that are relevant to each established research question will be presented, their implications for this thesis will be summarized, and their findings will be used to discuss and frame our contributions.

2.1 Understanding *Why* We Act: Human Motivation

What causes us to act as we do and why do we pursue goals that require us to invest a lot of effort, time, and costs? These and similar questions concern the core of understanding human behavior and have been related back to a concept called *motivation*. Motivation comes etymologically from the latin word *movere*, meaning *to move*. Thus, motivation can be broadly defined as the cause of movement or action, although it is hard to find a consensus on the definition in the literature. While some definitions are phenomenological (focusing on cognitive processes and experiences), others are more physiological (based on internal biological processes), energetic (seeing motivation as an energizing force) or functional (implying a certain goal-directed behavior), just to name a few [180]. This wide range of viewpoints originates from the equally wide range of concepts and theories that were established around the term in the past, as will be summarized briefly in the following section.

2.1.1 A Brief History of Motivation

The effort to explain why human beings act in certain ways, and what the underlying reasons are, was a subject of early philosophical works in ancient Greece [282]. Plato (428–348 B.C.), for instance, postulated the soul to be trifold and hierarchically ordered [282]. *Appetite*, being located in the abdomen [327], was considered the most primitive level, reflecting bodily needs such as thirst or hunger [282]. Above that, *courage* was considered to be located in the chest [327], being responsible for competitive aspects and feelings induced by the social surroundings, such as feeling ashamed or honored [282]. At the most sophisticated level, Plato considered *reason* to be in the head [327], representing rational processes such as decision-making and reflecting [282]. Separating human behavior into rational versus non-rational determinants, i.e. *instinct* and *reason*, was not only the main concept behind Plato's theories, but also the main idea behind the theories of other ancient philosophers such as Socrates (470–399 B.C.) [113] and Aristotle (384–322 B.C.) [327], forming the foundation for *mind-body dualism* [327].

Thousands of years later, this mind-body dualism was taken up by René Descartes (1596–1650) [312]. To explain motivation and behavior, Descartes distinguished between the *body*, which they considered passive and mechanical, responding to the environment through physiology, and the *will*, which they considered the active source of motivation, being capable of controlling the body and regulating its aspirations. Descartes assumed that the will interacts with the rather passive, bodily aspects of motivation through mechanical movements of the pineal gland. This assumption – that the will is translated to actions of the human body through physiology – and Descartes' idea to understand human motivation by understanding the will, marks a turning point: While understanding the will turned out to be similarly complex as understanding motivation from a philosophical standpoint, the idea that bodily actions and needs are regulated by the will through physiology led to subsequent research focusing on physiological and biological aspects rather than on philosophical ones [282].

Charles Darwin's theory of evolution can be seen as part of this development. For motivation research, Darwin's conceptualization of *instincts* is relevant: They considered instinct as a behavioral adaptation to the environment, as they found that animal behavior is not entirely based on experience but seems to be partially unlearned [282, 327]. In the new scientific field of psychology, which can be attributed to Wilhelm Wundt based on the work *Principles of Physiological Psychology* in 1873 [312], William James was the first scientist building upon Darwin's notion of instincts to explain human behavior [282]. James lists a number of instincts such as sucking, locomotion, imitation, hunting, fear or cleanliness and argues that there must be a certain stimulus, an impulsive root, that activates these instincts and translates them into behavior [164]. McDougall [220] regarded instincts as "prime movers of all human activity" (p. 41), making them the central cause of human behavior. They defined instincts as inherited or innate psychophysical dispositions that direct attention, excitement and actions towards a specific goal. They were convinced that without instincts, "the organism would become incapable of activity of any kind" (p. 42). Although there are many parallels to James' theory of instincts as causes for behavior, attributing all actions solely to instincts marks the major difference between them [282]. This focus on instincts as the only causes for behavior quickly became problematic, as researchers aimed to determine how many instincts existed, leading to lists with over 6,000 entities [282]. Also, criticism regarding the missing explanatory aspect of this theory emerged, since instincts do not explain behavior but rather serve as labels [282]. Ultimately, the theory of instincts was dropped [327].

Woodworth, who was convinced that psychology's goal should be to understand why people behave as they do [312], introduced the motivational concept of *drive* [364]. According to this theory, behaviors consist of two parts: a mechanism (how an action is done) and a drive (what induces the action). Its general assumption was that drive can be seen as "the power applied to make the mechanism go" (p. 37), i.e. drive was assumed to energize behavior. This conceptualization of drive as a form of energy was meant in a biological sense: Drive was seen as an energy that an organism releases, given a stimulus, to start mechanisms to accomplish actions [364].

The concept of drive was adapted by both Freud's and Hull's drive theories. Freud's drive theory [128] consists of four components. The *source* of drive comes from a bodily deficit, leading to an increased energy level in the nervous system causing bodily discomfort and anxiety (*impetus*). To reduce the bodily discomfort, the person seeks an *object*, which is utilized to fulfill the drive's *aim* and ensure bodily comfort. However, Freud's drive theory suffered from several issues including its weak empirical basis, its inability to make predictions and thus its lack of falsifiability, and its overestimation of biological aspects [282]. In contrast, Hull's drive theory was empirically grounded and stemmed from experimental research [282]. According to Hull [159], behavior is influenced by both the environment and the organism and follows – since both are part of nature – natural laws. Consequently, Hull considered drive to have a purely physiological basis. Drive, according to this theory, increased depending on the amount of time an organism is exposed to bodily deficits. As such, drive is a direct consequence of bodily needs, such as the need for food or water, the need for air, the need to avoid pain, the need for sleep, or the need for activity, and becomes stronger the longer an organism is exposed to a deficit in one (or more) of such needs. This definition of drive was particularly important, as it allowed making predictions of behavior and thus was falsifiable. However, Hull postulates that drive has no direction, i.e. it energizes behavior but does not guide it. Instead of drive, habit was considered to direct the drive into a certain direction. Based on Hull's experiments, habit strength was assumed to be formed by learning, and learning occurs through drive reduction: When an organism executes a response that removes a bodily deficit, the resulting drive reduction is a reinforcement for the corresponding behavior, and learning occurs. Consequently, Hull defined the strength of behavior as the product of habit times drive. Since this formula was not able to explain motivational forces outside the organism that are not caused by bodily deficits, Hull extended it in their later works [160]. They introduced another motivational concept besides drive which was called incentive motivation. Incentive motivation represents external stimuli through objects of value that are capable, similar to drive, of energizing organisms. As a consequence, they updated their formula such that the strength of behavior was a product of habit, drive and incentive motivation. Despite this adaptation of the formula, drive theory was not able to explain human behavior holistically and thus declined in relevance [282]. For instance, motivation was shown to be rooted in factors beyond bodily needs: People eat not only because of hunger, but maybe because they are curious about the taste of food [282]. Also, learning was shown to occur without drive reduction [282].

In contrast to Hull, Freud and Woodworth, Skinner refused to speculate about motivational forces inside the organism and focused solely on observable causes of behavior and the study of responses [312]. Skinner was not trying to explain behavior but rather to describe it through controlled experiments involving a stimulus and studying the response [312]. They believed that analyzing the causes of an action and its consequences is the best way to understand behavior and called this approach *operant conditioning* [320]. In contrast to what Skinner called respondent conditioning, in operant conditioning, no observable external stimulus is available; the stimulus appears spontaneously. While in respondent conditioning, an observable stimulus is presented to the organism, which reinforces the organism's action, the stimulus is not observable in operant conditioning, and operates on the organism's environment. In Skinner's experiments, they used an apparatus called a "Skinner box": a box in which rats were placed and where a lever was installed, which – once it has been triggered – led to food being dispensed. Instead of rewarding the rat after a desired behavior was performed, the rat learned what the desired behavior (pressing the lever) was by spontaneously performing it and thus operating on the environment. Based on these experiments, Skinner introduced the law of acquisition, stating that a positively reinforced stimulus increases the strength of operant behavior. According to Skinner, operant conditioning is the cause for learning and all behaviors are learned [312]. Consequently, all behaviors can be explained through past reinforcements [312]. However, this very assertion was challenged by other researchers, showing that the instinct of some animals was sometimes stronger than the reinforced behavior, resulting in unwanted actions [312]. Also, some behavior was found to be inherited, which speaks against operant behavior as the only source for action [312]. Furthermore, explorative behaviors of infants or people engaging in sports, artistic or other leisure-time activities do not seem to be reinforced by environmental stimuli [299]. In sum, there is a considerable amount of evidence speaking against the assumption that all behavior can be attributed to reinforcement and conditioning [299].

Maslow was not satisfied with the idea that human beings are solely passive organisms being moved by external stimuli as supposed by Skinner, nor were they

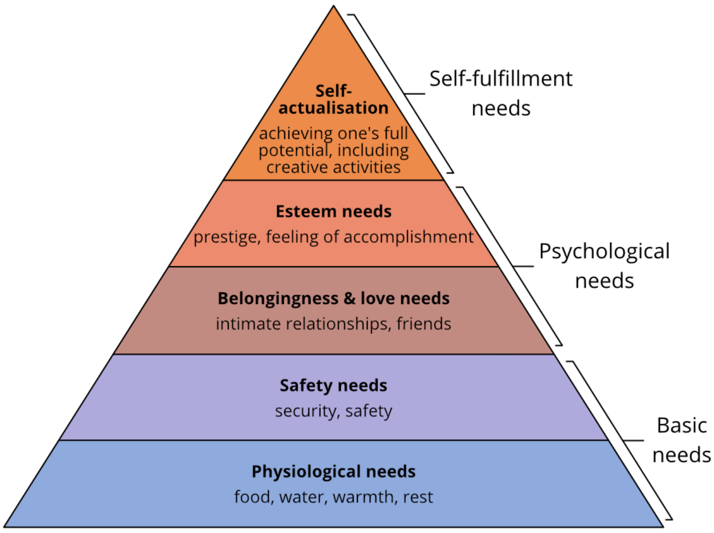


Figure 2.1: Maslow's Hierarchy of Needs, "Creative Commons Maslow's Hierarchy of Needs" by Androidmarsexpress is licensed under CC BY-SA 4.0

satisfied with the idea of behavior as a way of restoring biological equilibrium as was postulated by drive theorists [215]. They were convinced that behavior typically has more than one motivation, that one should be human-centric when aiming at explaining human behavior and that any motivated behavior should be interpreted as a channel through which basic needs are expressed. Furthermore, they considered human beings as active organisms, who strive towards *self-actualization*. Maslow suggested that all human beings are driven by *basic needs* which must be satisfied for their survival, health and growth towards self-actualization. They assumed that these needs are hierarchically ordered such that the needs at the bottom of the hierarchy are basic needs important for survival. Once satisfied, they become less important and the needs of the next higher level become prominent. This hierarchy, according to Maslow, consists of five levels (see Figure 2.1). At the lowest level, they propose *physiological needs*, which represent vital needs such as those for food, water, air, and so on. Once satisfied, *safety needs* become more prominent, representing needs for personal security, property, health or employment. On the third level, the need for *love and belonging*, including friendship, family, a sense of connection or intimacy is introduced, followed by the *esteem needs* such as reputation, recognition, strength or self esteem on the fourth level. On the highest level, Maslow proposed *self-actualization*, which is considered as the need for self-improvement, self-fulfillment or to become the best version of oneself. Maslow noted that the emergence of self-actualization "rests upon prior satisfaction of the physiological, safety, love and esteem needs" (p. 381) and thus strengthens the assumption of the hierarchical arrangement of basic needs. However, Maslow's theory did not meet the rigors of scientific research, which they themselves admitted, and had weak empirical support [312].

Also, the assumption that needs of a lower hierarchical level must be satisfied before needs of a higher level become prominent was frequently challenged by examples of human behavior that cannot readily be explained by this assumption, such as people staying up all night (and thus neglecting the physiological need for sleep) to finish work such as an essay, or people risking their safety to experience self-actualization [299].

To sum up, this brief historical overview of theories aiming to explain human motivation, which is far from being complete, demonstrates the high complexity of motivation. It also suggests that relying on one single *grand theory* to explain the causes of human behavior is impossible. Therefore, to cover the broad range of human motivation, research more and more refrains from single grand theories and rather focuses on *mini-theories* to explain specific facets of motivation [282]. One contemporary umbrella for such a set of mini-theories is Self-Determination Theory, which is introduced in the next section.

2.1.2 Self-Determination Theory

Self-Determination Theory ("SDT") [299] is an empirically-based macro-theory of human motivation and wellbeing, focused primarily at the psychological level. It assumes that organisms are inherently active and strive for psychological growth, engagement and wellness, and it focuses on which conditions support or thwart these processes. SDT is particularly concerned with *intrinsic motivation*, which refers to the type of motivation that energizes behaviors which are interesting, enjoyable and satisfying in themselves, and *organismic integration*, i.e. the process of integrating and internalizing extrinsic motives into the developing self, such that originally external behavioral regulations shift (or do not shift) towards internally perceived behavioral regulations. On a meta-level, SDT posits that intrinsic motivation and internalization processes as part of organismic integration are energized by satisfying three *basic psychological needs* – the needs for *autonomy*, *competence* and *relatedness*. Moreover, it states that motivation for a certain activity is not a uni-dimensional concept but rather that people are motivated by intrinsic motivation or several types of extrinsic motivation (differing in their degree of internalization), or are amotivated. While intrinsic motivation is characterized by inherently interesting or enjoyable activities, *extrinsic motivation* refers to activities induced by external sources, such as external rewards. *Amotivation* is characterized by a state in which there is no intentional motivation for any action. Out of these types of motivation, intrinsic motivation is the most autonomous form, which has been shown to be positively related to health, well-being and enjoyment. SDT is organized into six mini-theories, which are concerned with different aspects of motivation and related processes.

SDT has received substantial empirical support, has been used across a wide range of contexts and has been widely applied in the games research community to analyze the player experience, inform game design and create gameful systems that motivate people by using gameful design in non-game contexts [350]. Furthermore, SDT is considered as the most frequently used theory in gamification research [239]. In the following, we will introduce four of its six mini-theories, since these mini-theories will be used throughout this thesis to inform, guide, discuss and interpret our approaches and findings, thus providing a frame to the research that we conducted. SDT is particularly relevant for the research questions presented in Section 1.3 and serves as a theoretical basis for them, as will be explained in the following sections. We will omit goal contents theory, which is more concerned with general life goals and their effects on well-being and thus less helpful to explain motivation in specific (gamification) contexts. Also, we leave out relationship motivation theory, since it focuses on friendships and the quality of close relationships, which is less relevant for our research questions. The following description of SDT's mini-theories is based on [299].

Cognitive Evaluation Theory

Cognitive Evaluation Theory ("CET") is primarily concerned with *intrinsic motivation* and the social-contextual factors which support or undermine intrinsic motivation towards an activity. The theory was the first of SDT's mini-theories and can be seen as the starting point of SDT. It is rooted in first experiments by Deci [101], in which the effects of external rewards on intrinsic motivation were investigated in a series of three experiments. The first two studies found that when money was administered as a compensation for working on an interesting task, intrinsic motivation was lowered. On the other hand, the third experiment revealed that positive verbal feedback did not harm, but rather supported intrinsic motivation. CET explains these findings through a concept called *perceived locus of causality*. An *internal perceived locus of causality* means that actions are truly self-endorsed and are perceived as originating from oneself. An *external perceived locus of causality*, on the other hand, refers to actions which one feels orchestrated to perform by external regulators. CET proclaims that extrinsic rewards *may* lead to a change in the perceived locus of causality from internal to external, undermining the basic psychological need for autonomy. Whether extrinsic rewards undermine intrinsic motivation or support it is dependent on the *functional significance*, i.e. whether external rewards are perceived as *informational*, *controlling* or *amotivating*. When external rewards are perceived as *informational*, they can support the feeling of competence and thus have positive effects on intrinsic motivation. However, perceiving external rewards as supporting one's own competence can only occur when one's autonomy is not undermined. Thus, when rewards are perceived as *controlling* one's behavior, the need for autonomy is thwarted and intrinsic motivation is reduced. Lastly, extrinsic rewards that are perceived as impossible to reach, that signify incompetence or that are perceived as not valuable lead to *amotivation*, a state in which both intrinsic and extrinsic motivation is undermined.

As key take-aways from CET, we remark that intrinsic motivation can be undermined by external events relevant to the regulation of behavior, when they are perceived as controlling. In contrast, intrinsic motivation can be facilitated by external events which are perceived as informative and competence-supportive, e.g. through positive feedback. Lastly, external events that make subjects feel incompetent to perform actions or that are perceived as not valuable can lead to amotivation. Thus, for gamified systems, it is essential that the rewards and feedback mechanisms are perceived as informative instead of controlling, to positively affect the motivational impact on the user.

Basic Psychological Needs Theory

Basic psychological needs theory ("BPNT") proposes three basic psychological needs that are essential for driving humans', intrinsic motivation and organismic integration processes. It is concerned with the relationship between these basic need satisfactions and frustrations and the effects on well-being or ill-being. The first of these basic needs is *autonomy*, described as the need to regulate one's own actions. It refers to feeling completely volitional and having a sense that actions are self-endorsed and congruent with one's true inner values and interests. *Competence* refers to the basic need to feel in control of the outcome of an action, experience mastery and feel that one's own actions are effective. Lastly, *relatedness* refers to a sense of social belonging and reciprocal care. While this need is typically satisfied when having a feeling of being cared for by others, having the feeling of giving or contributing to others is equally important. BPNT also states that the satisfaction (or thwarting) of these basic psychological needs is dependent on social contextual conditions. As a consequence, contexts may support as well as fail to nourish each of the three basic needs, resulting in satisfactions and frustrations of them respectively, leading to supportive and detrimental motivational experiences. As a consequence, contexts in which certain needs are supported or thwarted play a major role in motivational experiences, and introducing certain measures in order to support psychological needs which are thwarted by the context may add positively to well-being and motivation.

To sum up, we may derive two major implications. First, there are three basic psychological needs – autonomy, competence and relatedness – which energize intrinsic motivation as well as organismic integration processes and are essential nutriments for well-being. Second, we learned that contextual factors can support or diminish basic needs' satisfaction and thus play a major role in the study of motivational experiences. Therefore, for gamified systems, it is important to support basic psychological needs, to facilitate more autonomous forms of motivation and to better understand how motivation in gamified systems is affected by the context or environment in which a gamified system is established.

Organismic Integration Theory

While CET is concerned with intrinsic motivation, Organismic Integration Theory (OIT) focuses on extrinsic motivation and the process of internalization, in which externally regulated behaviors can be integrated to the self. It differentiates extrinsic motivation on a continuum ranging from a fully external perceived locus of causality to an internal perceived locus of causality, and defines four types of extrinsic motivation that differ in how behavior is regulated. *External regulation* is the least autonomous form, being under full control of external factors or events. This type of regulation is characterized by the feeling that the reason for doing an activity is solely due to external contingencies, e.g. to obtain a reward or to avoid punishment. *Introjected regulation* refers to a regulatory style in which the aforementioned dependencies on external stimuli are decoupled to a substantial degree, but have not been completely accepted as part of the self. The driving forces for engaging in activities are intrapersonal, i.e. they come from controlling contingencies from within the individual. Behaviors which are motivated by introjected regulation are performed to avoid feelings of guilt or achieve feelings of approval. In contrast, *identified regulation* is characterized by increased feelings of autonomy towards a behavior due to the fact that the behavior and its regulation are consciously valued and perceived as personally important. Here, although the perceived locus of causality shifts towards internal, the activity itself has not been set in relation to one's own identity and thus is not yet fully internalized. The most autonomous type of extrinsic motivation which has an internal perceived locus of causality is *integrated regulation*. This type of motivation is characterized by having brought activities fully in congruence with one's values, needs and other aspects of one's self. Figure 2.2 provides an overview of the different motivations and regulatory styles within OIT.

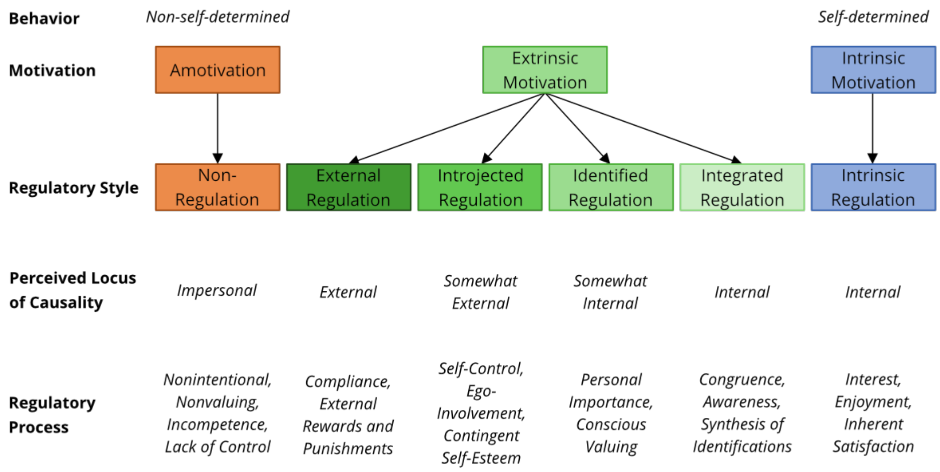


Figure 2.2: Motivations and regulatory styles as described by Organismic Integration Theory. Adapted from [299] (p. 192).

The process of internalization, i.e. integrating external events to oneself, is energized by satisfying basic psychological needs. This means that contexts which support the three basic needs when performing an action can facilitate internalization, while contexts which are controlling or do not support relatedness or competence needs will less likely lead to identification or integrated regulation. CET also notes that individuals do not necessarily move through all different types of regulatory styles when internalizing behaviors, nor is this process unidirectional, as the type of extrinsic motivation might change due to changes in the social-contextual environment. Thus, we learn that extrinsically motivated behavior involves different regulatory styles that differ in their relative autonomy and perceived locus of causality. Also, we saw that satisfactions of basic psychological needs facilitate one's internalization of a behavior into oneself and that the social-contextual environment and its effects on need satisfaction have an impact on the internalization of extrinsically motivated behavior. For gamified systems, these findings again emphasize the importance of supporting basic psychological needs and the need to study gamification in different contexts to better understand how gamification affects motivation.

Causality Orientations Theory

While CET and OIT describe *social-contextual factors* influencing intrinsic motivation as well as the process of internalization of extrinsic motivation, Causality Orientations Theory ("COT") describes global-level *individual differences* in motivational orientations. The theory suggests that people differ in their *causality orientations*, which are labeled *autonomous*, *controlled* and *impersonal*. It is important to note that people rarely have only one of the three orientations, but rather have some degree of each of them. Depending on their causality orientation, people give different functional significance to contexts. This is an important fact when reconsidering that CET states that it is the functional significance that makes a difference in whether external stimuli are perceived as informational, controlling or amotivating. Therefore, individual differences in causality orientations can have a direct impact on whether external stimuli support or undermine intrinsic motivation. When people are high in the autonomy orientation, they lean towards higher levels of intrinsic motivation and self-regulated behaviors, are taking interest, and usually use the identified and integrated regulatory styles in internalization processes. Autonomy-oriented people usually give informational functional significance to contexts, which means that they tend to give informational meaning to external stimuli that support basic needs satisfaction. Control-oriented people, on the other hand, are usually more receptive to external contingencies, tend to use external or introjected regulatory styles and usually have low intrinsic motivation. They lean towards giving controlling functional significance to contexts, which means that they tend to perceive external stimuli as controlling their behavior. This undermines their autonomy and thus decreases their intrinsic motivation. Lastly, impersonal-oriented people tend to see many obstacles in their way, lean towards feelings of incompetence and perceive a lack of control over behavioral outcomes. This results in a lack of intentionality and passiveness, and ultimately leads to amotivation.

COT further states that the individual differences in motivational orientations are the result of need satisfaction or frustration over time, i.e. that individuals learn or develop their own approaches to regulate their behaviors, based on whether they lived in a need-supportive and autonomous or in a need-thwarting and controlling environment. This means, for instance, that people being primarily exposed to controlling environments usually develop controlled motivational orientations, which consequently leads to interpreting new environments and stimuli as being controlling. Lastly, COT also notes that although motivational orientations are context-independent, certain contextual factors might trigger certain motivational orientations, which means that both contextual factors as well as motivational orientations should be considered to understand an individual's motivation in a given context. To sum up, we learn that there are global-level individual differences in motivational orientations, which have a direct influence on how contexts and external stimuli are perceived and consequently, whether they support or thwart intrinsic motivation. Also, it should be noted that to understand the type of motivation and the underlying regulatory processes, it is important to consider contextual factors and individual differences. For studying gamification, this implies that it is essential to study the individual person and the inter-personal differences in how gamification affects the motivation of users. Since the same feedback mechanisms and rewards used in gamified systems might be perceived as controlling by more control-oriented persons and informative by more autonomy-oriented persons, personalization is an important topic for gamification research. Consequently, the factors of importance, effects of personalization, and ways to elicit personalization factors are important research topics to consider.

2.1.3 Summary

We began this section by providing a brief overview of how motivation research evolved throughout history, leading to the conclusion that so-called "grand theories" are hard to formulate due to the complex nature of motivation. Therefore, research focused on mini-theories such as SDT, which was introduced afterwards. We described the integral role of SDT for this thesis and presented the most important mini-theories of it, which play an essential role for our research questions. That contextual factors can have an effect on need satisfaction and internalization processes, which we learned from both OIT and BPNT, underlines the importance of studying gamification in different contexts and understanding its effect on need satisfaction (**RQ1**). Considering CET, stating that intrinsic motivation can be undermined when external stimuli are perceived as controlling or amotivating, and COT, stating that the perception of such stimuli depends on motivational orientations tied to the person, studying personal factors (**RQ2, RQ3**) is integral to understand motivation in gamification. Consequently, **RQ1 – RQ3** are both grounded in empirical findings in the field (as described in Section 1.3) but also theoretically grounded due to their tight coupling to SDT. **RQ4** is more focused on gamification practice, i.e. allowing to disseminate the findings of this thesis, and therefore more implicitly based on SDT.

2.2 Theories Related to Motivation

In this section, we will introduce important theories and models that have been used in addition to SDT to guide the research conducted in this thesis as well as explain and interpret corresponding findings. We will also elaborate on the relationship of these models and theories to SDT.

2.2.1 Transtheoretical Model of Health Behavior Change

The Transtheoretical Model of Health Behavior Change [275, 276] ("TTM") is an integrative theory explaining the processes involved in health behavior change. It was developed by analyzing and synthesizing prominent theories of psychotherapy and behavior change, which is why it is labelled "transtheoretical". At its core, it proclaims that behavior change has a temporal dimension, i.e. that behavior change is a "a process involving progress through a series of six stages" ([276], p. 2), the so-called *stages of change*. Through identifying ten *processes of change*, the model also provides activities that people use to progress through these stages of change. Furthermore, it acknowledges that individuals weigh the pros and cons of changing their behavior as part of the *decisional balance* construct. It also states that *self-efficacy*, i.e the situation-specific confidence people have in being able to perform the target behavior, is a core pillar of behavior change. Lastly, the construct of *temptation* refers to the "intensity of urges to engage in a specific habit when in the midst of difficult situations" ([276], p. 3). We will elaborate further on these core constructs in the following. Also, we will provide insights on the relationship to SDT. TTM is relevant for the studies described in Section 4.3, Section 4.5, and Section 5.2.

Stages of Change

The stages of change construct describes the process of health behavior change by establishing a temporal dimension. At its core, the underlying assumption is that behavior change occurs over time and is seen as a process rather than an event. The stages of change construct states that behavior change involves progress through six stages. Since the sixth stage, termination, "may not be a practical reality for a majority of people" ([276], p. 2), and is less suitable in the domain of physical activity or weight control, only five stages of change were frequently assessed and investigated in past research [276]. The stage of change construct integrates attitudinal and behavioral changes, i.e. while the first two stages focus on the attitude to change, the third stage focuses on attitudes and more concrete intentions to change, whereas later stages put their focus more on the behavior as such. These stages are characterized as follows:

Precontemplation: People in this stage of change have no intention to change their behavior in the foreseeable future (which usually means in the next six months). As potential reasons for why people are in this stage, Prochaska and Velicer [276] mention a lack of information on the consequences of a certain behavior, or that people might have tried to change their behavior but failed, leading to them questioning their ability to change and becoming demoralized. People in this group usually avoid thinking or having conversations about their behavior and are frequently described as resistant or unmotivated.

Contemplation: In contrast to the precontemplation stage, people in this stage have intentions to change their behavior in the foreseeable future (usually measured as the next six months). They are more aware of the consequences of changing their behavior and weigh the pros and cons. This inner conflict between the advantages and drawbacks can lead to a state of ambivalence. Consequently, people might stay within this stage for a long time. This state is also referred to as "behavioral procrastination" ([276], p. 2).

Preparation: People in this stage have more concrete intentions to change their behavior in the immediate future (usually measured as the next month). They have a plan in mind, which they aim to execute (e.g. going to the gym, buying a self-help book or consulting a doctor). People in this stage are willing to change and are ready to take action.

Action: This stage is characterized by a concrete change of observable behavior, in contrast to changes in attitude or intentions. People in this stage have made modifications to their behavior, usually in the last six months. This is the first stage in the process of behavior change where observable outcomes can be detected. For example, people intending to do more sports go running twice a week; people intending to stop smoking have not had a cigarette for the last three months or people intending to lose weight have reduced their calorie intake from fat considerably for a certain amount of time.

Maintenance: People in this stage have been able to maintain a certain behavior usually for more than six months and are working to prevent relapse to an earlier stage. People intend to maintain the behavior change and have to put less effort in actively controlling their behavior, i.e. change processes (activities that people use to cope with progressing through the stages) are not applied as frequently as they are applied in the action stage.

Termination: People in this stage have zero temptation to relapse to an earlier stage and have gained 100% self-efficacy in the new behavior. Even when confronted with emotionally challenging situations, people will not return to their old behavior. This stage is regarded as unrealistic in many health-related areas such as exercise or weight control. In such health contexts, a “realistic goal may be a lifetime of maintenance” ([276], p. 2). Therefore, this stage has not been investigated in depth, compared to the other five stages mentioned above.

Besides on the aforementioned stages, the TTM also introduces the concept of relapse. Relapse refers to the return to an earlier stage, mostly from the action or maintenance stage. According to Prochaska and Velicer [276], relapse happens very frequently, but usually a majority of people relapse to a stage in which the intention to change is still persistent, instead of relapsing to precontemplation. It should also be noted that, although the stages of change are presented in a linear fashion, people may also progress non-linearly through these stages. People may re-cycle through the stages or relapse to earlier stages. Also, the stages of change themselves have been changed over time: While there used to be only five stages of change in which termination was replaced by relapse and preparation was not included [275], the authors clarified that relapse should not be seen as a stage on its own and introduced the preparation and termination stages [276]. Instruments to asses the stage of change of users have been developed for a broad range of health contexts including healthy eating [25] and physical activity [212].

Processes of Change

Processes of change refer to the activities being used by people to progress through the stages of change. Based on empirical research, ten such processes have been identified and received support. *Consciousness raising* is a process enhancing awareness about the causes and consequences of a behavior. Education and feedback are means to enhance awareness, for example. *Dramatic relief* is a process in which strong emotional experiences are induced (e.g. through media campaigns) which may lead to relief when action is taken. *Self-reevaluation* involves reflections on oneself and assessing one’s self-image (e.g. as a couch potato). In contrast, *environmental reevaluation* focuses on the assessing and reflecting on the effect of one’s own behavior on the environment. Also, becoming aware of the fact that one could serve as a role model for others and what impact this has on the behavior of other people might be part of this process. *Self-liberation* is a process in which people believe that they can change and commit to this belief. New year’s resolutions are an example of a self-liberation process. *Social liberation* refers to social means of supporting people’s health, or providing alternatives allowing people to engage in healthy behaviors. Examples include smoke-free zones or salad bars in a restaurant. *Counterconditioning* is a process of replacing unhealthy behaviors with healthier ones. Such healthier behaviors may serve as a substitute for unhealthy ones, e.g. using nicotine patches instead of smoking cigarettes. *Stimulus control* is a process focused on the stimuli triggering unhealthy behavior and replacing them with stimuli supporting healthier behaviors. Parking at a further walking distance from one’s workplace might be one way of adding a stimulus to support healthier behavior. *Contingency management* is a process of providing consequences to behaviors. Punishments and rewards are means to realize such consequences, although rewards (i.e. positive reinforcement) have been shown to be more effective than punishments. Lastly, *helping relationships* involves caring and feeling cared for by others and social support, e.g. through a therapeutic alliance or buddy systems.

Decisional Balance

This core construct of the TTM concerns a person's weighing of the advantages and disadvantages of changing their behavior. Initially, an eight-factor structure was used to conceptualize the pros and cons. However, in empirical research, a much simpler factor structure has been found, consisting of only two factors representing the pros and cons, respectively. The balance between these pros and cons has been found to vary when people pass through the stages of change in a meta-analysis across 48 health behaviors by Hall and Rossi [142]. The authors found that the cons of changing are more prevalent than the pros in the precontemplation phase. In the intermediate stages, i.e. contemplation and preparation, there seems to be a balance in pros and cons, which may lead to a certain ambivalence about changing the behavior. From the action stage on, the pros outweigh the cons of changing the behavior.

Self-Efficacy

Self-efficacy refers to the confidence people have in their own abilities to cope with (high-risk) situations without relapsing to their old behavior, i.e. how confident people are that they can maintain the new behavior. The construct is adapted from Bandura's self-efficacy theory [32], stating that higher levels of perceived self-efficacy are related to the success of people in adopting new behaviors, i.e. that self-efficacy is positively related to "whether coping behavior will be initiated, how much effort will be expended, and how long it will be sustained in the face of obstacles and aversive experiences" ([32], p. 1).

Temptation

Temptation refers to the urge of relapsing to old behaviors when being confronted with difficult situations. Such situations might be characterized by increased anxiety or emotional distress, but may also be evoked by social peers.

Relationship to Self-Determination Theory

Research has investigated the relationship between TTM and SDT. Mullan and Markland [237] found that the stage of change is related to the process of internalization explained as part of the Organismic Integration Theory of SDT, in the domain of physical activity. More specifically, the authors found that the behavioral regulation of participants in the later stages of change was more self-determined than those in the earlier stages of change (e.g. that people in the action or maintenance stages use more autonomous regulatory styles such as identified or intrinsic regulation). They conclude that the behavioral regulation becomes more self-determined with increasing stage of change. Similarly, Kennedy and Gregoire [174] investigated whether the type of motivation (as defined by SDT) would predict people's stage of change in the context of drug addiction. In line with Mullan and Markland, they found that participants having a more autonomous form of motivation and thus regulation were more likely to be in the higher stages of change than those having less autonomous motivation. These findings are also supported by Kushnir et al. [191] in the domain of gambling. The authors found that higher autonomous forms of motivation predicted greater intentions for change and thus higher stages of change. In contrast, people using more controlling regulatory styles and having more external forms of motivation were less likely to have advanced their stage of change.

2.2.2 Achievement Goal Theory

The concept of achievement goals originates in the independent and collaborative work of several researchers in the late 1970s and early 1980s [118]. Here, the works by Ames [22], Maehr and Nicholls [209], Dweck [112] and Nicholls [250] are considered as fundamentals of achievement goal theory [118]. The overarching understanding of achievement behavior was that "achievement behavior is defined as behavior directed at developing or demonstrating high rather than low ability" [250] and that achievement goals are seen as the purpose for engaging in achievement behavior [118]. The aforementioned works had in common that they established two types of achievement goals, differing in how competence is defined. The overlap of these works was high enough that they could be subsumed as the *dichotomous achievement goal model* [118]. Achievement Goal Theory is relevant for Section 4.3.

Dichotomous Achievement Goal Model

In the dichotomous achievement goal model, two types of achievement goals are distinguished: *mastery* and *performance* goals, which differ in how competence is conceived [118, 250]. The purpose of mastery goals is developing competence and mastery [112]. Thus, mastery goals focus on learning and improving oneself. In contrast, performance goals focus on gaining positive judgments from others [112]. Thus, the purpose of performance goals is demonstrating (rather than obtaining) normative competence [118], often involving social comparison. Both types of goals are conceptualized as *approach* goals in the dichotomous goal model [118], meaning that they focus on approaching success rather than preventing failure (e.g. getting a lot of right answers on an exam rather than avoiding incorrect answers [118]).

Trichotomous Achievement Goal Model

The trichotomous achievement goal model emerged out of the observation that effects of performance goals on intrinsic motivation were inconclusive [116, 118]. Research assumed that performance goals would undermine intrinsic motivation. However, a substantial amount of studies, using a dichotomous achievement goal model (without differentiating performance goals in terms of approach/avoidance), did not find evidence for this assumption. Therefore, Elliot and Harackiewicz [116] extended the dichotomous achievement goal model by adding an approach/avoidance dimension to performance goals. Thus, the trichotomous achievement goal model consists of three goal types: *mastery*, *performance-approach* and *performance-avoidance*. It was found that the performance-approach goal type facilitated intrinsic motivation, similar to mastery goals, while performance-avoidance goals were more likely to undermine intrinsic motivation [116]. However, it was also found that perceived competence plays an important role in the effect of performance goals on intrinsic motivation: Butler [67] found that performance goals are more likely to undermine intrinsic motivation when perceived competence is low.

2x2 Achievement Goal Model

In the 2x2 achievement goal model [117], the mastery goal type is divided into approach and avoidance types, similarly as was done in the trichotomous achievement goal model for the performance goal type. This leads to four types of achievement goals, spread along two dimensions: the *valence* and *definition* dimension. *Mastery-approach* goals are positively valenced (i.e. focused on approaching success) and defined as absolute or intrapersonal. *Mastery-avoidance* goals differ in how competence is valenced, i.e. they focus on avoiding failure (e.g. striving to avoid failing to learn course material, or for people in the later stage of their careers, striving not to do worse than before [117]). The *performance-approach* and the *performance-avoidance* goal types differ on how competence is defined, since they rely on normative regulations rather than intrapersonal or absolute ones.

3x2 Achievement Goal Model

Lastly, in the 3x2 achievement goal model [118], another type of definition for competence is added. In this model, the mastery goal is split into three types differing in their definition of competence: *task-based*, *self-based* and *other-based* goals. In line with the 2x2 achievement goal model, there are two types of how competence is valenced for each definition of goal type: *approach* and *avoidance*. Thus, there are six types of achievement goals, as illustrated in Figure 2.3. Task-based goals focus on attaining task-based competence, i.e. evaluating how well or badly a task was solved, given a predetermined objective. As an example, reaching 5000 steps per day would count as a task-based goal. Self-based goals,

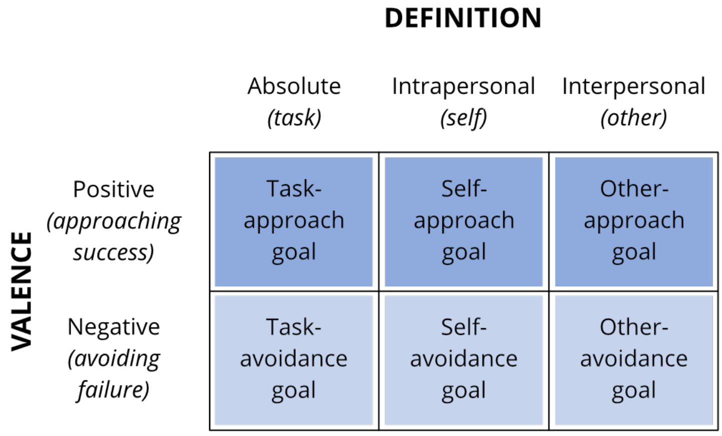


Figure 2.3: Goal types of the 3x2 Achievement Goal Model. Positive and negative refers to how competence may be valenced. Absolute, intrapersonal and interpersonal represent the definitions of competence. Based on [118].

in contrast, focus on one's own person by evaluating the current goal in relation to one's previous performance. For example, reaching a higher daily step count than yesterday would be seen as a self-based goal. Self-based goals focus more on self-improvement instead of task-based competence. While the standard of competence is mastery for both task-based and self-based goals, other-based goals focus on normative performance. With other-based goals, one's own performance is evaluated in relation to the performance of other people. An example of a performance goal would be reaching a higher daily step count than other people.

Task- and self-based goals are similar (which is why they have previously been considered as one, called mastery goals): they both have an "evaluative standard that may be used privately and at one's own discretion in the acquisition of competence information" ([118], p. 2). Thus, the evaluation of both these type of goals is not dependent on the performance of others. However, there are differences in the two types of goals upon closer examination concerning their cognitive representation: Task-based goals are rather simple, as they only require representing a task and evaluating to what degree the task has been accomplished. Also, the standard of competence is inherent in the task and more decoupled from oneself, allowing for immediate feedback during task engagement. In contrast, self-based goals are more complex and require more cognitive capacity to be represented. They require the ability to simultaneously represent two outcomes (the past and the current, not yet available outcome) and compare them across a temporal dimension. While self-based goals might be optimally challenging, since the evaluative standard is defined by oneself, they might also lead to self-worth concerns, since the regulatory focus shifts away from a process orientation in task-based goals to a regulation depending on oneself.

Relationship to Self-Determination Theory

In general, it has been assumed that mastery goals facilitate intrinsic motivation while performance goals undermine intrinsic motivation [116]. For instance, Heyman and Dweck [154] found that mastery goals (called learning goals in their article), which focus on improvement, learning and gaining competence, are associated with enhanced intrinsic motivation. This is explainable by SDT, since aspects like self-improvement might be related to more autonomous regulatory styles and an internally perceived locus of causality. Heyman and Dweck [154] also found that performance goals focusing on the (normative) evaluation of competence are more likely to undermine intrinsic motivation and may lead to attributing failure to one's incompetence and anxiety. However, the assumption that performance goals generally affect intrinsic motivation negatively was not empirically supported: It was found that less than half of the experiments on the relationship between performance goals and intrinsic motivation supported the assumption that performance goals undermine intrinsic motivation [116]. Instead, Elliot and Harackiewicz [116] found that the valence of performance goals has a substantial impact in this regard. Performance-avoidance goals in particular were associated with negative effects on intrinsic motivation. In addition, Butler [67] emphasized that perceived competence plays a major role in the effect of performance goals on intrinsic motivation. When perceived competence is low, intrinsic motivation is more likely undermined.

Ultimately, Gillet et al. [136] found more recently that the type of motivation underlying achievement goals has a major impact on the effects of such goals on affective and behavioral outcomes. They were able to show that the reasons for engaging in achievement goals, especially whether these reasons are autonomous or controlling, have a stronger impact on the aforementioned measures than the types of goals themselves. Their results revealed that people engaged in each type of goal in the 3x2 achievement goal model due to different, controlled or autonomous, reasons. Notably, they found that performance goals (other-based goals) were associated with higher satisfaction when being pursued due to autonomous reasons. These findings might explain why past research found negative effects of performance goals and positive effects of mastery goals on intrinsic motivation: Performance goals might be seen as controlling or out of reach (when perceived competence is low), leading to more controlled forms of regulation or even amotivation, according to SDT. In contrast, mastery goals, emphasizing potential to learn and develop without introducing normative or external evaluative pressure, might lead to more autonomous forms of regulation, since it might be easier for individuals to identify with such goals.

2.2.3 Hexad User Type Model

The Hexad user type model aims to explain user motivation within gamified systems by introducing six different user types [342]. To date, the Hexad is the only model specifically developed for gamified systems (in contrast to player type models focusing on player behavior in games) [261]. It was initially developed by Marczewski in the book "Even Ninja Monkeys Like to Play" [213]. The model is not based on observation and has not been empirically established. Instead, the model is based on SDT and practical design experience in the domain of gamification [342]. SDT is thereby the foundation for the six user types, since these user types are personifications of intrinsic and extrinsic types of motivation.

To investigate its validity, Tondello et al. [342] created a survey instrument to assess the degree to which a user belongs to each of the six user types. This implies an important attribute of Hexad: the user types should be seen as traits, meaning that a user can seldom be classified as one specific type. Instead, users – although often showing a tendency towards a user type – are in most cases motivated by all types to a certain extent [342]. The 24-item survey developed by Tondello et al. showed promising reliability in the first publication in 2016 [342], yet an empirical validation was still lacking. However, the authors demonstrated that Hexad user types and preferences for gamification elements are correlated. This is an important finding, as it suggests that knowledge of a tendency towards a certain user type allows tailored gamification elements for that type to be offered. In 2018, Tondello et al. [337] solved the issue of the lack of validation, adapting two items of the initial scale slightly and providing a validation of the Hexad scale in English and Spanish. Next, we will elaborate further on the six user types established by the Hexad model and their underlying motivational factors. The Hexad user type model is relevant for the studies described in subsequent chapters of this thesis (Section 4.2, Section 4.4, Section 4.5, Section 5.2, Section 5.3, Section 5.4, Section 6.2, and Section 6.3).

The Six User Types

Marczewski [213] introduced six user types. These user types differ in the degree to which they are driven by intrinsic or extrinsic motivational factors, as defined by SDT. Regarding the intrinsic sources of motivation, the Hexad includes the basic psychological needs postulated by SDT, namely competence, autonomy and relatedness. In addition, it considers *meaning* as an additional intrinsic source of motivation. Although meaning is not officially a basic psychological need according to SDT, it has been suggested as a candidate need [299]. The Hexad establishes the following user types [213, 342]:

Philanthropists ("PH") are socially-minded, like to bear responsibility, and enjoy sharing their knowledge with other users. Also, they are altruistic and like to give, provide help and support others without expecting a reward. Overall, they are driven by *purpose or meaning*.

Socializers ("SO") are socially-minded, too. However, they are more interested in social interaction, fostering social connections with other users of the gamified system and communicating with others. Consequently, *relatedness* is their main motivation.

Free Spirits ("FS") like acting without external control and having freedom to express themselves. They also like to engage in exploration and creation processes and customize the gamified system to their needs and preferences. *Autonomy* is most important for them.

Achievers ("AC") are motivated by *competence*. They like to overcome obstacles, make progress in gamified systems by reaching goals or completing tasks and engage in difficult challenges to prove themselves and obtain a feeling of mastery.

Players ("PL") focus on their benefits and seek to earn rewards for doing activities within a gamified system. They represent the least autonomous form of motivation within the Hexad model by focusing exclusively on *extrinsic rewards* and external stimuli regulating their motivation.

Disruptors ("DI") like to disrupt gamified systems to achieve negative or positive changes for others or themselves. Negative changes include cheating behavior, but Disruptors might also work to improve the system. Related to the SDT need for autonomy, Disruptors are ultimately driven by triggering *change*.

Relationship to Self-Determination Theory

The six user types established by the Hexad model can be seen as personifications of different types of motivation described by SDT [342]. Intrinsic motivation is reflected through the different user types related to basic psychological needs, while one user type is specifically focused on extrinsic motivation. Consequently, SDT, or more specifically organismic integration theory and basic psychological needs theory, are at the heart of the Hexad model, and the Hexad as such could be seen as a gamification-focused layer on top of them. Thus, the Hexad model follows the assumption that individual motivational orientations exist (proclaimed by causality orientations theory as part of SDT) and could be seen as a proxy conceptualizing inter-personal motivational orientations and differences in the domain of gamification.

2.2.4 Summary

In this section, we have introduced important models and theories guiding our research in relation to the development of research questions and hypotheses but also regarding the interpretation and discussion of our findings and deriving implications. We have also demonstrated the tight coupling of these models and theories to SDT. First, we have introduced the transtheoretical model of health behavior change, adding a temporal dimension to motivation of intentional behavior change through the stage of change construct. This dimension allows us to formalize and assess intentions to change behavior and evaluate the impact of this temporal dimension of the perception and effectiveness of gamification elements. We have also seen that the stages of change construct is tightly related to the internalization process introduced by SDT. Next, we introduced achievement goal theory and the various types of achievement goals. We have seen that the different types of goals establish different types of standards for how competence is defined and valenced. Again, we showed the relationships to SDT, since these different types of goals have been shown to stimulate different motivational outcomes, either supporting or thwarting intrinsic motivation. Finally, we introduced the Hexad user type model which can be used to formalize user motivation in the domain of gamified systems. By establishing six user types differing in the degree to which they are motivated by intrinsic or extrinsic regulations, it contributes a theoretical basis to investigate preferences and motivations of users in gamified systems. Since the Hexad model builds upon SDT and its sub-theories, its close relationship to SDT is obvious and was discussed at the end of this section.

2.3 Gamification and Game-Like Approaches in Different Contexts

To investigate the research questions of this thesis, we conducted studies in different behavior change contexts. Therefore, we will present past research that was conducted in these contexts in the following. First, we will start by presenting related works making use of gamification or game-like approaches in the domain of physical activity and nutrition. These works are relevant for the study of contextual factors when adding a public display to a gamified mobile application, as presented in Section 3.2; for establishing design guidelines for fitness goals to increase their persuasiveness (Section 4.3); for studying the impact of behavior change intentions and Hexad user types on the perception of gamification elements in the fitness context, as part of Section 4.5; to analyze the actual effects of behavior change intentions and Hexad user types on behavioral and psychological outcomes when running on a treadmill (Section 5.2); to investigate the long-term effects of personalizing gamified systems based on Hexad user types in a course booking system for a gym (Section 5.4) and to investigate the role of Hexad user types in explaining preferences for gamification elements in the domain of healthy eating (Section 4.4).

Afterwards, we will present past research about gamification being applied in the field of hand washing. This field of study is relevant for the *Germ Destroyer* system, encouraging proper hand washing in shared bathrooms, which is presented in Section 3.3. Lastly, we will focus on the domain of advertising and shed light on how gameful approaches have been used in this area. This is relevant for our studies about the potential of gamification to enhance advertisements' user experience and acceptance (Section 3.4).

2.3.1 Healthy Lifestyle

Physical activity and nutrition are core determinants of a healthy lifestyle [139]. Given that overweight and obesity are the most prevalent health issues today [103], encouraging users to be physically active and cultivate healthy eating habits has been investigated via numerous interventions in the past. Often, gamification or game-like approaches have been used to enhance users' motivation in this regard [7,146,315].

Physical Activity

For instance, Consolvo et al. [88,89] developed a mobile application called "UbiFit Garden", giving users feedback on their physical activity. The application uses a virtual garden as a metaphor to reflect what kind of activity has been performed and how frequently certain activity goals were completed. Such activity goals are reflected by flowers and butterflies, which may appear and grow. The virtual garden is visualized on the mobile phones of participants, without having them open any application. The different types of activity (e.g. cardio, strength or flexibility) are differentiated by the color or type of flower being shown. To infer the types of activity and track the user's performance, the system uses a small wearable containing several sensors such as an accelerometer or a barometer. In a field test with 28 participants, the authors found that the virtual garden visualization helped participants maintain their physical activity levels, while these decreased among participants who did not receive the virtual garden visualization. This finding was also supported by qualitative data from the interviews with participants, where a majority of 25 participants liked the ubiquitous presence of the virtual garden. After analyzing potential environmental influences, the authors came to the conclusion that it was the game-like virtual garden which contributed most to helping users maintain their fitness levels. In another study by Consolvo et al. [87], the focus was on investigating whether sharing step counts with social peers, compared to seeing only one's own step counts, has an influence on the number of steps taken per day. The authors present "Houston", a fitness application running on a mobile phone to track steps. The app has two versions. The "personal" version uses daily step goals and visualizes progression towards these goals over the last seven weeks. The "sharing" version offers the same set of features but also allows users to see the progress made towards daily step goals by others. Based on a three-week user study with 13 participants, the authors found that participants who received the "sharing" version were more likely to reach their daily step goal. They also presented several design requirements, one of which was supporting social influence.

Zuckerman and Gal-Oz [371] developed "StepByStep", a smartphone application to motivate people to increase their walking behavior. The authors developed three versions of the application. In one version, the application provided information on the users' walking time, set a daily step goal and provided real-time feedback towards this goal. In the second and third version of the application, gamification elements such as virtual rewards (points) and social comparison were used, respectively. In the second version, users were rewarded with points based on walking time and goal attainment. In the third version, users were shown a leaderboard, on which the step counts of other users were ranked. In the course of two user studies, the authors found that the versions offering additional gamification elements (points, social comparison) were just as effective as the first version, in which only goals and real-time feedback were used. They furthermore found that certain gamification elements were very effective for some participants, but not for all. As a potential reason, the authors state that interpersonal differences might explain the absence of effects in the gamified conditions. They also acknowledge the role of specific contextual factors as essential antecedents for engaging gamification. StepStream [230] is a system that serves the purpose of motivating adolescents to increase their physical activity levels. The system consists of a pedometer counting steps, which are uploaded to a website employing several gamification elements on a daily basis. This website has a social stream, showing achievements when users reach their daily step goals. The authors performed a user study with 42 middle school students which revealed that the system did not lead to an increase in step counts. As reported by the authors, participants were living in an urban community with low walkability. Thus, their intention to perform physical activity might have been low, and social comparison might have been unsuitable to motivate this population effectively. However, the results showed that the participants improved their attitudes about fitness and physical activity, and also reported a stronger sense of social support.

Chen and Pu [79] set out to investigate the different types of social gamification elements in the domain of physical activity. They compared the effectiveness of social collaboration, social competition and hybrid settings combining competitive and collaborative settings. They developed a smartphone application called "HealthyTogether" which pairs users to exercise together. In a user study with 36 participants (each participant was paired with another, resulting in 18 dyads), they found that collaboration and hybrid settings outperformed competition. In a similar fashion, Gui et al. [137] investigated social gamification elements in pre-existing social networks on the social networking app *WeChat*. Instead of pairing unknown users, the authors investigated whether existing social peers stimulate an engaging environment motivating physical activity through a qualitative user study with 32 participants. The authors concluded that sharing fitness data with already existing social networks can stimulate users to keep tracking their fitness activity. They also found that sharing fitness data with their peers has great potential to improve the social relationships of participants. Social gamification elements such as social comparison as well as tailored coaching messages and self-monitoring were compared in the context of step counting by Klein et al. [179]. The authors present "Active2Gether" [179], a smartphone application utilizing the aforementioned gamification elements to increase the walking behavior of users. In a user study by Middelweerd et

a. [229], two versions of the system were created and compared against FitBit, a commercially available fitness application. In the first condition, all gamification elements were activated. In the second condition, only self-monitoring and social comparison were activated. In the third condition, participants were given the FitBit application using self-monitoring. In the user study with a total of 104 participants, the authors found that the effect sizes for active time per day were larger in the second condition and smaller in the first condition, compared to the third, FitBit, condition. None of the differences was statistically significant. Understanding determinants of behavior in gamified systems and studying interventions accounting for individual differences is suggested as future work.

Further investigating the role of social factors, Cercos and Mueller [74] developed a system called "Watch your Steps", visualizing step counts on a public display. They conducted a preliminary study with 15 users over eight weeks. It was found that participants started socializing, discussing their step counts in front of the public display. Also, the public display led to an increased usage of the pedometers and more perceived motivation to be physically active. In the context of stair climbing, Meyer et al. [228] also introduced a public display to serve as a motivator. They developed the "ActiStairs" system, which was installed in a public mall to encourage stair climbing, and were interested in evaluating its acceptance. Through observations and interviews, the authors found that there were differences in the perception of the system across different target groups. Younger users, for instance, engaged more with the system than older users, and users who were less physically active liked the system more than those who were more physically active. Overall, the authors concluded that ActiStairs was successful in increasing awareness for stair climbing.

Fish'n'Steps [204] combined public and personal displays to investigate which of these two experimental conditions was more effective in increasing users' step counts. The system used game-like features by linking the user's step count to the growth and emotional state of a virtual fish. The authors conducted a user study in an office environment. While all participants were able to see their personal fish tank, half of them were additionally grouped in teams. Each team had their own fish tank, in which all of the team members' virtual fish were living. The public display, which was installed at a prominent location in the office, visualized the fish tanks of each team, thus using social comparison. The user study with 19 participants revealed no differences in the number of steps taken between these two conditions. As a potential reason, the authors stated that the absence of an effect might be due to the fact that participants had little chance to socialize. Following the idea of ambient public displays to encourage physical activity, Nakajima et al. [241] investigated so-called persuasive ambient mirrors. In one of the four case studies presented, a system is introduced in which virtual ambient paintings change their appearance based on how physically active users are. For example, there is a landscape painting on which a tree grows, depending on how active users are. In two user studies, one with 6 participants and the other with 8 participants, no effects on the walking behavior could be found. The

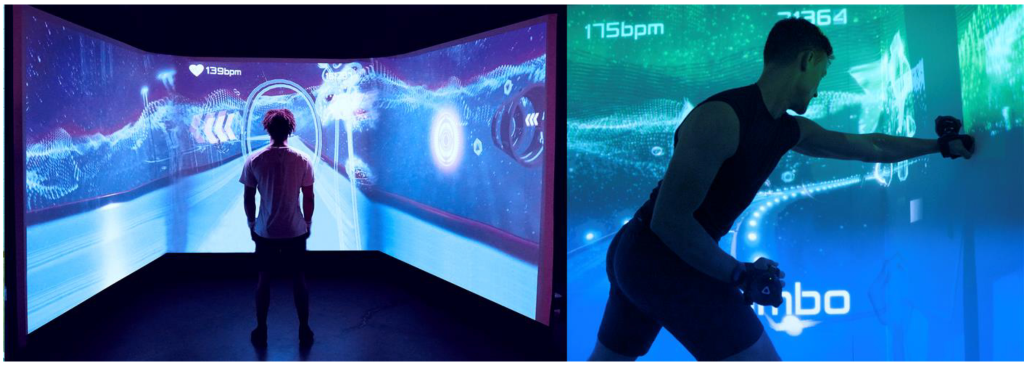


Figure 2.4: The *ExerCube* system. Left: Three walls on which the exergame is projected upon. Right: Interaction with the exergame through touching the walls. All images taken from [214].

authors speculated about the role of behavior change intentions in this context and stated that the type of feedback might need to be tailored to the stage of behavior change of the users.

While the aforementioned systems introduced game elements to a non-game context and thus can be considered as representatives of gamification in the context of physical activity, there is also a substantial amount of work on *exergames*, which can be defined as "interactive video gaming that stimulates an active, whole-body gaming experience" ([41], p. 1) or simply as "games that combine play and exercise" ([46], p. 1). For instance, Chittaro et al. [81] created a mobile exergame to encourage walking called "LocoSnake". The exergame builds upon the well-known game Snake and allows users to control the snake by walking. Users can first select a rectangular area on the map in which they would like to play. Next, their GPS position is mapped to the movement of the snake within the rectangular area, and their goal is to move the snake towards fruits to make progress in the game. Based on a within-subjects user study with 15 participants and two conditions (one in which participants played the exergame and one in which they walked without playing), the authors found that "LocoSnake" positively affected perceived exertion and was perceived as enjoyable. Another exergame was presented by Martin-Niedecken et al. [214]. They implemented an immersive and adaptive fitness system called "ExerCube". The system consists of three walls on which an interactive game is projected. The user stands within these walls and has to perform various fitness exercises in a dynamic way to make progress in the game (see Figure 2.4). Based on a user study with 40 participants, the authors showed that the ExerCube setup yields experiences of flow, enjoyment and motivation, which are comparable to those achieved in personal training with a human trainer. Similarly, Kosmalla et al. [188] implemented an interactive slackline training assistant. The system consists of a wall-sized projection in front of a slackline and a unit to track the user's position on it. Multiple exercises are presented to the user, and the system provides real-time feedback as the user performs each exercise. Their results of a lab study with 12 participants show that the gameful design led to an enjoyable experience and that the system was an effective alternative to a personal trainer.

The design and perception of exergames has also been researched among older adults. Gerling et al. [134] developed an exergame called "Silver Promenade", which takes into account age-related impairments and adapts its design to older adults' needs and preferences. The exergame itself is a collaborative game for three players in which virtual walks form the main game mechanic. The players have different roles. The *walker* performs walking movements on the Wii balance board to walk through the virtual world, the *shaker* has to shake the Wii Remote to count objects on the way and the *pointer* tries to catch objects by pointing the Wii remote at them. Based on a user study with 18 participants, the paper concludes that the exergame enhanced social interaction and was perceived as enjoyable. Furthermore, design opportunities were outlined, recommending consideration of the accessibility of exergames to older adults, relying on interaction metaphors that are known from the real world, and providing different roles and complexities to account for inter-personal differences. In a follow-up work by Gerling et al. [135], common age-related changes and impairments are discussed to derive game design recommendations, suggesting that physical impairments mainly affect the design of the user interface, while cognitive impairments and older adults' lower levels of gaming experience have more of an effect on the game mechanics of exergames for older adults. Brauner et al. [55] also designed and implemented an exergame for older adults. In the game called "Fruit Salad", apples and carrots are to be collected with a game avatar, which can be controlled by performing the corresponding gestures to pick up a carrot from the ground or grab an apple from the tree. Fruit Salad consists of three levels. In the first level, only carrots need to be picked, while in the second level, only apples need to be grabbed. In the third level, combinations of the two actions need to be performed. In the user study with 21 older adults, the authors found that participants could improve their performance when playing through the three game levels, and that most of them considered the exergame fun and would play it again.

An important aspect of exergames and game-like approaches to encourage physical activity is goal setting. While such goals are often static and do not change in the aforementioned systems and interventions, research has demonstrated that designing for dynamic goals is important, since they might change over time, as will be demonstrated in the following. Niess et al.[254] focus on fitness tracker goals and emphasize that they are evolving. To formalize this, the authors introduce the "Tracker Goal Evolution Model". In essence, the model states that qualitative goals (such as the general desire to do more sports) are built upon internalized hedonic and eudaimonic needs related to the anticipated benefits of tracking, which can be translated into quantitative, concrete fitness goals, often expressed by numbers. Epstein et al.'s [119] lived model of personal informatics is related to this. It also emphasizes the fact that motivations, goals and needs while using self-tracking technologies are subject to change over time. Also, Li et al. [201] follow the assumption that the motivation of fitness tracker users changes. According to their model, behavioral intentions are the cause for these changes. Accordingly, the model states that fitness tracker users progress through five phases (related to the stages of change introduced in Section 2.2.1), which pose different challenges to the user, affecting their motivation.

As we have seen, game-like approaches have been used to both encourage physical activity and guide users when performing physical activity. In fact, the research in this domain is substantial, with physical activity being one of the most targeted domains in gamification research [185, 186]. Therefore, recent literature reviews focused exclusively on the outcomes, approaches and issues of gamification in this domain. Koivisto and Hamari [185], for instance, conducted a systematic literature review on gamification encouraging physical activity at the end of 2018. After applying their filtering criteria, they included 16 papers. They analyzed which gamification elements have been used and found that goals and points were used most frequently. In addition, they analyzed which outcome measures have been used and found that subjective reports on physical activity were most frequently applied. Based on the systematic review, the authors identified gaps in research, highlighting that the effect of single gamification elements needs further investigation, since studying a combination of elements does not allow conclusions to be drawn on which aspects of the gamified system have led to the detected results. We will contribute toward closing this gap by studying the factors moderating the perception of single gamification elements in Chapter 4. Furthermore, we will contribute by studying the actual impact of adapting the set of gamification elements to these factors, as presented in Chapter 5. Moreover, the authors state that past research either used objective *or* subjective measures, which, taken on their own, do not allow the effects of gamification to be understood holistically. Thus, they recommend utilizing both types of measures to better understand the determinants of effects in gamification research. We follow this recommendation to better understand both contextual factors when applying gamification (Chapter 3) as well as individual factors (Chapter 4). As well, the authors consider potential novelty effects in research on gamification to encourage physical activity as problematic, since most studies had rather short study durations. We also contribute on this issue by investigating the long-term effectiveness of gamification in the context of a fitness course booking system, presented in Chapter 5.

Instead of focusing solely on gamification, Aldenaini et al. [7] conducted a systematic literature review about persuasive technologies for physical activity. They included 170 papers between 2003 and 2019. They found that roughly half of interventions were successful in encouraging physical activity. This means, as a consequence, that roughly half of interventions were only partially successful, or were even unsuccessful. The authors speculate that inter-personal differences might be part of the reason why interventions were not fully successful and emphasize that future work should be conducted on tailored persuasive strategies. This motivates studying potential factors moderating such differences and the actual effects of tailored gamification, to which we contribute in Chapter 4 and Chapter 5. Based on the literature review, the authors also recommend investigating the effectiveness of persuasive strategies across contexts, to get more insights on context-specific factors which may affect the perception and effectiveness of certain strategies. We will contribute toward this issue in Chapter 3. In line with Koivisto and Hamari [185], the authors highlight that future work should conduct long-term studies, investigate the perception of both multiple strategies and single strategies on their own and make use of mixed methods to get more holistic insights on factors affecting the perception and effectiveness of persuasive strategies. Lastly, they call for a diversification of the target audience, to which we contribute by studying the perception of gamification elements among older adults, when investigating age as a potential factor moderating the perception of gamification elements in Chapter 4.

Healthy Nutrition

Gamification and game-like approaches have also been used to encourage users to eat healthily. For instance, Schaefbauer et al. [306] developed a mobile application called “Snack Buddy”. This application offers several features to allow families to track their snacking behavior and gives feedback on the healthiness of other family members’ snacks. In addition to offering an informational interface where snacking behavior could be reviewed, a gameful interface was implemented. In this gameful interface, an avatar was shown and a narrative was used to encourage healthy snacking behaviors. When eating healthily, the avatar in the gameful interface makes progress in its life goals, i.e. when a healthy snacking behavior was adopted, the avatar could, for example, attend college or buy a house. The underlying core game mechanic was so-called snack healthiness points: When eating a healthy snack like a carrot, users received ten healthiness points, while rather unhealthy snacks like Hot Cheetos were worth only one healthiness point. Based on a twelve-week field trial, in which five families were assigned to a control group whereas five families were placed in the intervention group, the authors found that the system led to a decrease in the number of snacks, and that participants appreciated its social and gameful features. Bomfim et al. [47] created a gameful application called ”Pirate Bri’s Grocery Adventure”, which aimed at improving the food literacy of users holistically by fostering internalization of healthy shopping behavior. The gameful application could be played at home or at the grocery store. It allowed users to create their own character and provide information on their preference towards either salty or sugary food. In the game, the users are presented with challenges, which – depending on their preferences – are intended to stimulate their competence and motivate them to buy less sugary/salty food when grocery shopping. With the help of a virtual character, users are encouraged to create shopping lists. When at the store, they can add products to the virtual shopping cart by either scanning the barcode or entering the product manually. When leaving the store, the virtual character provides an overview on the items bought and stimulates reflection by announcing whether the users mastered certain challenges or reached their goals. Based on an exploratory 3-week long field study with 24 participants, the authors compared the gameful app against a non-gameful app from the app store. They found that the gameful app increased the participants' knowledge about food and had direct impacts on what they bought in the store. For instance, it was found that the gameful app was more successful in lowering purchases of ultra-processed foods than the non-gameful app. In addition, both apps were successful in increasing the amount of fruits and vegetables purchased.

Increasing fruit and vegetable consumption was also the goal of Jones et al. [170]. They investigated the effectiveness of a game-based intervention in an elementary school. This game-based intervention consisted of a narrative, which was supported by public displays providing further information on the narrative's characters. At the beginning of the game, there was a gathering with students, in which they were told the narrative, i.e. that heroes need their help in capturing members of the "villainous VAT (vegetation annihilation team)". Thus, students were told each day before lunch to eat a little more fruit or vegetables than normal. In an alternate-treatments study with 251 elementary school students, in which the phases of intervention and baseline were alternated, the authors found that on intervention days, fruit and vegetable consumption increased by 39% and 33%, respectively. Moreover, teachers reported that students enjoyed the game and recommended the use of such a game-based intervention in other schools. Similarly, Chang et al. [76] investigated a combination of a public display and a mobile application to encourage healthy food choices in a company cafeteria. On the public display, a collaborative food challenge was presented on a daily basis. Also, normative feedback on the joint progress in the daily challenge by all users currently eating in the cafeteria was visualized. To generate this feedback, the reported real-time lunch food consumption was used. To report their lunch intake, users could use a mobile application to input what they had for lunch together with their progress in the daily challenge. Using the app, users could also compare the nutritional content of their food to established guidelines as well as to the food other users had. The authors conducted a three-week-long field evaluation, having one week as the baseline and two weeks for the intervention, with 171 registered users. The findings showed that the public display was successful in attracting users and capturing their attention. Also, it was found that the social and normative feedback elements were effective in encouraging app use. Lastly, positive effects on self-awareness and self-reflection were found.

In a recent literature review from 2020 by Chow et al. [85], 43 publications were analyzed regarding their effectiveness in changing the eating behavior of children through gameful design. They found that the goals of the interventions could be categorized into four themes: *increasing fruit and vegetable intake*, *modifying snacking behavior*, *encouraging food exploration* and *promoting healthy eating*. Their analysis shows that game-based approaches mostly have positive effects on fruit and vegetable intake, increase knowledge on healthy eating and positively affect attitudes towards healthy eating. This indicates that the positive experience emerging from games can be transferred to the context of healthy eating through gamification and serious games. In terms of game elements being used in successful interventions, the authors emphasize that a compelling storyline of heroes completing quests seems to be particularly successful. They also highlight that most interventions used rewards such as points, badges or currency, which might be perceived as controlling. Thus, they might only lead to short-term effects on behavior, which might not last when these external stimuli are removed. In this regard, the authors conclude that the role of such game elements, as well as their perception, should be investigated further. We contribute toward this demand by investigating the perception of such gamification elements in the context of healthy eating, and to what extent different user types affect their perception, in Chapter 4. In line with the literature review by Koivisto and Hamari in the context of physical activity, the authors state that there is a need to understand the impact of single game elements, since the approaches studied so far used a wide range of different game elements, which makes understanding the impact of each individual one difficult. We will also contribute to this by investigating single gamification elements in the healthy eating domain in Chapter 4.

2.3.2 Hand Washing

In this section, we focus on the important role of hand washing for public and private health, which strategies to encourage proper hand washing have been investigated and what interventions there are for improving hand washing in the field of human-computer interaction.

Hand Washing as a Means to Improve Public and Private Health

The importance of hand washing for personal and public health was first noted back in the 1800s by Semmelweiss [178]. It has been found to be one of the most effective ways of preventing the spread of diseases [178], such as food poisoning, flu or diarrhea [4, 114] and is especially important in shared bathrooms [97]. It is also considered the most cost-effective way to improve public health [97], and research has demonstrated that improvements in hand hygiene are directly associated with lower rates of infectious illnesses [4, 97].

Especially today, considering the ongoing COVID-19 pandemic, washing one's own hands properly is essential: Pogrebna and Kharlamov [271] have found that the hand-washing culture of a country is a good predictor of the magnitude of COVID-19 spread, and partially explains why the virus spreads faster in certain countries than in others. Although it is simple, effective, and of substantial importance for health, hand washing, especially after using the bathroom, is often not done at all [169] and is mostly performed for a very short time [166]. This is problematic, since hand washing duration has been shown to be a key factor for properly removing microorganisms [322]. Past research suggests that we should wash our hands for 15–20 seconds, as this duration is most efficacious [166]. This proposed duration was confirmed by many countries and organizations, such as the Centers for Disease Control and Prevention [72]. It is also supported by Smith et al. [322]. They found that a 20-second hand wash removes bacteria and microorganisms more effectively than using a gel sanitizer with 70% alcohol. They also suggest that hand washing for less than 5 seconds is potentially worse than not washing at all. Also, hand disinfection may lead to bacteria developing antibiotic resistances [4], which is why hand washing should be preferred over hand disinfection in everyday life. However, Borchgrevink, Cha and Kim [49], who conducted field observations of 3,749 people in restrooms located across a college town, found that less than 6% of people washed their hands for longer than 15 seconds. Thus, they conclude that the general guidelines as recommended by the Centers for Disease Control and Prevention are not being practiced. Furthermore, a study by Wirthlin Worldwide and Bayer Pharmaceutical [363] aimed to provide answers to the question of how many people wash their hands after using a public bathroom. They also investigated whether self-reported hand washing behavior corresponds to the actual behavior in a public bathroom. To do this, they conducted phone interviews with 1,004 adults, and found that 94 percent reported that they always washed their hands after using public bathrooms. However, in observational studies in which the hand washing behavior of 6,333 people across five American cities was observed, they found that 32% of subjects did not wash their hands at all after using public bathrooms. These findings are supported by Guinan et al. [138]. The authors investigated the question "Who washes their hands after using the bathroom?" by conducting observational studies in the public bathrooms of middle and high school students. Based on observing 120 subjects, they found that 42% of the female and 52% of the male participants did not wash their hands at all when using a public bathroom. Such gender differences have been found consistently across different studies [49].

Ways to Improve Hand Washing Adherence

These aforementioned findings highlight that there is a need to increase hand washing adherence as well as hand washing duration in public bathrooms. This is important, as we have seen that washing one's hands properly and for a long enough time has great potential to improve personal and public health. Consequently, research has investigated how hand washing adherence could be improved and how people's attitude or behavior could be changed toward washing their hands properly.

Kinnison et al. [178], conducted observational studies with 599 participants in the public bathrooms of two shopping malls. They were interested in which factors have an influence on hand-washing behavior. They investigated whether gender, time of day and the presence of a sign reminding people to wash their hands has an impact on the behavior of people, among other factors. In general, they found that less than one third of participants washed their hands properly (i.e. in a way that would reduce contamination, meeting the guidelines of the Centers for Disease Control and Prevention). They also found that females were more likely to wash their hands properly than males (44.8% vs. 17.9%). However, regarding the presence of signs and the time of day, no significant differences could be found. Similarly, Johnson et al. [169] conducted observational studies in public bathrooms on a university campus to investigate whether there are differences regarding gender and the presence of a sign reminding people to wash their hands. They found that 61% of the women and only 37% of the men washed their hands properly after bathroom use, when no signs were present. When signs were introduced, the number of participants washing their hands properly increased significantly among women to 97%. For men, however, the number of subjects washing their hands did not increase, with only 35% washing their hands properly after public bathroom use in that case.

In line with the previous works, Edwards et al. [234] were interested in learning more about the factors predicting hand washing adherence. In contrast to the previously reported works, the authors investigated the impact of having an observer present, in addition to factors like gender and time of day. They conducted observational studies in public bathrooms on a university campus. To investigate the impact of having an observer present, the observers either hid in the toilet stalls such that no observer was visually present, or were standing at the sinks, and hence were visually present. A total of 184 people were observed. They found similar effects regarding gender as reported before, i.e. female participants were more likely to wash their hands than male participants. Also in line with previous findings, no effect regarding time of day was observed. However, they found that there was a significant effect regarding whether an observer was visually present or not. Among participants who had no observer visually present, 70% washed their hands. In contrast, out of those who had a visually present observer, 90% washed their hands after using the public bathroom.

The effect on hand washing adherence of having others present while in the bathroom was also investigated in a study by Nalbone et al. [242]. They conducted observational studies in a men's public bathroom at a casino. Similar to the study by Edwards et al. [234], the observer hid in the toilet stalls to give the illusion that others were alone in the bathroom to investigate the observer effect. Overall, 93 subjects were observed. The study revealed that 90% of the people using the public bathroom washed their hands while an observer was visually present. However, when no observer was visually present, only 44% washed their hands. Curtis et al. [97] conducted a review on the promotion of safe hygiene. They highlight that social norms and disgust are key motivations for hand washing and report positive effects of studies conducted in a public restroom and in a train station, which relied on promoting disgust to increase hand washing adherence. Regarding social norms, the authors provide an example from a public bathroom of a motorway service station, which installed signs saying "Is the person next to you washing with soap?", thus providing normative feedback. This intervention was successful in encouraging hand washing behavior.

Hand Washing in Human-Computer Interaction

In the domain of human-computer interaction, hand washing and improving associated behaviors has been targeted as well. For instance, Arroyo , Bonanni and Selker [26] recognized the potential of interactive systems installed at the sink. They conceptualized a variety of prototypes aiming at providing assistance and inducing behavior change. For example, they present ”HeatSink”, a system which can be installed at the water tap to provide visual feedback on the water temperature. It uses RGB LEDs to indicate whether the water is warm or cold. Also, ”SeeSink” automatically determines the right temperature of the water, depending on what is presented to the sink. ”CleanSink” encourages hand washing compliance by using a CCD camera to check whether hands are present, and provides feedback when enough time has passed by a flashing illumination in the stream of water. However, it should be noted that the use of optical sensors (CCD cameras) in the bathroom constitutes a substantial privacy threat for users [233]. Lastly, they present “Waterbot”, a system to encourage saving water, that can be attached to the water tap. The system uses RGB LEDs to provide visual feedback about water consumption, and auditory feedback is used to reinforce closing the tap. Since the paper is mainly focused on providing several design concepts for feedback and assistance systems at the sink, no user studies are reported to investigate the actual impact of such systems.

Asai et al. [27] investigated ways to encourage people to sanitize their hands. They developed three systems. In all three systems, a public display was used highlighting the importance of clean hands and attracting the attention of people passing by. In the first system, a Wii Balance Board was used to detect when users were in front of the sanitizer dispensers. Upon detecting a user, the system prompted users to sanitize their hands. Once they pushed the pump of the santitizer dispenser, a message was shown stating “Thank you for your cooperation”. This system was installed in a cafeteria of a public university for five days. The study revealed that the system was effective, as the number of people sanitizing their hands increased from roughly 2% to roughly 23%. The second system used the same feedback mechanisms, but relied on an optical sensor installed at the entrance of a hospital instead of a Wii Balance Board to detect whether people were passing by. A field test over four days in the public hospital revealed similar results as the first system, with hand sanitizing rates growing from roughly 2% without the system to more than 25% with the system. In the third system, the authors used a different feedback mechanism. Here, they relied on augmented reality, i.e. augmenting the camera view of the real world with digital content. They used a camera directed at the sanitizer dispenser and showed the live view of the camera on a public display. In this view, viruses and bacteria were visualized, which disappeared once users made use of the sanitizer. In a brief user study in a city hall over two days, the authors found that the system had only a small effect, increasing the number of people who sanitized their hands from about 1% to 3-5%. As a potential reason, the authors suggest that the context had a major impact on whether users were inclined to sanitize their hands or not.

However, considering the aforementioned work by Smith et al. [322] showing various drawbacks of hand sanitizer use in daily life, washing hands seems more conducive to a healthy lifestyle.

Targeting food workers, Mondol and Stankovic [233] developed a system called "Harmony" providing real-time feedback on the quality of hand washing. It reminds workers to wash their hands and is capable of storing hand-washing data for further use and analysis. The system uses smart watches to detect hand washing, determine its quality and provide feedback. Moreover, so-called "alert zones" are installed. When workers enter these zones, the watch checks whether the workers have washed their hands recently. If this was not the case, the watch prompts the worker to wash their hands. If the workers have washed their hands but with low quality, a quality prompt is presented on the smartwatch. Otherwise, no prompt is triggered. These alert zones are detected by using Bluetooth beacons. However, the research was more focused on the technical aspects of the system. In particular, the main contribution of the paper was the system itself and its classification algorithm to detect hand washing and infer its quality. In a system evaluation, the authors were able to demonstrate a high accuracy and robustness of their system. Similarly, Corato, Frucci and Di Baja [90] contribute a system to increase hand washing adherence in the work context. They focus on training surgery staff in proper hand washing. The virtual training application, which they developed to teach surgery staff to follow established hand washing procedures before accessing the operating theater, made use of augmented reality and relied on a camera to detect hands and washing gestures using color-based segmentation. On a monitor, the view of the camera is shown together with a semi-transparent overlay showing a video on how to wash one's hands properly. In the paper, no user evaluation is presented, and the use of a camera again raises certain privacy concerns.

Based on the findings of the past research, we develop a gamified system to encourage hand washing of sufficient duration in public bathrooms, which will be presented in Section 3.3. In contrast to previous research, we do not use any type of optical sensors so as to avoid privacy issues, and we make use of gamification elements to induce behavior change. Besides these rather system-focused contributions, we also provide insights on the effect of the employed gamification strategies on users' motivation through the lens of SDT by conducting both laboratory and in-the-wild user studies.

2.3.3 Advertising

Having presented related works from the field of physical activity, nutrition and hand washing, we now focus on the field of advertising. This section represents the last context in which we studied *why* and *how* gamification influences motivation and behavior by measuring variables related to SDT. These works are relevant for Section 3.4.

Advertising and its Effects on the User Experience

When online advertising emerged, the potential effects on the user experience were greatly neglected, and designers of websites and interactive services regarded advertisements as design constraints to which they had to adapt [292]. However, research has shown that providing a good user experience of advertisements is essential, as demonstrated by Rohrer and Boyd [292]. They reported findings from a collection of user feedback regarding the perception of ads shown on websites in the Yahoo! network. Their main findings were that deceptive ads had a negative impact on the user experience, that intrusive ads were more annoying but also positively affected perceived entertainment, and that ads that undermined user control negatively affected the user experience. Similarly, Kim et al. [176] investigated whether perceived entertainment value of advertisements affects the users' purchase intentions. They conducted a questionnaire-based study, distributing questionnaires among offline customers who had previously purchased from online shopping websites. Based on 264 responses, the authors were able to show that perceived entertainment affects trust towards websites, and thus buying intentions, positively. Their results highlight that perceived informativeness and entertainment are the most important factors of websites regarding purchasing decisions, which supports the assumption that a gamification approach in this context might be worth exploring.

Enhancing the user experience of advertisements was also the goal of Visuri, Hosio and Ferreira [353]. The authors highlight the increasing importance of creating entertaining and enjoyable advertisements to enhance the user experience as well as the ad effectiveness. They focus on mobile advertisements and try to find new formats that are advantageous in terms of enjoyment but also effectiveness. They propose a new non-disruptive ad type, which can be easily moved and deactivated, and thus returns control over ads to the user. This new ad type consists of a movable icon which does not take up too much screen space. In a user study with 10 participants, the authors found that besides having positive effects on ad effectiveness, substantial benefits for user experience were identified, since the proposed ad type facilitated interaction with the ad because it allowed users to control where the ad is placed on the screen and when to deactivate it. Gaining back control over ads can also be achieved by using so-called "ad blockers", i.e. software that prevents ads from loading. Such ad blockers are increasingly used to mitigate the negative effects of online ads, as described by Miroglio et al. [231]. The authors received usage reports of the Mozilla Firefox web browser, which they used to select data from users that had either installed an ad blocker or had not. After applying several criteria to select a set of eligible users, the authors analyzed the data of 16,414 users in the test group (who had installed an ad blocker) and 15,724 users in the control group (who had not installed an ad blocker). They compared these groups regarding several variables related to their engagement with the web, such as active time spent in the browser or number of pages visited. The results show that users who had an ad blocker installed visited more web pages than users without ad blockers, suggesting that using an ad blocker seems to be beneficial to the users' engagement with the web. Also, users who have an ad blocker installed spend more time actively in the browser. Considering these findings, it seems worthwhile to investigate whether giving users the option to playfully deactivate ads might lead to similar positive effects on enjoyment. These aforementioned findings highlight that enjoyment is very important to the acceptance and user experience of ads. Also, they show that autonomy is important, i.e. that giving users control over ads may improve the user experience.

Interactivity in Advertising

The role of interactivity of advertisements has been investigated regarding ad effectiveness as well as user experience. Liu and Shrum [205] state that interactivity in advertisements may create engaging experiences. They investigated under which conditions interactivity can enhance advertisements, but also how interactivity might negatively affect ads. The authors propose a multidimensional definition of interactivity to operationalize the term and apply this definition to previous research. They highlight that interactivity in online ads positively relates to user learning and user satisfaction. Consequently, positive effects on the recognition and recall of ads are likely. Also, they emphasize the crucial role of feeling in control for user experience. In contrast, Risden et al. [288] investigated the role of interactivity empirically by comparing ad effectiveness between television advertisements and interactive web advertisements. They created an interactive website as well as a TV advertisement for the same products. They ensured that the TV ad conveyed the same messages as the web ad; the only difference was that the ads on the web were interactive, i.e. animated anchors were displayed on a website, which, when the user rolled over them, showed a message for the ad and played a sound. In a study with 70 subjects aged 12-13 years old, the authors found that participants were more likely to mention a product that was advertised interactively than a product shown on TV, supporting the positive effects of interactivity.

Positive effects of interactivity were also demonstrated by Campbell and Wright [71]. They were interested in studying the interplay between interactivity, personal relevance and attitudes towards ads. To investigate this, they conducted two experiments: a survey study with 97 participants as well as a laboratory experiment with 118 participants. In the first study, the authors tested four conditions in a 2x2 factorial design (high/low personal relevance, high/low interactivity). The results of this first study revealed that the level of personal relevance and interactivity of the ad significantly influenced attitudes. In a second study, the same procedure was followed but in a more controlled setting, to investigate whether the results from the online study could be replicated. The results again showed that the perceived level of interactivity and the perceived personal relevance positively affected users' attitudes towards the ads and their featured products, as well as towards the host site. This indicates that interaction might not only lead to positive effects for the ad or the advertised product, but also for the website hosting such interactive ads. Related to this, Sundar and Kim [329] investigate the effect of interactivity, animation and ad shape on attitudes towards the product and brand being advertised. They conducted a laboratory study with 48 participants, in which 12 different ads (differing in the aforementioned attributes) were placed on news websites. Their results show that interactivity promotes user engagement with ads, which is expected to lead to positive attitudes towards the ad and the product as well as enhance user involvement with the product or brand. In addition, they also found that animation leads to similarly positive effects on user engagement, but negatively affects product involvement, i.e. participants had problems recalling product information with animated ads, which might be attributable to the distractive potential of animated ads. Lastly, the results show that the shape of the ad had an influence on the perception as well. Overall, square-shaped ads were favored by participants and were shown to be more engaging. However, this type of ad was less successful in conveying product information, compared to banner-shaped ads.

Game-Like Approaches in Advertising

Given the aforementioned results showing that the user experience and enjoyment of ads is considered as a core dimension of ad performance and that interactivity adds positively to the effectiveness and perception of ads, game-like approaches in advertising seem a promising research direction. Consequently, researchers have followed this direction and investigated whether game-like advertising could transfer the positive feelings and emotions induced by games to the advertised product or brand [45]. Also, a line of research analyzes whether the interactivity of digital games has an impact on players' memory and thus enhances ad effectiveness [369].

A prominent approach called "In-Game Advertising" (IGA) integrates brands or products into digital games as a means to (subtly) affect memory of the brand or product [331]. However, in IGA, the focus is still on the game and on entertainment; not on the product or brand being advertised. According to Terlutter and Capella [331], several IGA characteristics influence attitudes towards the advertised product or brand. For instance, a high congruence between the advertised product or brand and the game leads to improved memory, but also to negative attitudes towards the brand or product. While the focus is on the game in IGAs, it shifts towards the brand or product in so-called "advergames" [331]: Such advergames are designed specifically to promote a product or brand. They are usually free of charge, easy to install and play and offer quick rewards.

Regarding IGA, Nelson [245] investigated the (short- and long-term) effectiveness and the perception of the approach. They conducted two studies. In the first study, 20 subjects played the racing game Gran Turismo 2 on a gaming console. The game has several in-game ads, e.g. for car brands. Five months after playing the game, the subjects were contacted and asked about which brands or products they remembered. Directly after the game session, participants were asked questions regarding their perception of IGA and also which brands and products they recalled. The first study revealed that most players do not consider the practice of brand placements as deceptive and that product placements could add to the realism of games. Regarding brand or product recall, participants could remember 4.53 brands on average directly after playing the game, whereas this number dropped to a maximum of two brands or products being recalled after five months. In the second study, 16 participants again played a racing game. However, this time all players had the same car, and advertisements were placed in the game. The procedure followed was the same as in the first study. They found that players again did not perceive product placements negatively. It was also found that players recalled 25% to 30% of brands immediately after playing the game. Again, a drop was found five months later, with participants recalling only one or two product placements. Similarly, Yang et al. [369] investigated implicit and explicit memory for brands in two sports video games. They conducted a study with 153 participants, using two games that participants were asked to play. The first was a soccer game while the second was a racing game; both games included several brand and product placements. After playing the game for 20 minutes, participants were confronted with several tests assessing their implicit and explicit brand and product memory. The results indicated that both explicit and implicit memory were affected positively, compared to a control group. However, stronger effects were found regarding implicit memory.

Regarding advergames, Ho et al. [319] investigated different ways of placing brands and their effect on implicit and explicit memory. They compared three types of brand placement. In associative placement, the brand is placed in the background; in illustrative placement, it is placed prominently and in demonstrative placement, the user may even interact with the product or brand. They conducted a user study with 150 participants, in which each participant was randomly assigned to either one of three intervention conditions (according to the aforementioned types of brand placement) or a control group. Their findings indicate that the type of placement has an effect on implicit and explicit measures of memory. Explicit memory was decreased in the associative placement condition, compared to the other two conditions. However, at the same time, associative placement led to higher implicit levels of memory than the other two types of brand placements. Consequently, the authors show that subtle product placements lead to a lower level of explicit but a higher level of implicit memory.

In the context of marketing campaigns, so-called alternate reality games, i.e. games that "blur the distinction between a player's experience in the digital world inside the game and the real world outside the game" ([177], p. 1), have been used to promote products and services. One prominent example is the alternate reality game "I Love Bees" which had more than one million players [177]. It was launched in 2004 to promote the release of the Xbox game "Halo 2". For less than a second, a URL was shown in the official trailer of the game, pointing users to a website. This website seemed to be taken over by a mysterious artificial intelligence program from another universe. The website had various hidden clues which, together with information provided through different channels and the real world, helped players to make progress in the game. Step by step, the players understood that they had to combine virtual gameplay with events in the real world, to help an artificial intelligence that was stranded on Earth find its way back to the Halo world. Similarly, "The Beast", another alternate reality game launched in 2001 to promote the Steven Spielberg movie "Artificial Intelligence", attracted over three million players worldwide [177]. In this game, players had to collaboratively solve a murder mystery story including characters from the movie. The first clue was printed on the movie posters, which prompted millions of people to form a whole community around the puzzles, who discussed and heavily engaged with the story to solve the crime.

Lastly, gamified advertising has also been researched. Bittner and Shipper [45] investigated the use of gameful elements in advertisements for sports products. Here, it needs to be considered that the authors did not investigate any interactive system, but used the term "gamified products" for banner ads in which the product slogans suggested playful features of the product. They conducted an online study with 101 participants. Participants were shown gamified and non-gamified ads for sports products in a within-subjects design. The results revealed that the purchase intentions of gamified products were predicted by different factors than the purchase intention of non-gamified products, such that gamification could be useful to products being influenced by social pressure or uncertainty (e.g. to decrease the (negative) influence of other people). They also found that gamification might be more suitable for people who had experience with digital games and that with increasing age, the positive effect of gamification on purchase intentions seemed to vanish. Finally, it can be concluded that gamified products may positively support intrinsic motivation, which in turn leads to an increased enjoyment of the advertised product.

As revealed by the related works presented above, enjoyment and the user experience are very important for successful advertising. This motivates the approach of using gamification in this domain. Also, we have seen that game-like approaches, such as advergames or alternate reality games, can be very successful in marketing and advertising. However, we have also learned that a lack of user control is one of the main issues with advertisements, relating to a lack of autonomy. Therefore, it will be interesting to investigate to what extent this lack of autonomy might have an influence on the success of gamification in this context. We will elaborate on and contribute to these questions in Section 3.4.

2.3.4 Summary

In this section, we presented related work on the use of game-like approaches in several behavior change support contexts such as physical activity, healthy eating, hand washing and advertising. These domains were considered due to the fact that the interventions presented in Chapter 3 are situated in these contexts. We learned that gamification, when using a one-size-fits-all approach, i.e. a static set of gamification elements for every user, may lead to positive effects on the user experience and may evoke changes in the behavior of users. These are important findings, since they demonstrate the power of using game-like approaches to enhance the motivation of people.

However, we also saw that gamification is not always successful. In fact, almost half of the interventions considered in systematic literature reviews were only partially successful or even unsuccessful. The underlying reasons for this phenomenon are still mostly unknown. A common problem related to this is that most gamification interventions are not theory-driven and lack insights about the reasons why gamification does or does not work. We will contribute toward closing this gap in gamification research by investigating gamification in different contexts, and elaborating on the effects of gamification on behavioral and psychological outcomes through the theoretical lens of SDT. We are thus able to contribute insights on the role of contextual factors: As we have seen that e.g. a lack of control, and thus a lack of autonomy, is a core problem in online advertising, it will be interesting to see whether gamification approaches utilizing gamification elements which support the need for autonomy are capable of overcoming this context-inherent lack of autonomy and thus increasing users' motivation. In contrast, we saw that there might be a lack of knowledge regarding proper hand washing. As a consequence, a gamified system targeting this lack of knowledge by providing informative feedback might be able to enhance the context-inherent lack of competence and thus add to users' motivation to wash their hands properly. Whether such assumptions hold, and to what extent gamified systems may help to enhance users' motivation to become more physically active, wash their hands properly or engage in online advertisements will be discussed in Chapter 3.

2.4 Factors in, and Individualization of, Gamification

As we have seen in the previous section, gamification was shown to have mostly positive but also negative outcomes on psychological and behavioral measures. Therefore, gamification research has focused on understanding what makes gamification successful (or not) [239]. Since most empirical studies investigated a combination of gamification elements, making it hard or even impossible to understand the effect of individual gamification elements [146, 315], one research direction to follow was investigating the effect of individual gamification elements on psychological and behavioral outcomes. We will start by presenting related research in this area. However, as we will see, although research in this area revealed important findings regarding the motivational impact of gamification elements, the findings still differ and contradict each other to a certain extent. Therefore, gamification research has mainly focused on two approaches to tackle this problem [260]: allowing users to adjust which gamification elements should be used in a gamified system, and modeling user preferences to allow for adapting gamified systems to the user. The former approach will be presented first and can be referred to as *customization*, i.e. the user adapting the system to their needs, while the latter is called *personalization* (presented afterwards), i.e. the system adapting to the user. Regarding personalization, research investigated which factors mediate the perception of single gamification elements to allow for such adaptations. The works presented in this section are mainly relevant for the studies we present in Chapter 4 and Chapter 5.

2.4.1 The Effect of Gamification Elements on Intrinsic Motivation

To what extent single gamification elements affect intrinsic motivation has been studied extensively by Mekler et al. [223, 224, 225]. For instance, Mekler et al. [225] compared the effects of points and meaningful framing on intrinsic motivation and task performance in an image tagging task. They set up an online platform in which users were randomly distributed in one of four conditions (meaningful framing and points, framing and no points, no framing and points, no framing and no points). On this platform, participants were asked to provide tags describing the mood of 15 abstract paintings after completing a practice trial explaining the task. Participants in the condition with meaningful framing received the information that their tags would help to improve affective image classification and thus help to advance science. Participants in the condition in which points were given received 100 points per tag provided. The current score was shown to participants in the upper right corner of the platform. In the experiment with 172 participants, the authors found that participants in the points conditions provided significantly more tags and that framing significantly increased the quality of tags. Regarding intrinsic motivation, it was found that it did not matter whether points, meaningful framing or both were provided. However, when neither meaningful framing nor points were provided, intrinsic motivation was significantly lowered.

Using the same image-tagging task, Mekler et al. [223] investigated the effect of points, a leaderboard and levels on task performance and intrinsic motivation. The authors stated that these gamification elements were seen as extrinsic incentives and thus could threaten intrinsic motivation. However, at the same time, the authors acknowledged that such gamification elements also provided positive feedback which might be perceived as informational rather than controlling. Thus, such gamification elements could also lead to more autonomous forms of motivation. To shed light on this aspect, they ran an online study with four conditions and a total of 295 participants: a control condition without any gamification elements, a condition in which participants received 100 points for each tag, a condition in which fictitious users were shown on a leaderboard together with the participant and a condition in which participants could reach levels (a progress bar indicated the progress towards the next level). The results showed that all gamification elements led to participants providing significantly more tags than in the control condition. Also, the level and leaderboard conditions outperformed the points condition. Regarding intrinsic motivation, no effects were found at all, i.e. the control condition did not differ compared to any of the conditions using gamification elements. The authors speculate that the gamification elements did not have detrimental effects on intrinsic motivation because they were not linked to any pressuring external events. Also, they speculate that the gamification elements did not lead to an increase of intrinsic motivation because they might not have provided enough informational feedback on the participants' performance. It was also mentioned that participants voluntarily participated in the study, which suggests that their intrinsic motivation towards participation was already reasonably high; this might also be a possible explanation for the absence of effects on intrinsic motivation.

In the most recent work using the same image tagging platform by Mekler et al. [224], the authors were interested in better understanding the moderating factors that have an influence on whether gamification elements enhance or diminish intrinsic motivation. Similar to the previous work, they conducted an online study with 273 participants who were confronted with the same gamification elements as before, or were assigned to a control condition without gamification elements. However, to better understand what mediates the effect of these gamification elements on intrinsic motivation, they assessed the causality orientation of participants (control-oriented vs. autonomy-oriented; cf. causality orientations theory described in Section 2.1.2). The results show again that gamification elements led to an increased task performance with participants providing significantly more tags in all gamification conditions than in the control condition. Also, as before, the levels and the leaderboard condition outperformed the points condition. Regarding the causality orientation, no significant interaction effects between causality orientation and gamification elements was found. However, a main effect of causality orientation was significant: autonomy-oriented participants provided significantly more tags than control-orientated participants. In line with the previous results, no effect (neither a main effect of gamification elements nor an interaction effect of gamification elements and causality orientations) on intrinsic motivation was found. In line with SDT, a main effect of causality orientations on intrinsic motivation was found: Autonomy-orientated participants were significantly more intrinsically motivated than control-orientated participants.

In contrast to the results by Mekler et al., Hanus and Fox [147] found that badge and leaderboard gamification elements detrimentally affected intrinsic motivation over time. They conducted a 16-week user study to investigate the effects of gamification elements on intrinsic motivation in an educational setting. In their study, 80 students from two different classes of the same course were recruited. Data was gathered four times during a 16-week-long intervention. While one class received no gamification elements at all, the leaderboard and badge gamification elements were introduced to the second class. The findings showed that students in the gamified class had decreased motivation and satisfaction, ultimately leading to lower final exam scores. When interpreting these results it should be considered that participants could decide for themselves in which class to participate. Thus, they were likely interested in the subject of the class and might have been rather autonomously motivated. When confronted with competition and badges, these were likely perceived as controlling, affecting their motivation negatively. The fact that unlocking badges was mandatory in the gamified course seems to have added further to the aforementioned problem. Furthermore, the authors emphasized that other factors, foremost among them intra-personal factors, might play an important role in how gamification is perceived and how it affects intrinsic motivation.

The findings of Sailer et al. [303], who investigated the effects of gamification elements on need satisfaction, contradict the findings by Hanus and Fox to a certain extent. Sailer et al. conducted a randomized controlled online study in which participants were confronted with an order-picking task, i.e. participants had to find products for an order in a storage depot. There were three conditions: In the control condition, no gamification elements besides points were activated. In the first experimental condition, badges, a leaderboard and a performance graph were added. In the second experimental condition, avatars, a meaningful story and teammates were introduced alongside the points. Overall, 419 participants took part in the study. Regarding the competence need satisfaction and the perceived meaningfulness of the task, the authors found a significant effect between the control and the first experimental condition. The badge, leaderboard and performance graph gamification elements significantly increased both the perceived competence of participants and the perceived meaningfulness of the task. Also, they found that participants in the second experimental condition had significantly higher relatedness need satisfactions than those in the control condition and in the first experimental condition. This shows that having avatars, meaningful stories and teammates as gamification elements enhances social relatedness. These findings demonstrate that gamification elements can positively affect psychological need satisfaction and thus intrinsic motivation. In line with these findings, Xi and Hamari [365] show that gamification elements "can have a substantially positive effect on intrinsic need satisfaction" (p. 1). They conducted a survey-based study in Xiaomi and Huawei online gamified communities with 824 participants, assessing the effect of immersion-related gamification elements (such as avatars or narratives), achievement-related gamification elements (such as leaderboards or badges) and social-related gamification elements (such as collaboration and social networks). The study revealed that immersion-related gamification elements increased perceived autonomy and choice. Also, achievement-related gamification elements were shown to enhance autonomy, competence and relatedness need satisfaction. Finally, social-related gamification elements were indicative for increased satisfaction of competence, autonomy, and relatedness needs.

Considering all of the studies on the effects of gamification elements on intrinsic motivation presented in this section, it becomes obvious that there is a broad range of outcomes: The work by Mekler et al. [223, 224, 225] revealed no effects on intrinsic motivation, while the study by Hanus and Fox [147] showed that gamification elements undermined intrinsic motivation. The opposite effect was found by Sailer et al. [303] as well as Xi and Hamari [365]. Here, gamification elements substantially enhanced basic psychological need satisfaction and intrinsic motivation. Thus, it seems like other factors besides the gamification elements themselves play a major role in this regard. Individual factors were frequently mentioned as potential mediating factors, leading to research focusing on how individual factors can be accounted for. Here, two main general approaches – customization and personalization – emerged and have been studied, as will be described in the following.

2.4.2 Customization

One approach to account for the diversity in the perception of gamification elements is to allow users to adapt the gamification elements offered by a gamified system to their needs through *customization*. This approach is particularly promising regarding satisfying the basic psychological need for autonomy, as it gives users more control over the gamified system. It also follows the idea of meaningful gamification posited by Nicholson [251], according to which gamification should support users in building meaningful connections between the gamification elements and the users' personal objectives. Therefore, users should be put at the center of gamified systems by involving them in the design process of gamified systems. Nicholson also calls for customization options allowing users to adapt the system to their needs and objectives.

Customization was also shown to be beneficial in the domain of games. Research has, for example, found that allowing players to customize their avatar has positive effects on player identification with the avatar and the game experience. For instance, Birk et al. [42] investigated the effects of offering customization options in the context of avatar creation in an infinite runner game. In a study with 130 participants, two conditions were compared against each other. In one condition, participants received a random avatar in the game, while in the second condition they were allowed to customize their avatar. Their results showed that customization stimulated avatar identification, with significant differences in identification measures. Moreover, the results revealed that customization improved the satisfaction of the autonomy need as well as immersion. Regarding intrinsic motivation, customization and its improvement in identification significantly enhanced enjoyment and effort, and led to players playing the game for a longer time. Also, measures of positive affect were predicted positively by avatar identification and thus customization. In a similar study, Turkay and Kinzer [348] investigated the effect of avatar customization on player identification with the avatar in a massively multiplayer online game. They conducted a laboratory study with 66 participants over a duration of two weeks. Participants were invited four times during the study duration to play the game, totaling approximately ten hours of playing time per participant. Furthermore, participants were randomly assigned to one of two groups: customization, allowing players to customize their avatar, and no customization, where players were given a pre-configured avatar. The results revealed that participants in the customization group identified with their characters significantly more than players in the no customization group. This was found to potentially be the cause for a higher sense of presence in the game among participants in the customization group. The authors suggest that the positive effects found in the customization group were due to the increased autonomy need satisfaction, similar to Birk et al [42].

In gamification research, customization was frequently treated as its own gamification element enabling "the self-expression of the users through the creation and decoration of their virtual space, their avatar or their character and the personalization of some aspects of the system's interface" ([182], p. 6). However, Lessel et al. [198] went a step further by considering customization more globally in what they called *bottom-up gamification*. Here, users could adjust which gamification elements would be activated in the gamified system, adjust the gamification elements regarding their goals and visual appearance and even combine gamification elements as they saw fit – during the run-time of the system. They conducted three user studies: In an online survey with 75 participants, they found that participants were open to the bottom-up approach and could imagine defining their own gamification elements. In an interview study with eight employees in a manufacturing company, the authors found that, in line with the online study, the concept of bottom-up gamification was well received. They also conducted an in-the-wild study with 20 participants, who were using a task management application which allowed them to create their own gamification setups with a substantial amount of customization options. In this study, the perception of bottom-up gamification and its effect on task performance were analyzed. The results showed that significantly more tasks were gamified than not gamified by participants; that participants did not use a wide range of possible combinations of gamification elements, since simple gamification setups containing a small set of gamification elements were mainly used; and that participants tended to stick with the same gamification elements. This suggests that the broad range of combinations was overwhelming for participants. However, on a descriptive level, participants perceived the concept of bottom-up gamification well.

In a follow-up work, the effectiveness of bottom-up gamification was investigated by Lessel et al. [199]. The authors used an image classification task on a website to investigate whether bottom-up gamification and the potentially increased autonomy results in higher task performance (see Figure 2.5). In an online study with 106 participants and five conditions (control without gamification and four gamification conditions differing in the amount of choice and customization options), no effects were found between the conditions. However, when considering only users who actively made use of bottom-up gamification (i.e. those who activated a gamification element manually, or switched gamification elements), a significant effect regarding task performance was found. Users

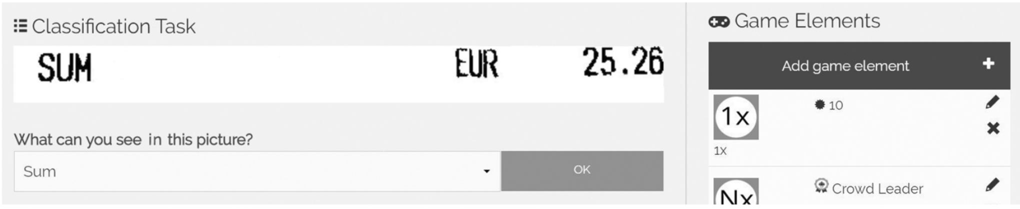


Figure 2.5: Microtask user interface and gamification element selection as part of *bottom-up gamification*. Taken from [199].

who were in adaptable gamification conditions (allowing them to switch, adapt or select gamification elements) solved significantly more tasks than those in other conditions.

To better understand the capability of users in coming up with motivating gamification setups, Lessel et al. [197] conducted an online study with 140 participants, in which they were asked to describe textually how they would gamify a certain scenario (participants were confronted with one out of four scenarios). The textual responses were analyzed qualitatively by inductive coding with two raters. The authors found that the range of gamification elements used was very broad, showing that participants contributed diverse concepts to motivate themselves with gamification elements. The analysis further revealed that participants not only used a wide range of gamification elements but also subjectively considered them as motivating. However, since it was not possible to investigate whether the proposed concepts would actually be motivating for participants if implemented, Schubhan et al. [311] bridged this gap by allowing participants to textually describe a gamification concept which would motivate participants in the context of image classification to provide more tags, implemented the proposed concepts and presented the participants with their self-created gamification setup. In an online-study with 71 participants, these self-created gamification concepts were compared against a no-gamification control group, and two top-down gamification setups with fixed, pre-defined gamification elements in a within-subjects study on an image-tagging platform, similar to Mekler et al. [223, 224, 225]. It was found that participants receiving top-down gamification provided more tags for images than participants in the control condition, and that providing participants with their self-created gamification setups increased the number of tags significantly, compared to the control and top-down gamification conditions.

While the aforementioned investigations show that customization can positively affect the perception and performance of users in gamification systems, the underlying reasons for the effect remain unclear – do users select gamification elements which are particularly suitable for them, or is the mere choice the cause for the positive outcomes? To shed more light on this question, Lessel et al. [200] investigated whether a very simplistic form of customization – the choice to enable or disable gamification – leads to positive outcomes. They again used an image tagging platform, on which participants could select whether they would like to enable gamification or not, and compared this group against a group of participants who were given no gamification and a group in which gamification was activated by default. The online study with 77 participants revealed that choice had an impact on the performance of participants. Those who were able to decide whether to enable gamification or not, and who decided to disable it, provided significantly more tags than users in the control condition without gamification. Although no further differences were found, it can be concluded that offering a choice may positively impact those who would like to disable gamification and does not negatively impact other users.

Lastly, Tondello and Nacke [339] used an image tagging context to investigate whether participants select gamification elements that match their Hexad user type. Also, they analyzed whether the possibility to customize the gamification setup has an effect on the performance of users, i.e. the amount of tags provided. They conducted two studies, differing in their recruitment strategy, with 252 participants, in which participants were asked to interact with a customized gamification setup (where participants were able to select which gamification elements they would like to activate on the platform) or a top-down gamification setup (in which participants were given a static set of gamification elements). Their results revealed correlations between the selected gamification elements in the customized condition and participants' Hexad user type scores. Also, users who could customize their setup provided more tags than those in the control condition without customization.

2.4.3 Personalization

Another way to account for individual differences in gamified systems is *personalization*. In contrast to customization, users do not adjust the system to their needs, but the system adjusts to the user by modeling the user's preferences and considering their personal characteristics. Orji et al. [260] called this approach "system-controlled" and referred to customization as "user-controlled". They compared these two approaches in a study with 1,768 users. Participants appreciated that customization gave them a sense of control, freedom and a personal touch. However, the main finding was that users preferred personalization over customization, i.e. the perceived persuasiveness of personalization was significantly higher than the perceived persuasiveness of customization. Reasons included that customization was seen as difficult, too time-consuming and distracting users from the main goals of the system. Also, users seemed not to trust their own knowledge on what would be best for them, questioning whether they were able to select gamification elements that would actually suit their preferences. These findings show that researching personalization is important as it provides a potential solution to the problem of individual differences in gamification, and is preferred over customization by users. In this section, we will provide an overview on which factors have been studied to personalize gamified systems and what the outcomes of these investigations were.

Demographic Factors

Demographic factors like gender and age are among the most commonly used factors to personalize gamified systems [182]. For instance, Tondello et al. [338] conducted an online study with 188 participants, in which they were asked to rate their perception of commonly used gamification elements in a general context. Next, an exploratory factor analysis of people's preferences was calculated to find clusters in the preferences for gamification elements. For each of these clusters, the authors analyzed to what extent demographic factors were related to them. For gender, the authors found that women scored significantly higher in clusters related to immersion and customization, and were more likely to accept help when interacting with a gamified system. Men were shown to score significantly higher in clusters related to socialization and collaboration as well as altruism. Regarding age, it was found that clusters related to rewards, incentives, customization and altruism were negatively correlated with age, i.e. they decreased with increasing age. When interpreting these results, it should be kept in mind that the sample was skewed towards younger participants (74% were 30 or younger; the maximum age was 71).

The effect of age and gender was also investigated by Oyibo et al. [264]. They analyzed to what extent the persuasiveness of persuasive strategies such as rewards, social comparison, social learning and competition were influenced by these factors using textual descriptions of the persuasive strategies in the form of questionnaire items, to which participants were asked to provide their level of agreement in a general context (similar to Tondello et al. [338]). Based on the responses of 323 participants, they found significant gender differences: female participants were significantly less inclined towards rewards and competition than male participants. Also, age-related differences were found: younger adults perceived competition, social learning and social comparison as significantly more persuasive than older adults. However, it should be noted that the authors used age 24 as the boundary for splitting participants into "young" and "old" groups. Furthermore, Orji [257] focused specifically on the role of gender in the perceived persuasiveness of persuasive strategies commonly applied in the health domain. They conducted a large-scale study with 1,108 participants and created storyboards illustrating persuasive strategies such as social comparison, rewards or cooperation in the context of healthy eating and losing weight. For five out of eight persuasive strategies, significant differences were found between male and female participants. Females perceived cooperation, customization, personalization, praise and simulation as significantly more persuasive than males. It is important to note that personalization was perceived as the most persuasive strategy for both females and males and that customization emerged as least persuasive. In a follow-up work by Orji et al. [258], both gender and age were considered as potential factors influencing the perceived persuasiveness of Cialdini's persuasion strategies – reciprocity, scarcity, authority, commitment, consistency, consensus and liking. Instead of using storyboards, the authors used textual descriptions of these strategies as part of an online questionnaire with

1,108 participants. The results revealed that female participants were generally more open to all persuasive strategies, compared to male participants. Also, differences on the level of the single strategies were found. Female participants perceived reciprocity, commitment and consensus as significantly more persuasive than male participants. Regarding age, the authors grouped participants into three age groups: 18–25, 26–35 and above 35. However, for the analysis only the latter two groups were considered. While no main effect of age was found, interaction effects between age and the strategies were identified: Adults perceived commitment as significantly more persuasive, while younger adults perceived scarcity as significantly more persuasive.

Denden et al. [102] investigated age and gender as factors moderating the perception of gamification elements in the education domain. They set up online courses using the course management system Moodle to teach object-oriented programming and basic software skills. In these courses, eight game elements were implemented: points, levels, a progress bar, a leaderboard, avatars, badges, feedback, and chat. 90 undergraduates, who took part in these courses and used the online platform for almost a month, were asked to fill out questionnaires assessing their perception of gamification elements and their preferences; 83 completed them. Their findings show that all gamification elements but the chat were perceived well and that female participants liked badges more than male participants. Busch et al. [66] investigated the effect of gender on the perceived persuasiveness of ten persuasion strategies in an online study with 592 participants. In line with previous research, storyboards were used to explain commonly used persuasive strategies such as rewards, social comparison and cooperation. In contrast to the previously described studies, Busch et al. decided to use a dimensional approach to assess gender identity. Participants scored on two dimensions: femininity and masculinity. Their results show that – while there were only two significant differences regarding gender – femininity was positively associated with eight strategies: competition, suggestions, simulation, self-monitoring, reward, praise, personalization and customization.

Focusing on age as a factor that potentially affects the perception of gamification elements, Kappen et al. [172] investigated exercise motivations in older adults. They conducted semi-structured interviews with 19 older adults and a focus group study with professional trainers and older adults leading an active lifestyle (participants were aged 50 years and up) to identify barriers and challenges in designing gamified applications encouraging physical activity. They found that social interaction with other people in similar age groups is important and that social comparison with younger people might reduce confidence. Also, the fear of incompetence was shown to be a strong motivator to engage in physical activity exercises. Overall, it was found that intrinsic motivations were more prevalent among older adults and that extrinsic motivations were not as valued by older adults. Birk et al. [43] also found that social interaction is more and more important with increasing age and that older adults refrain from performance-orientated goals when playing games. They conducted a cross-sectional study with 2,747 participants (ages 18–55) playing games to learn more about how play preferences change with increasing age. They found that as age increases, motivation to play and experiences while playing games become more intrinsically-focused (i.e. players are less motivated by performance goals and report higher levels of enjoyment and enhanced basic psychological needs satisfaction). This is also reflected in player styles: With increasing age, player styles shift towards completion-focused styles and tend to refrain from of performance-oriented playing styles.

Personality

Besides demographic factors, research has also considered personality traits as a potential mediating variable in the perception of gamification elements and persuasive strategies. For instance, Jia et al. [168] investigated the influence of Big-5 personality traits [219] on the perception of ten gamification elements such as points, badges, leaderboards and rewards. They used videos of a researcher interacting with gamification elements and textual descriptions of gamification elements to explain how they worked. They asked participants to rate their perceptions in a survey with 248 participants. Their results revealed several linear relationships between personality traits and the perception of gamification elements, based on calculating multiple regressions. For instance, they found that extroversion was positively associated with points, levels and leaderboards, that agreeableness was predictive of the perception of challenges, that conscientiousness was positively associated with levels and progress, that emotional stability was negatively associated with points, badges, progress and rewards, and that openness was negatively linked to the perception of avatars. In a follow-up work, Jia et al. [167] again investigated personality traits as a factor in how people perceive gamification elements, but focused specifically on the perception of leaderboards. Similar to research presented earlier, they used storyboards to explain the different types of leaderboards and asked participants to rate their perceived enjoyment based on these storyboards. In an online study with 286 participants, they found that more extroverted users, and users scoring high on the agreeableness factor of the Big-5 model, perceived leaderboards more positively.

To investigate the influence of personality traits on the perception of persuasive strategies in mobile health applications, Halko and Kientz [141] conducted an online survey with 240 participants. They created storyboards depicting eight different persuasive strategies, which can be categorized into four types of strategies: instruction style (authoritative and non-authoritative), social feedback (cooperative and competitive, see Figure 2.6), motivation type (extrinsic and intrinsic) and reinforcement type (positive and negative). After presenting the respective storyboard to participants, seven questions regarding their perceived enjoyment, likelihood of use, helpfulness, potential to improve the quality of life, perceived time saving, and perceived ease of use, as well as a free-text field for general

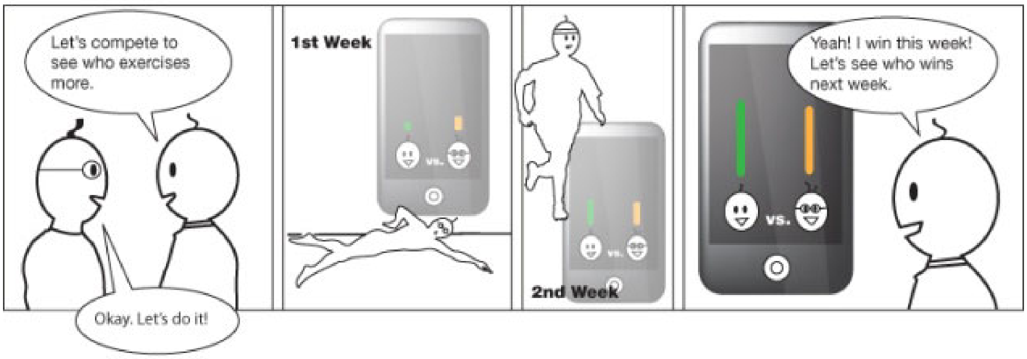


Figure 2.6: Storyboard used to illustrate the persuasive strategy *competitive social feedback*. Taken from [141]

comments, were provided. The authors analyzed the relationship between the persuasive strategies and Big-5 personality traits by calculating bivariate correlation coefficients. It was found that neuroticism was positively correlated with negative reinforcement and also negatively correlated with cooperation. Conscientiousness was positively correlated to competition and cooperation, while agreeableness was correlated positively with competition and negatively with both negative and positive reinforcement. Extroversion was negatively linked to extrinsic strategies and negative and positive reinforcement. Lastly, openness was positively correlated with authoritative and competitive strategies and negatively linked to extrinsic, intrinsic and negative reinforcement strategies.

Orji et al. [259] followed a similar approach as Halko and Kientz [141]. They sought to investigate the role of Big-5 personality traits in the perception of persuasive strategies in the domain of unhealthy alcohol consumption. Again, storyboards were used to explain ten persuasive strategies which are also commonly applied in gamification: customization, simulation, self-monitoring, goal-setting, personalization, punishment, reward, comparison, competition and cooperation. In an online study, in which 660 participants were confronted with these storyboards and asked to rate their perceived persuasiveness, the authors found that personality traits explain how persuasive certain strategies are perceived to be. For instance, it was found that all persuasive strategies were positively associated to the extroversion trait, and that extroversion and agreeableness were the personality traits most responsive to the persuasive strategies. In contrast, openness and neuroticism were found to be positively associated with the lowest number of strategies, with openness being negatively linked to almost all strategies. Tondello et al. [338], who conducted an online study on the perception of commonly used gamification elements (as presented earlier), also considered personality traits, age and gender. They found that extroversion is positively correlated to socialization and assistance, that neuroticism is positively correlated to incentives and that openness is positively correlated to customization.

Buckley and Doyle [62] investigated the effect of learning styles and personality traits in the context of a gamified learning intervention. Learning styles were operationalized using the Index of Learning Styles [120], and personality traits by relying on the Big-5 model. They used an existing platform, the National Tax Forecasting Project, on which students have to think about and forecast the outcome of the national budget in a gameful matter. They are given virtual money, which they can use to invest in the outcome they consider most likely to happen, regarding their predictions. As a tool for learning, the platform thus requires users to reflect on their decisions, search for information about the national budget and gather information from experts to make informed decisions to invest their virtual money well. Several gamification elements are integrated in this platform, such as achievements, avatars, leaderboards, levels, and more. This platform and the potential moderating role of personality traits and learning styles regarding the perception and effects of gamification were investigated in a class of 158 undergraduate students, of whom 95 filled out the surveys completely. It was found that active learners perceive gamification more positively, that participants with a global learning orientation enhanced their performance, that extroversion was positively correlated to the perception of the intervention and gamification, that conscientiousness was negatively correlated to the perception of gamification and that emotional stability was positively correlated to performance.

Nasirzadeh and Fathian [244] focused their research on the finance domain – they investigated the perception of gamification elements in banking and the role of personality traits in this regard. In an online study with 412 participants they collected information on the participants' Big-5 personality traits and their perception of commonly used gamification elements, among others. Overall, a considerable number of correlations were found; the authors focused on the strongest ones in their results. They found that higher levels of extroversion were associated with a more positive perception of avatars and a more negative perception of leaderboards. High agreeableness was positively related to lottery and countdown elements and negatively associated with social interaction. Participants scoring high on conscientiousness were associated with higher ratings of epic meaning, whereas they were associated with more negative perceptions of points and badges. Neuroticism was positively correlated to the countdown, information, reward, penalty and epic meaning elements. Openness was particularly positively associated with badges and competition.

Bartle's Player Types

Besides personality traits and demographic factors, player typologies play a major role in personalizing gamified systems and have even been identified as the most used factor in a recent literature review by Klock et al. [182].

Bartle's player typology [34] is one of the earliest player type classifications for video games [107]. It was established based on Bartle's observations about the underlying reasons why players play Multi-User Dungeon ("MUD") games. Bartle analyzed bulletin-board postings about the question of what players want out of a MUD. Based on a qualitative analysis of the answers, there emerged two dimensions along which reasons for playing can be categorized: action vs. interaction and player orientation vs. world orientation. Within these dimensions, four player types were established, each with different motivations and behaviors in MUDs: *Achievers* like gathering points and leveling up; *Explorers* are motivated by exploring the game world, looking for interesting features and figuring out how things work; *Killers* are motivated by imposing themselves on others and attacking other players; and *Socializers* are interested in social interaction with other players. However, over the years, significant criticism of the model emerged. For example, the typology is based on motivations and preferences of MUD players, which limits its generalizability to other games or gamification [36]. Besides the lack of generalizability, the lack of empirical validation of the model is even more concerning, as it poses a threat to use of the model for scientific purposes [36, 65].

Despite these criticisms, the model is still frequently used for personalization purposes or to inform the design of systems, also in the field of gamification [182]. Akasaki et al. [5], for instance, investigated the perception of gamification elements across three case studies and found that people are diverse, so that relying on one static gamification approach is not capable of accounting for users' various motivations and needs. In their second case study, 13 participants were interacting with a sharing economy service, which used gamification elements such as badges, collecting, ranks and a narrative. The authors analyzed preferences for gamification elements, depending on the Bartle's player type of participants. They found that Achievers and Killers preferred collecting and badges, while Explorers preferred collecting and the narrative. Fernandes and Junior [122] relied on Bartle's player types to select gamification elements in the context of e-government services. They expected that points, levels and badges should be positively perceived by Killers and Achievers and that levels, ranking and quests should be relevant for Killers, Achievers, Explorers and Socializers. However, the user study was not focused on evaluating whether these assumptions actually hold. Similarly, Fuß et al. [130] used Bartle's player types to inform the design of a gamified application made for university courses. They integrated a gamification element for each player type to motivate a broad range of users. As such, they added tasks differing in their difficulty for Achievers, a high score for Killers, social profiles and cooperation for Socializers and collectibles for Explorers.

BrainHex Player Types

The BrainHex player typology [238] is based on a series of demographic game design studies and neurobiological research. It introduces seven player types: The *Seeker* is motivated by exploring the game world and curiosity. The *Survivor* finds pleasure in frightening situations in games. The *Daredevil* is motivated by excitement, likes to take risks and seeks thrills. The *Mastermind* is driven by strategic planning and enjoys solving puzzles as well as making the most efficient decisions. The *Conqueror* is mainly motivated by challenges and enjoys defeating difficult enemies, achieving victory and mastering difficult situations. The *Socializer* is driven by social interaction and enjoys communicating with other players, helping them and hanging around with them in the virtual world. Lastly, the *Achiever* is similar to the Conqueror in the sense that they both like to complete goals. However, the Achiever is more goal-oriented while the Conqueror is more challenge-oriented. The Achiever is mainly driven by goal completion. Although more than 50,000 players completed the BrainHex survey and self-selected their player type based on textual descriptions, the typology itself has, similar to Bartle's typology, severe issues regarding its psychometric properties [336]. For instance, Busch et al. [65] found that only two types – Socializer and Achiever – could be differentiated as part of a confirmatory factor analysis and that the results are not stable over time, i.e. there are issues related to the test-retest reliability. Also, when trying to use player types of the BrainHex model as predictors for the game experience, no significant relationships between BrainHex player types and the game experience when presented with suitable game mechanics could be found in a study by Busch et al. [64]. Furthermore, Tondello et al. [341] re-analyzed the over 50,000 responses to the BrainHex survey by conducting an exploratory factor analysis, and were only able to discriminate three rather than seven stable factors. These findings show that the seven-factor structure postulated by BrainHex cannot be supported, detrimentally affecting the scientific support for BrainHex.

Nonetheless, despite the lack of validity and the aforementioned issues regarding the test- and re-test reliability, the model was and is still used for personalization purposes, also in the domain of gamification. Orji et al. [262] investigated whether relationships between the perceived persuasiveness of ten commonly used persuasive strategies and the BrainHex player types exist. They used storyboards illustrating the persuasive strategies in the context of healthy eating. In an online study with 1,108 gamers, they identified the best strategies for each player type. Achievers were most motivated by cooperation; Conquerors by competition and comparison; Daredevlis by simulation; Masterminds by self-monitoring and suggestion; Seekers by customization, Socializers by cooperation; and Survivors by self-monitoring and suggestions. Monterrat et al. [235] investigated whether tailoring gamification elements used within a gamified online learning environment to learn French grammar rules, based on the BrainHex class of users, has an impact on their perception of the system. They conducted a study with 280 participants, where they were randomly assigned to a group in which they received two suitable gamification elements based on their BrainHex profile, or two unsuitable gamification elements. The results showed that participants in the group with suitable gamification elements spent significantly more time on the platform. However, no differences in the enjoyment of the platform were found between the groups. Similarly, Lavoué et al. [193] conducted a study on the effectiveness of adapting gamification elements based on BrainHex player types in a web-based learning environment. They used the same platform as Monterrat et al. [235], teaching French spelling and grammar to learners, and recruited 266 participants. These were randomly assigned to one of three conditions: In one condition, participants received adapted gamification elements, based on their pre-assessed BrainHex class. In a second condition, participants received gamification elements which were counter-adapted. In a third group, no gamification elements were integrated. They found that among the learners using the platform particularly frequently, those who received adapted gamification elements spent significantly more time in the learning environment. Furthermore, participants receiving counter-adapted gamification elements reported higher levels of amotivation.

Hexad User Types

When investigating user preferences and personalizing gamified systems, the Hexad user typology (see Section 2.2.3) stands out for numerous reasons. First, it is the only model which has been specifically developed for gamified systems, instead of games [261]. Second, a survey has not only been developed to assess Hexad user type scores [342] but has also been empirically validated [337]. Third, Hexad user types have been demonstrated to be predictive in explaining preferences for gamification elements [342] across various domains. Lastly, the Hexad model has been shown to be favorable over personality traits or BrainHex player types in predicting preferences for gamification elements [144]. We will elaborate on these aspects further in the following.

When introducing the first questionnaire to assess Hexad user types to the field, Tondello et al. [342] also investigated potential correlations between the score in each of the six factors of the Hexad and the perception of commonly used gamification elements. They also aimed to investigate whether the recommendations for gamification elements established by Marczewski [213] could be empirically supported. They conducted an online study with 133 participants, in which they were asked to fill out the Hexad questionnaire and rate their perception of 32 gamification elements. To rate the perception of the gamification elements, each element was presented textually and participants had to rate how motivating they were on a 7-point scale. They found that overall, the user types were positively correlated with the gamification elements which were suggested by Marczewksi. As an example, Socializers were correlated with gamification elements such as social competition, guilds or teams and social networks; Free Spirits were shown to be positively associated with gamification elements such as unlockable content, exploratory tasks and customization; Achievers were correlated with elements such as challenges, badges or levels; Players with elements such as points, leaderboards and badges; and Disruptors with elements such as voting mechanisms and anarchic gameplay.

Orji et al. [261] examined to what extent preferences and the perceived persuasiveness of persuasive strategies aiming at preventing unhealthy alcohol consumption are explainable by considering Hexad user types. They conducted an online study, presenting storyboards explaining each persuasive strategy to participants, with a total of 543 subjects. The authors found that a user's Hexad type plays a considerable role in how persuasive participants perceived certain strategies to be. Players were most positively associated with competition and reward, and Philanthropists with simulation; Disruptors were negatively linked to goal-setting, personalization and simulation; Free Spirits were only weakly associated with personalization, Socializers with all strategies and Achievers with none. A similar goal but a different application domain was investigated by Kotsopoulos et al. [189]. They were interested in analyzing which gamification elements are suitable for which Hexad user type in the domain of energy conservation at work. In line with Tondello et al. [342], they conducted an online survey in which 99 participants were confronted with several gamification elements along with a textual description/definition of how the gamification elements work. When analyzing bivariate correlations between the perception of gamification elements and the factors of the Hexad model, they found significant positive correlations between Socializers and points, badges, rewards and roles; between Philanthropists and badges and roles; between Free Spirits and points, badges, progression, status, levels and roles; between Disruptors and status; and between Players and points, badges, leaderboards, status, and rewards.

Mora et al. [236] used Hexad user types to personalize learning experiences. Hence, besides health and energy conservation, they extended the application context of the Hexad model to the education domain. They investigated whether gamified learning experiences are more motivating when tailored to the Hexad type of users. In an online course with 81 students, they were asked to complete the Hexad user type questionnaire. Afterwards, one group received a one-size-fits-all gamification setup, while another group received gamification elements tailored to their Hexad type. When comparing both groups, no statistically significant differences were found. When analyzing descriptive data only, it was found that behavioral and psychological outcomes seemed to be enhanced in the personalization group. However, it should be considered that the distribution of participants among both groups was heavily unbalanced, with only 21 participants receiving the one-size-fits-all setup. In the online study with 188 participants conducted by Tondello et al. [338], which was presented previously, the clusters in the preferences for gamification elements were also analyzed regarding potential correlations to Hexad user types. They found that Free Spirits were positively correlated with immersion; that Philanthropists were positively associated with immersion, progression and altruism; that Achievers were positively linked to socialization, immersion, risk and reward, progression, and altruism; that Players showed positive associations to socialization, risk and reward and incentives; that Socializers were strongly correlated with socialization, and moderately linked to assistance and altruism; and positive relationships existed between Disruptors and immersion as well as risk/reward.

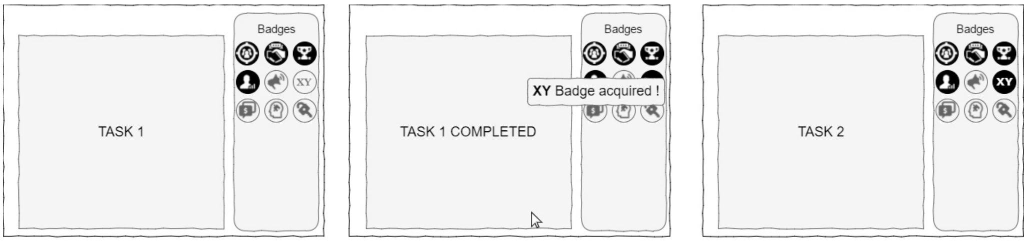


Figure 2.7: Storyboard for the gamification element badges. Taken from [144].

The Hexad model was also compared against the BrainHex model and personality traits regarding the extent to which these models are capable of explaining interpersonal differences in the perception of gamification elements by Hallifax et al. [144]. The authors conducted an online study with 300 participants, in which storyboards were used to illustrate commonly used gamification elements in a general context. The storyboards depicted the completion of a general task and receiving feedback from a certain gamification element. Figure 2.7 shows an exemplary storyboard for the gamification element badges. The authors presented pairs of gamification elements and asked participants which one of them they perceived as more motivating. When comparing the Hexad model, the BrainHex model and personality traits regarding which one was most relevant to explain user preferences for gamification elements, the authors found that the relationships between the factors and the gamification elements for the Hexad model were the most consistent with the definitions of the six user types and that the Hexad had more influence in explaining the perceived motivational impact of gamification elements than both other models. They assume that this is because the Hexad model was specifically designed for gamification (which, again, was not the case for other models) and because most of its user types are based on the well-established SDT. Therefore, they recommend using the Hexad model to tailor gamified systems.

In another study by Hallifax et al. [143], the authors compared three tailoring approaches in an educational context: the Hexad model, motivation of participants to engage in learning tasks, and a combination of both. They conducted a study with a gamified learning platform for mathematics. Participants were randomly assigned a gamification element. When analyzing the data, participants were split into two groups for each tailoring strategy, depending on whether the randomly assigned gamification element matched the recommended element of the strategy or not. Data was collected for six weeks among 4 high schools and 258 participants aged 13 to 14 years. One of six gamification elements were activated for each participant: an avatar, badges, progress, a leaderboard, points or a timer. The results reveal that different tailoring approaches can lead to different outcomes: Hexad-based tailoring was found to lead to more engagement but a lower performance and had no effects on motivation, while an adaptation based on initial motivation led to an increased intrinsic motivation to learn. When combining both these approaches, the results from motivation-based tailoring can be maintained and extended by an increase in fun and excitement to learn.

2.4.4 Summary

In this section, we presented related work contributing to the question of why and how gamification works. First, we have seen that gamification research focuses more and more on analyzing the impact of individual gamification elements, instead of investigating whether a set of combined gamification elements leads to certain beneficial outcomes. This allows us to pinpoint the effect of individual gamification elements on both psychological and behavioral measures. However, we have also seen that the outcomes of individual gamification elements still differ substantially. This has led to gamification research investigating factors having a moderating influence on how single gamification elements are perceived. We will also contribute to this by considering the impact of several factors on the *perception* of individual gamification elements, as part of Chapter 4.

Next, we presented related work about how to account for interpersonal differences in the perception of gamification elements. Here, two main approaches emerged: customization (the user adapting the system to their needs) and personalization (the system adapting to the user). Regarding the former, research has shown that the increased autonomy is appreciated by users. However, we have also seen that customization and the large amount of available options can be overwhelming, and that customization may be considered as difficult and complex, too time-consuming, or distracting, which is why participants preferred personalization when asked about their preferences. Hence, in this thesis, we focus on personalization and investigate which factors are relevant to adapt the set of gamification elements to the user (Chapter 4), and what effects such an adaptation approach has on psychological and behavioral outcomes (Chapter 5).

In the related work on personalization, we have seen that various factors have been considered – demographic factors, personality traits and player typologies such as Bartle's or BrainHex. Regarding demographic factors, we will extend previous research on the age factor by contributing insights from a considerably older sample on their preferences regarding gamification elements than was considered before, as well as by contributing qualitative insights, complementing the primarily quantitative studies conducted in the field of gamification in Chapter 4. Regarding trait models such as personality (Big-5) and player typologies, none of these was theoretically grounded to be used for the purpose of adapting gamified systems, and both player typologies lack a proof of validity. In contrast, the Hexad user type model is based on SDT, has been empirically validated and is the only model to date which has been developed specifically to model preferences in gamified systems. Moreover, in a direct comparison regarding the predictive quality of personality traits, BrainHex player types and Hexad user types, it was shown that Hexad user types are advantageous and thus should be used for personalization research in gamification. Therefore, we consider the model to investigate its predictive quality in explaining user preferences in health domains which have not been investigated before in Chapter 4, and contribute to open questions regarding the actual effects of tailoring based on the model on the behavior and motivation of users in Chapter 5. Moreover, in previous sections we have seen that behavior change intentions may change (as postulated by the Transtheoretical Model presented in Section 2.2.1) and that fitness goals change over time (see Section 2.3.1). Therefore, we extend the factors investigated in the literature by investigating the impact of behavior change intentions in the fitness domain on the perception of gamification elements in Chapter 4 and its effects on behavioral and psychological measures in Chapter 5.

2.5 Subtle Assessment and Prediction of Personal Characteristics

In this section, we present past research which has focused on predicting personal characteristics such as player, user, or personality traits without relying on traditional questionnaires. Such research is relevant for the work presented in Chapter 6, where our goal is to allow for subtle assessments of Hexad user types to enable personalization of the set of gamification elements without detrimentally affecting the gameful experience by asking users to complete questionnaires. First, we will present approaches that focused on predicting personality traits and highlight the potential of usage data from smartphones as a proxy for personal preferences. Next, we provide relevant examples from the gaming and gamification domain, showing the general feasibility of such an approach.

2.5.1 Predicting Personality

Phillips et al. [269] investigated the relationship between personality traits and self-reported usage of mobile phones. The authors conducted a study with 112 participants in which they were asked to fill out questionnaires assessing their mobile phone usage. By using multiple regressions, they found that time spent playing games on mobile phones was associated with low scores on agreeableness. Also, Lane and Manner [192] investigated whether smartphone application usage is explainable by personality using self-reported data from an online questionnaire with 233 participants. The authors found that personality traits explain self-reported mobile phone application use. For instance, they found that extroverts reported greater use of gaming apps and less use of productivity apps. Also, for less conscientious people, communications, productivity, and utilities apps were less important. Participants scoring high on the neuroticism trait regarded travel apps as more important. Using a list of smartphone apps installed on a user's smartphone to predict user traits such as spoken languages, countries of interest or relationship status was researched by Seneviratne et al. [316]. The authors collected data from 200 smartphone users and applied machine learning techniques to infer personal information. They were able to achieve over 90% precision for most user traits by using the list of installed apps and information such as app categories from the Google Play Store. Chittaranjan et al. [80] investigated potential relationships between behavioral characteristics derived from smartphones and personality traits. They collected the smartphone data of 83 individuals over a period of 8 months. Based on their results, the authors conclude that features obtained from smartphones can be used to predict Big-5 personality traits with an accuracy of up to 76%. In a recent study by Stachl et al. [326], smartphone behavioral data such as calls, texting, contact entries, played music, GPS locations, app starts and installations, photo events, and similar data was collected from the smartphones of 624 participants over a duration of 30 days. Overall, the model consisted of 1,821 predictors obtained from smartphone data. The authors used machine learning approaches and were able to show that the Big-5 personality traits openness, conscientiousness and extroversion could be predicted with a precision well above a baseline model, which constantly returned the mean score in the corresponding training set. However, emotional stability could only be partially predicted, and agreeableness not at all.

Instead of relying on smartphone data, Triantoro et al. [344] aimed to turn filling out a validated survey into a fun and engaging experience. They investigated whether Big-5 personality traits can be predicted when using a gamified survey and compared it against a non-gamified survey. In the gamified version, the Big-5 survey items were transformed from Likert scales into binary, gameful decisions while the non-gamified questionnaire used the traditional scale. They conducted an online study with two conditions–the gamified survey and the traditional survey–and 694 participants. Based on a path analysis, the authors found that the Big-5 responses that were assessed in a gameful way could be used to predict the actual Big-5 responses of the traditional survey.

2.5.2 Predicting Player Types and Gameful Experience

Research was also conducted on whether preferences, experience or behavior in games and gameful systems could be predicted. Mahlmann et al. [210], for instance, explored whether player behavior in the game *Tomb Raider: Underworld* could be predicted based on previously collected gameplay data. The authors were interested in predicting whether players would stop playing the game, whether they would complete it, and how much time they would spend playing it. The authors received access to the gameplay data from approximately 203,000 players of the game, of which 10,000 players were selected for an initial study. After cleaning the data (e.g. to remove players who completed the game and started it again), a total of 6,430 players were considered for the analysis. In the remaining set of gameplay data, 30 features were selected to represent player performance in the first level of the game. Another 25 features were added for players who also completed level 2. The accuracy to predict whether players completed the game was 48% when using data from level 1 only and increased to 77% when also adding features from level 2. Also, regarding the completion time, the authors were able to show that it can be predicted significantly better than random guessing. However, given that the best absolute average error was 84%, there was still room for improvement.

Hadiji et al. [140] were interested in predicting player churn in free-to-play games. They received gameplay data from five different free-to-play games over a period of five months from about twenty million play sessions. The authors created a model which formalizes player churn as a binary decision; it predicts whether players have churned or will return. They used the number of sessions, the number of days since the player has signed up for the game, the current absence time, the average play time per session, the average time between sessions, and other features regarding in-game purchases and spending behavior. By employing different prediction classifiers such as naive Bayes and decision trees, the authors show that the aforementioned features can be used to predict churning with high accuracy (the highest F1 score (0.95 for predicting churn; 0.78 for predicting return) was obtained when using decision trees). Regarding the prediction of player traits, Toker et al. [334] investigated whether player types can be predicted based on automatically collected data in social network games. They relied on the BrainHex player typology and collected gameplay data from the Facebook game *Pot Farm*. They asked players of the game to voluntarily complete a survey to asses their BrainHex type and, after cleaning the data, considered 1,899 participants. Besides the gameplay data, personal information was extracted from the participants' Facebook profile, such as music interests. In a preliminary analysis, they solely investigated whether music interests could be used to predict BrainHex player types, without considering any gameplay features. They found some weak evidence that music interests are predictive of BrainHex types.

Instead of relying on social network games, Li et al. [202] investigated relationships between features extracted from user profiles on the gaming platform Steam and preferences for gamification elements. They received data from 60,267 Steam users including their number of games, unlocked badges, screenshots, reviews, guides, friends and more. The authors conducted an exploratory factor analysis on this pool of data and found nine factors which they labeled elites (a factor indicating a user's striving to become the elite of the steam community), achievers (representing a preference towards mastering games), providers (representing users who like to share their artwork and game guides with others), completers (similar to achievers, but focusing more on gameplay than on achievements), improvers (who like to improve games by providing reviews and their thoughts to developers), traders (representing those who like to buy and sell game-related virtual items), belongers (focusing on social interaction) and nostalgists (who like to retain their gameplay memories by e.g. taking screenshots). The authors hypothesized about potential connections to different types of motivations, as introduced by SDT, and gamification elements. Thus, predicting which gamification elements might be suitable for users could potentially be done by considering

Steam profiles of users.

2.5.3 Summary

In this section, we have learned that predicting personality traits based on smartphone data is promising, as multiple studies have found considerable relationships between these factors. Moreover, we have learned that personality can be assessed in a gameful way by using binary choices instead of relying on Likert scale items. Furthermore, we have seen that in-game behavior and even player traits and gamification preferences can be assessed without directly measuring the outcome variable. These results are important for our research presented in Chapter 6 due to the following reasons. First, they show that smartphone data is linked to personality traits. Considering that personality traits are linked to Hexad user types [342], one could assume that smartphone data could also be used to predict Hexad user types. This would allow dynamic adjustment of gamified systems in a subtle way, without asking users to complete questionnaires. This approach of assessing gamification preferences without explicitly querying users is also motivated by research demonstrating that gamification preferences could be assessed by utilizing Steam profiles or by using data from social network games. Second, research has shown that turning a validated questionnaire into a game-like approach has potential and allows for assessing personal characteristics in a more engaging way. This could be very helpful in gamified apps, where designers would like to tailor the experience but at the same time would like to avoid detrimentally affecting the gameful experience. Thus, whether Hexad user types could be assessed in a gameful way is a promising question to address. We will investigate these two aspects – whether smartphone data can be used to predict Hexad user types and whether Hexad user types can be assessed in a gameful way – as part of Chapter 6.

Chapter 3
Gamification in Behavior Change Contexts

In this chapter, we will present results from studies investigating the effect of gamification on motivation and behavior of users across three different behavior change contexts: physical activity, public health, and marketing/advertising. All three of these are commonly studied behavior change contexts, as revealed by a literature review by Hamari, Koivisto and Pakkanen [145].

First, a system encouraging physical activity by using a gamified smartphone application and a public display component is presented in Section 3.2. Here, we will provide insights on how *social relatedness* can be increased by gamified systems and to what extent this increase has an effect on behavioral measures. Second, we present the *Germ Destroyer* system in Section 3.3, which aims to increase hand washing duration in shared bathrooms by utilizing gamification elements. With this system, we provide insights on how *competence-enhancing* feedback can lead to positive effects on motivation and behavior as well as to what extent the gamified system is able to elicit positively valenced emotions, which support intrinsic motivation. Last, we present the results of a study comparing different gamification approaches in the context of online advertisement in Section 3.4. Here, we found that gamification can increase intrinsic motivation in a context in which *autonomy* needs in particular are thwarted. Also, we will see that in this context, participants prefer a gamification concept that utilizes gamification elements supporting the need for autonomy.

By investigating the impact of gamification on behavior and (intrinsic) motivation in concrete behavior change contexts, we are able to provide answers to **RQ1**, i.e. the question of how gamification affects motivation in such contexts. Section 3.2 is based on [14], Section 3.3 on [15] and Section 3.4 on [8].

3.1 Motivation

Gamification is often used to promote behavior change and can be seen as a form of persuasive technology (see Section 1.1.3). However, the fact that the outcomes of gamification in behavior change contexts are mixed (see Section 2.3) calls for a better understanding of how gamification works. In this regard, the application context has been considered as a potential explanation. However, a lack of theoretical grounding (both in the rationale of the studies and the operationalization of relevant variables) in gamification studies has made it hard to compare gamification studies across different contexts and pinpoint how gamification affects motivation. By informing our hypotheses and the dependent variables being measured in all our studies based on SDT, we have a consistent operationalization and thus are able to contribute insights on how gamification affects motivation in different behavior change contexts, which allows us to reason about the influence of the context itself. We considered two health-related behavior change contexts, physical activity and hand washing, since health is the most prominent context in which persuasive technologies are investigated [145]. In addition, we investigate gamification in the context of online advertising, another commonly used behavior change context [145].

Moreover, as we have seen in Section 2.1.2, SDT posits that the context may inherently support or thwart certain basic psychological needs. Therefore, it will be interesting to see to what extent such contextual affordances have an effect on the success of gamification approaches. As we will see, it seems that contexts in which certain needs are inherently thwarted benefit from gamification approaches supporting these unmet needs. For example, it seems that gamification approaches utilizing gamification elements to support the perceived autonomy of users work best in contexts in which autonomy is inherently thwarted (such as when the user is required to consume online advertisements). However, besides reasoning on the impact of context across the studies that we performed, we also contribute important findings within each context, adding to ongoing efforts in gamification research to better understand why and how gamification works and whether gamification has an effect on the motivation and behavior of users in behavior change contexts. By combining qualitative and quantitative study methodologies, laboratory and in-the-wild studies as well as subjective and objective measurements, we are able to contribute a holistic picture on whether and how gamification affects motivation, how gamification is perceived, and whether and why gamification leads to changes in behavior.

3.2 A Gamified Mobile App and a Public Display to Encourage Walking

In this section, we focus on encouraging physical activity. Given that physical inactivity is one of the main health concerns [279], leading to a wide range of health problems, including cardiovascular diseases, obesity and numerous other chronic illnesses [56], research on how gamification could be used to increase motivation and change behavior is important. More specifically, we developed a system consisting of a smartphone application, a public display and wrist-worn fitness trackers. The smartphone application is used to gather step counts from the fitness trackers and used several gamification elements such as social comparison or achievements. The public display utilizes the same gamification elements, but adds an additional layer of visibility and social interaction, since it is installed in a prominent spot in a gym or meeting room. In this study, we were interested in understanding to what extent gamification elements have an impact on *social relatedness*, one of the basic psychological needs (see Section 2.1.2). In this regard, we also wanted to investigate if it makes a difference whether these are either shown in a mobile app or on a public display. Also, we wanted to analyze whether the gamification elements that we used are capable of changing the behavior of users. To do so, we conducted an in-the-wild study with our system. Lastly, we also wanted to gain a better understanding of how social gamification elements are perceived in this context and which potential issues might detrimentally affect the outcomes of gamification by conducting semi-structured interviews.

3.2.1 Concept and System Design

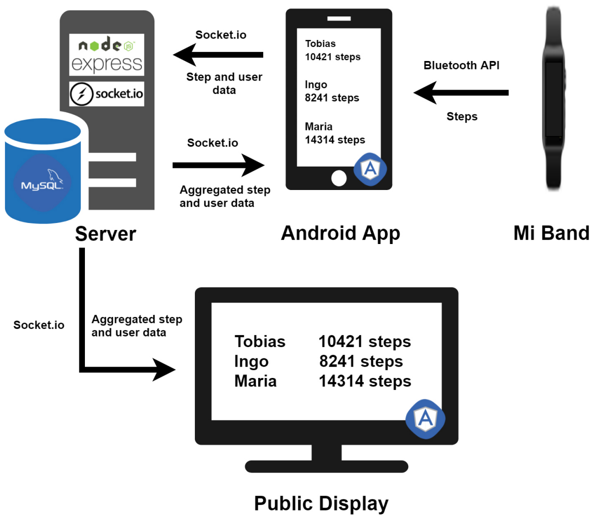


Figure 3.1: System architecture

To investigate the effects of introducing a gamified public display on participants' motivation to walk, we propose a system design consisting of two main components: a mobile app connected to a wrist-worn pedometer and a website which is shown on a public display. Figure 3.1 illustrates the architecture of our system. Both main components share the same user interface elements and game elements. To mediate the communication between smartphone app and public display as well as between different users, a server was implemented using WebSockets. This section explains all these components and illustrates reasons for why we included certain gamification elements in our system.

Smartphone Application

We implemented a native Android application to read steps from the fitness tracker. The app allows users to pick a nickname as well as choose a color that is used to identify themselves (e.g. on the graphs). On the main view (see Figure 3.2b), each user is shown their current step count, their step count of the past seven days on a line chart, and information about whether the connection to the fitness tracker is established. Moreover, users are able to see the current step counts and nicknames of all other users in the app underneath each other together with the time passed since the last step update. Other users' step counts of the past seven days are also visualized on the graph. Moreover, the number of times users reached their daily step goals is shown on a trophy icon next to each nickname. In the top bar, users are able to trigger the synchronization manually by pressing a corresponding button. In addition, a button is available to highlight oneself on the public display, to allow users to easily find themselves.

To keep the step counts as up to date as possible, and minimize issues related to data loss (cf. [151, 278]), the synchronization process to read step counts is triggered frequently. The app notifies users whenever a synchronization process was not successful for the last ten hours. Therefore, the app installs a background service that synchronizes step counts periodically (at least every 15 minutes). In order to avoid users seeing their outdated step counts (which was an issue in [74]), the background service synchronizes whenever the device is unlocked. To read step counts and configure the Mi Band, we developed an API based on the open source app "GadgetBridge"[8], as no official API is available for the MiBand. Our API[9] provides methods to synchronize step counts, disable the LED notifications (to ensure that no feedback is given during the baseline phase of our study) and to read basic information such as the battery level. After a successful synchronization, the aggregated step count for every minute of the day is returned, which is sent to the webserver (distributing the step updates to all clients), whenever the smartphone is connected to the internet. The main view is implemented as a web application. We decided to use a website for the user interface as it allows us to show the same interface on the public display without having confounding factors caused by a different visual appearance, which would have reduced the comparability of the app and the public display.

[8] GadgetBridge: *Freeyourgdget/GadgetBridge*, `https://bit.ly/2T6g3Bt` (last accessed: 2021-12-01)
[9] BandSynchronization: *m-altmeyer/BandSynchronization*, `https://bit.ly/3cmaCoN` (last accessed: 2021-12-01)

Public Display

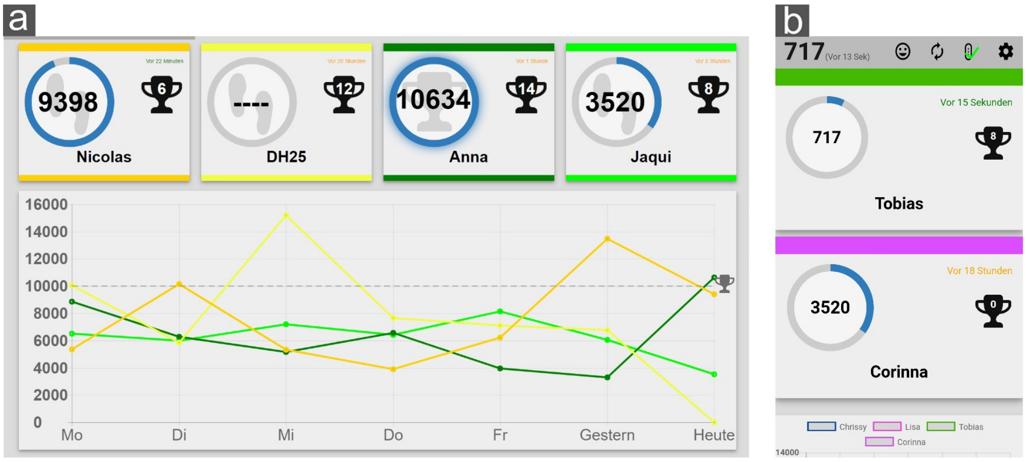


Figure 3.2: User interface of the public display (a) and the mobile app (b)

The public display shows the same website as the mobile app, i.e. the interface looked similar and the game elements were the same. The only difference is that four users are shown next to each other on the public display instead of underneath each other as in the mobile app (see Figure 3.2). For each user, the same data as in the smartphone application is shown (step count, progress towards the daily goal, total goals reached and time since last updated). In line with the smartphone application, a graph showing the step history of the past seven days of each user is shown under the user panels. The public display cycles automatically through all users (every 15 seconds four different users are shown), as no direct interaction with the display itself is possible. We used a Raspberry Pi 3, connected to a screen, as public display, which, after booting, opens the same website as the smartphone application in a full-screen browser window.

Server

The server is used to host the web page, handle real time communication with the app and the public display and store the step- and user data. Whenever the smartphone application sends new step counts to the server, the step counts are parsed and stored in the database. Afterwards, the step data is distributed to all clients (smartphone application and public display).

Fitness Tracker

We use the Xiaomi Mi Band 1[10] for several reasons: First, the Mi Band does not have a display, which is important for the baseline phase of our study. Second,

[10] Mi Global: *Mi Band - Understand your every move*, `https://bit.ly/3zfhZZl` (last accessed: 2021-12-01)

it has a battery life of at least 30 days, meaning that participants do not need to recharge the battery, which could lead to data loss or participants forgetting to wear the band. Moreover, the Mi Band is water-resistant and thus can be worn even when swimming. It is very lightweight (an important requirement participants established in [89,204]), able to store step data up to ten days (which further reduces data loss) and was shown to be precise compared to other commercial fitness trackers [115,203]. Finally, the Mi Band is cheap [115], having a price of around 14$.

Gamification Elements

Based on past research, we included the following gamification elements (see Figure 3.2):

Journaling Having a history of past activities was a design recommendation from Consolvo et al. [87]. This is confirmed by Ashford et al. [28], stating that interventions giving feedback on past performance and the performance of others lead to the highest levels of self-efficacy, which is beneficial to encourage physical activity. By recording their physical activity, users learn how much they usually walk. A history of past activity reminds them how much they walked in past days, and challenges them to reach the same step count again. Our system provides this by recording the user's steps, and allowing the user to see their daily step count, as well as their history for the past seven days. Inspired by [74], we chose graphs to display the step counts of the past week for every user. We limit the history to one week, in order to make the display more easily comprehensible, and to make the users focus on the current week, thereby reducing the chance of being demotivated by poor performances in past weeks.

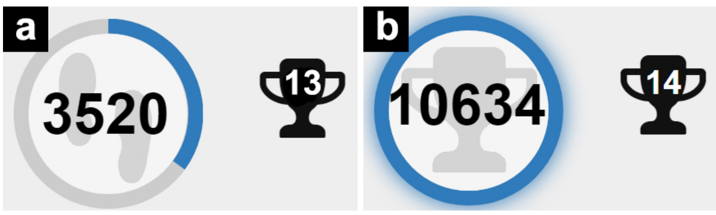


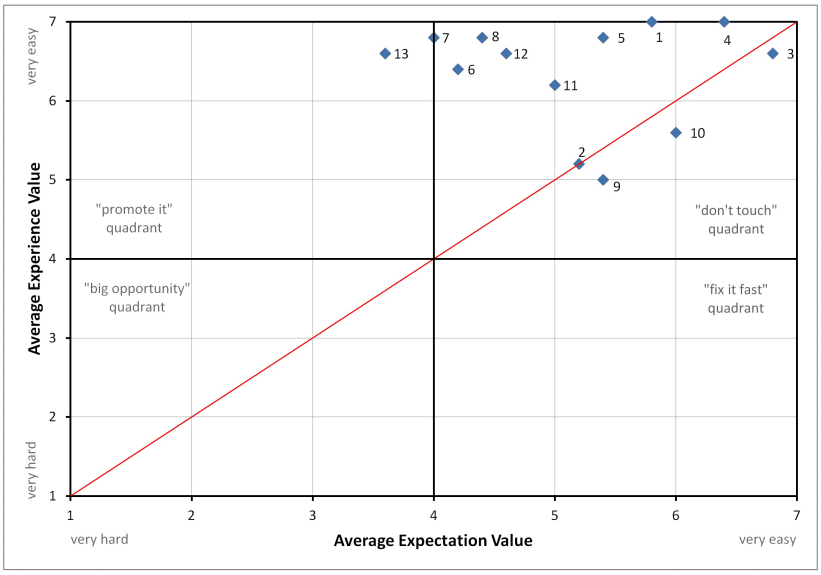
Figure 3.3: User panel showing a users' step count, the progress towards the daily goal and how often a daily goal was reached so far (a) and visualization of a user who has reached the goal (b)

Achievements, Goals and Progression According to [87], the activity level performance is very important to users, i.e., most participants were motivated by knowing their progress towards the daily goal and appreciated receiving recognition for it. The importance of daily step goals is also emphasized in [74,88,89]. In line with [74], we decided to use 10,000 steps as the daily goal to be reached. This appears to be reasonable, as it is the lower bound for healthy adults to be considered "active" according to Tudor-Locke et al. [346]. In addition, Le Masurier et al. [194] found that individuals who achieve 10,000 steps/day are more likely to meet the physical activity guidelines promoted by the Centers for Disease Control and Prevention and the American College of Sports Medicine [266]. To also visualize the progress towards a goal, we decided to use a circular progress bar, in the middle of which the current step count is shown (see Figure 3.3a). As soon as the user reaches the step goal, a blue glow is added to the circle and a trophy is shown. Next to the trophy, the number of days on which the user has reached the step goal is displayed (see Figure 3.3b).

Social Comparison Social comparison means that users are given information about how well their performance compares to the performance of others, and has the potential to motivate people to perform a target behavior [126]. Since facilitating social comparison is another recommendation from Consolvo et al. [87], our system employs it by showing the step history of all users to all other users and those coming by the public display. Besides triggering the positive effects of social comparison, we also expect that people start socializing around the data at the public display, as was reported in [74]. We decided to show the current daily step counts and the daily step counts of the past seven days to all participants to encourage social comparison and social influence, which were shown to be motivators in the physical activity context [28, 87, 127, 343]. As direct competition may also have detrimental effects on improving step counts [146], especially when users perform differently and have different abilities [79], we decided not to use a ranking. Instead, participant's step counts were shown in random order.

Usability Test

To ensure that our system is intuitively usable and to avoid inducing confounding variables related to a bad usability in the evaluation, we performed a usability test (5 participants, as proposed by [253], age 25 on average, 2 self-reported as female, 3 as male) using the German version of the System Usability Scale (SUS) [60], a think-aloud approach [252] and the "Expectation Measure" method [6]. In the latter, users are first asked how hard they expect a task to be in the system (without knowing it) and later have to complete the task within the application and rate how hard they found the task to actually be. The core idea of this method is that some tasks are inherently harder perceived as others but that this perception is often idiosyncratic and thus by comparing expected and actual difficulty, these differences can be taken into consideration. Also, by comparing the expectation to the actual effort, one can find what elements work well, and which need improvement. We identified 13 tasks to cover all functionalities of our system (see Figure 3.4 for a list of all tasks), which had to be rated before interacting with the system and afterwards. The users did not receive a compensation and had prior experience with using smartphones (M=5.2, SD=1.22) and computers (Mean 5.2,



Tasks:
1 Register yourself within the app.
2 Find out whether you are currently connected to your fitness tracker.
3 Find out how many steps you have taken today.
4 Find out how recently the step count was read.
5 Read out your step count from the Fitness tracker again.
6 Estimate how many steps a user took on Saturday.
7 Find out how recent the step count of a user is.
8 Change your username.
9 Try to find yourself on the public display by using the app.
10 Find out, what the desired goal is.
11 Find out how often a user has reached his daily goal.
12 Find out which user has reached the daily goal most often.
13 Find out whether a user reached his desired goal on Wednesday.

Figure 3.4: A scatter graph visualizing the average expectation- and the average experience rating of every task. Tasks are represented by 13 numbers with each task being explained beneath the graph.

SD=1.78) (rated on a 7-point Likert scale). One of them had used fitness trackers before.

The SUS revealed an average score of 87.5 (SD=5.90), indicating that the overall usability of our system is "excellent" [33]. The Expectation Measure revealed that the average expected difficulty for each of the tasks participants had to perform was either higher than or roughly as high as the experienced difficulty (all tasks are above or very close to the red line in Figure 3.4). Most of the tasks are placed in the upper right "don't touch" quadrant, with task 13 even being in the upper left "promote it" quadrant, since it was expected to be more difficult than it turned out to be. Task 2, 3, 10, and 9 were on or slightly below the red line, meaning that they were perceived to be (roughly) as hard as expected. Overall these results further support that our system provides a very good usability, as all tasks were in the quadrants that do not require adaptations of the system. Nevertheless, we found some minor issues through the think-aloud approach (e.g. language inconsistencies or problems interpreting icon symbols), that were fixed for the in-the-wild evaluation.

3.2.2 User Study

The goal was to investigate the effects of introducing a gamified public display in addition to a gamified mobile app on motivation and behavior. We decided to run a within-subject study: After the baseline phase, in which no feedback was given at all (i.e. participants were not able to see their step counts), an app-only phase followed, in which participants had access to the mobile app, as presented above. Afterwards, the public display was introduced in addition to the app. We tried to find evidence for the following hypotheses:

H1: Step counts are higher in the app-only phase compared to the baseline

H2: Step counts are higher in the public display phase (the public display is present in addition to the app) compared to the baseline

H3: Step counts are higher in the public display phase compared to the app-only phase

H4: Introducing the public display has a positive effect on social relatedness and thus intrinsic motivation

H1 is motivated by the fact that receiving feedback and using game elements should motivate participants to reflect on their behavior and challenge them to beat their own step counts [126, 146]. Even though there is related work in which introducing game elements did not increase step counts significantly (e.g. [204,230]), we expect that the app, showing step counts of other participants, leverages participants' natural drive to compete [126] and thus leads to increased motivation as was reported in [87,127,343]. The motivation for **H2** follows the same reasoning as for **H1**, since the game elements and the interface are similar. However, social recognition, social pressure and social influence may additionally encourage participants [126], motivating **H3**. Moreover, the socialization around step data, which was reported in [74], may have positive effects on intrinsic motivation, as the feeling of social relatedness might be increased, motivating **H4**. We assume this to have an influence on participants' motivation since social relatedness is one of three dimensions of intrinsic motivation according to the Self-Determination Theory (SDT) [298].

Procedure and Method

Since we wanted to reduce chances of participants forgetting to wear their fitness trackers (which is a major problem in similar studies [151, 278]), we decided to conduct the study over a timespan of four weeks so that participants were not required to charge the battery of their trackers. The duration of the baseline phase, in which the mobile app did not show any step data nor game elements, was one week. This was followed by a one-week app-only phase, in which the app automatically unlocked and all functionality and game elements were active, and a two-week-long phase in which the public display was introduced in addition to the app (public display phase). The public display phase was longer, since we wanted to ensure that participants came across the public display multiple times.

We recruited members from a fitness center and from the student council who regularly (at least twice a week) visit the space in which each of the public displays were placed (one was placed in a prominent spot in the fitness center (approximately 100 visitors per day) and the other one was placed in a prominent spot in the faculty room (frequently visited by students)). In the baseline and app-only phase, the display was turned off. The two populations (fitness center and student council) were separated, i.e. each public display only showed step data from participants of the corresponding population. To investigate **H1**, **H2** and **H3**, step counts were measured and participants were required to wear the fitness trackers throughout the whole study. The fitness trackers did not give any kind of feedback regarding participants' steps.

Before the study started and between the study phases, participants were required to fill out online questionnaires. These included demographics (only once before the study) and the German version of the Balanced Measure of Psychological Needs scale (BPMN) [317] by Neubauer and Voss [246] (after the app-only and the public display phase). This scale can be used to determine the grade to which the users' needs for competence, autonomy, and relatedness are fulfilled for a specific context. We used it to compare the effects of the mobile app and the public display on intrinsic motivation, as we expected the social relatedness to be higher when the public display is present (cf. **H4**).

The questionnaires moreover included the short German version of the International Physical Activity Questionnaire ("IPAQ-short") [37]. This validated survey was shown to be a reliable method for measuring moderate physical activity (including walking) [208]. It uses the Metabolic Equivalent of Tasks (MET) [37] to calculate physical activity. The IPAQ was used to validate the step data from the fitness trackers and check whether participants subjectively support the step counts measured, giving additional information to investigate **H1**, **H2** and **H3**. Moreover, we logged the number of app starts and asked participants how often they were in proximity to the public display in the online questionnaires to infer how frequently they interacted with the system.

After the intervention, we met with participants to conduct semi-structured

	Baseline	App Only	Public Display	sig.
Step Count	M=9976.88 SD=3345.08 Mdn=9254.50	M=10624.13 SD=3376.29 Mdn=9303.50	M=11603.08 SD=3614.22 Mdn=10991	$p < .01$*
MET-min / week	M=4594.33 SD=4108.48 Mdn=2967	M=5034.33 SD=4079.14 Mdn=3915	M=5672.33 SD=4390.98 Mdn=4330.50	$p < .05$*
BMPN Autonomy	-	M=32.92 SD=2.88 Mdn=33	M=34.08 SD=3.37 Mdn=33.50	-
BMPN Competence	-	M=28.50 SD=3.61 Mdn=28.50	M=29.67 SD=5.48 Mdn=29.50	-
BMPN Relatedness	-	M=27.58 SD=2.84 Mdn=27	M=30 SD=3.98 Mdn=28.50	$p < .05$**

Table 3.1: Overview of the main results for every study phase. MET refers to the IPAQ Metabolic Equivalent Times [37], BMPN to the Balanced Measure of Psychological Needs scale [317]. * Friedman ANOVA was performed and Wilcoxon signed rank tests for post-hoc analysis. ** Wilcoxon signed rank test was used.

interviews. This was done to learn more about reasons for the effect of the public display in this setting. The interviews were directly transcribed and the transcripts were coded using a directed content analysis [156], i.e. we went through the transcripts to find themes related to the perception of the public display and to find reasons why it encouraged people to walk more. We then counted for each theme how often it was mentioned and exemplary selected statements related to each theme [96]. The following questions were asked (additional questions were asked when the answers revealed interesting aspects): *Do you think that the public display influenced the number of steps you walked? Why do you think so, especially considering that the app showed the same data?*; *Were there discussions about the public display? Did other people ask you questions about the display or your step data?*.

Results

We recruited 16 participants, of which we had to remove four (one participant lost the band after the baseline phase, from two participants we only had step data from one day in the baseline phase due to technical problems with the band, and one participant reported that the band stopped working after two weeks). Of the remaining 12 participants (8 from the fitness center, 4 from the student council) having an average age of 29.58 years (SD=5.93, Mdn=27, Min=24, Max=42), 6 self-reported as female, 6 as male.

We removed single days from our sample on which the synchronized step data for a user was not complete, i.e. we inspected the timestamps of the data and made sure that there was no data missing. Therefore, we needed to exclude 67 days overall (5.58 days per participant on average). The main reasons for why step data was incomplete were that either participants forgot to wear the fitness tracker the whole day or because the fitness tracker itself had technical issues. Thus, 269 single days were analyzed: 70 days in the baseline phase (first week), 74 in the app-only phase (second week) as well as 125 in the public display phase (third and fourth week). As revealed by a Shapiro-Wilk test, step data, data from the IPAQ-short and the "relatedness" subscale of the BMPN were not normally distributed. Therefore, if not otherwise specified, Friedmann tests were performed and Wilcoxon signed-rank tests with a Bonferroni-Holm correction for post-hoc analysis. Main results are shown in Table 3.1.

Interaction with the System Participants frequently opened the mobile app in the intervention phases (2.46 times/day in the baseline (SD=3.56, Mdn=1), 4.38 times/day in the app-only (SD=3.44, Mdn=3.75), 3.75 times/day in the public display phase (SD=1.86, Mdn=3)). These numbers differed significantly ($\chi^2(12)$=8.977, p=0.01) between the baseline and the app-only phase (Z=-2.18, p=0.029) but not between the app-only and the public display phase (Z=-0.36, p=0.721). This indicates that the gamification components, which activated after the baseline phase, increased participants' interest in the smartphone app. The fact that there is no significant effect between app-only and public display phase is not surprising, as the app showed exactly the same as in the app-only phase. During the public display phase, participants were in proximity to the public display 5.20 times (SD=3.20, Mdn=4) on average. Overall these results suggest that participants actively and frequently interacted with the mobile application and that it was possible to be influenced by the public display, as they were in its proximity multiple times.

Step Counts and Walking Behavior On average, participants walked 9976.88 steps per day in the baseline phase, 10624.13 in the app-only phase and 11603.08 in the public display phase (see Table 3.1). Each participant's step count for the study phases is shown in Figure 3.5. A Friedmann ANOVA among repeated measures showed that these step counts differed significantly between the study phases ($\chi^2(12)$=12.17, p=0.002). The number of steps was statistically significantly higher in the public display phase than in the baseline (Z=-2.90, p=0.004, r=0.48) as well as significantly higher in the public display phase than in the app-only phase (Z=-2.43, p=0.015, r=0.41). However, no significant difference was found between the baseline and the app-only phase (Z=-1.65, p=0.099). While these results support **H2** and **H3**, we could not find supporting evidence for **H1**. The

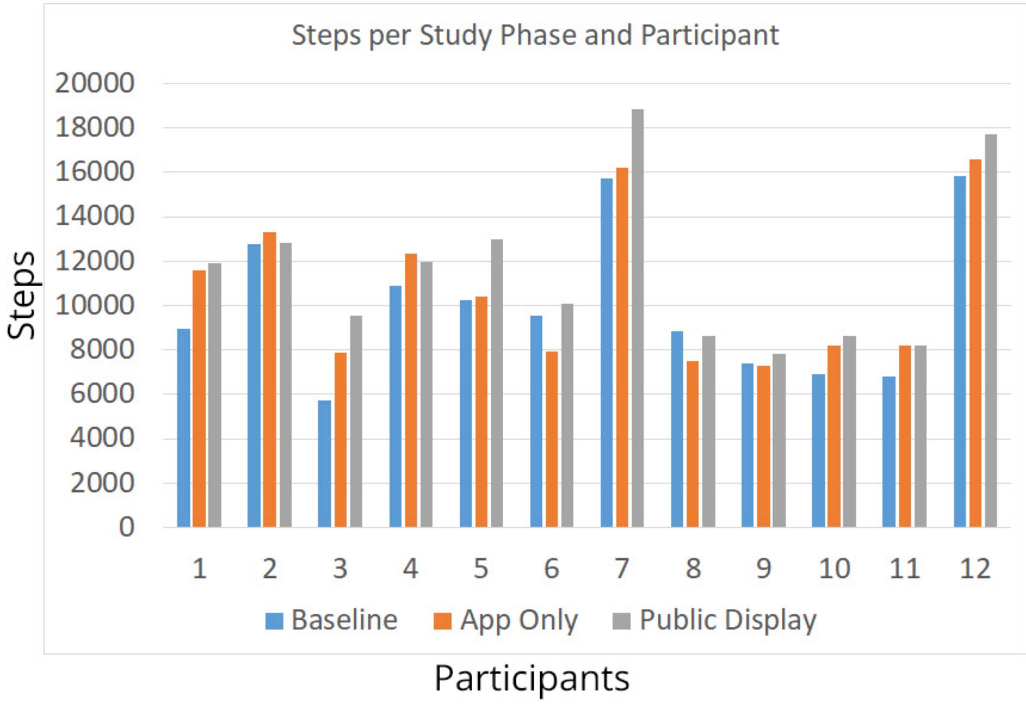


Figure 3.5: Step counts in each study phase. The x-axis represents participants, the y-axis step counts

results show that the introduction of the public display has a positive effect on the walking behavior of participants, as the number of steps is significantly higher in the public display phase than in the app-only phase and the baseline. Finding that the mobile application alone does not lead to a significant increase in step counts is unexpected, but also in line with previous research (e.g. [204]), although there exist research finding positive effects [87].

The self-reported physical activity (which was assessed using the IPAQ-short) validates the aforementioned effects (results are shown in Table 3.1). The MET-min/week values differed significantly throughout the study phases ($\chi^2(12)$=8.71, p=0.013). Post-hoc tests indicated that the MET-min/week values were significantly higher (Z=-2.67, p=0.008, r=0.45) in the public display compared to the baseline phase. These were also significantly higher (Z=-2.29, p=0.022, r=0.38) in the public display than in the app-only phase. In line with the actual step counts, there was no significant difference between baseline and the app-only phase (Z=-1.41, p=0.16). Finding the same effects between the study phases as for the step counts, the IPAQ MET-min/week values support **H2** and **H3** even further, i.e. that the public display had an effect on the walking behavior of participants, as not only step counts but also self-assessed physical activity shows significant positive results.

Balanced Measure of Psychological Needs Average scores for each subscale of the BMPN scale are shown in Table 3.1. To test for significant differences between the app-only and the public display phase, we used paired samples t-tests for the competence and autonomy subscales (as this data was normally distributed) and a Wilcoxon Signed-Rank Test for the social relatedness subscale (as this data was not normally distributed). We did not find any significant effects for the competence ($t(11) = -1.13$, $p = 0.28$) nor the autonomy subscale ($t(11) = -1.19$, $p = 0.26$). However, the social relatedness increased significantly in the public display phase ($Z=-2.25$, $p=0.024$, $r=0.38$), supporting **H4**. This indicates, as expected, that the public display encouraged socialization and led to higher feelings of social belonging, which in turn might have affected intrinsic motivation [298] and thereby might have been the cause for the increase in physical activity.

Semi-Structured Interviews When asked whether the public display influenced the number of steps walked, all but two participants agreed. Asked for reasons, two main themes emerged:

- **visibility for outsiders**, mentioned by ten participants
- **being confronted about the own performance by others**, mentioned by seven participants.

Concerning the first theme ("visibility for outsiders"), participants stated that they were additionally motivated because the public display was visible to people without the mobile app, thus reaching a broader audience: *"I was walking more steps because other people could see it. I agree that this was also true for the app, but the display was visible to more people, even those that do not have a fitnesstracker nor the app"* (P3). Related to this, the app was perceived as a closed system, in which participants were among themselves, while in contrast the public display was visible to everyone, as P5 explains: *"On the one hand they [users in the app] were in the same boat and on the other hand I didn't know all of them. But the display was visible to more people and I knew some of them"*. Here, it becomes also visible that the relationship to other users might play a role in measuring how effective a public display using social comparison is in motivating users.

Concerning the second theme ("being confronted about the own performance by others"), 7 participants stated that the fact that people were able to directly confront users of the system (by knowing their nickname or asking them about their performance directly in front of the public display), when they are in proximity to the public display, motivated them (*"I think the public display led to increased pressure. You always knew that whenever you go to the box [the fitness center], it could be that someone would confront you with the number of steps that you have. Having this in mind, I didn't want to be exposed as performing particularly worse than others"* (P10), *"I think the possibility that other people could approach me was a big reason to care about my steps. Especially knowing that I have to justify myself, when I have a low step count"* (P2)). Related to this, 4 participants explicitly stated that they wanted to avoid being talked about when having low step counts (*"I told myself that I need to walk more to not attract attention and thus have other people talking about me"* (P6)).

From the answers to the question whether there were discussions about step data, we learned that only two participants actually were confronted directly with their performance (P4 reported that a coach told them that they are keeping an eye on their steps when they are not around and P2 stated that they were confronted with their steps multiple times, especially when their steps were low, which increased their motivation to walk). However, all participants reported that the display was a conversation starter (which is in line with the finding that social relatedness increased significantly, cf. **H4**). Other people (not taking part in the study) asked questions about how the system worked and were joking about reasons for why participants performed well.

We also asked the two participants who reported that the public display did not affect their performance (P9 and P11) for reasons. P11 stated that they had the feeling that they walked more steps after the app was introduced. However, it was not possible for them to further increase their step counts because of their occupation and their daily routine. They also reported that they felt bad seeing other players performing better than them (*"After some time I did not pay much attention to the display anymore since it was just not possible to walk as many steps as they did, because of my job and so on. So I stopped paying attention to the display"*). P9 told us that they do not need public recognition to perform well. However, they liked the socialization around the public display, which they perceived to be encouraging for other participants (they never heard of someone being blamed because of a low step count).

3.2.3 Discussion and Limitations

The study revealed that introducing a gamified public display in addition to a mobile app has positive effects on users' walking behavior (**H2**, **H3**) as was shown by participants' step counts and the IPAQ MET-min/week values. Since the public display and the mobile app shared one interface having the same game elements, it is very probable that the public display led to this increase in step counts. As hypothesized (**H4**), one reason for this is the higher feeling of social relatedness, increasing the chance to positively affect intrinsic motivation [298]. This was supported by participants indicating that the public display encouraged socialization and increased their motivation to walk because of a higher visibility for outsiders and because they could be directly confronted with their performance by others. Related to this, participants reported feelings of social recognition and accountability. Although these findings are partially more related to peer pressure than socialization, they align well with the increased social relatedness, as the feeling of belonging increases the power of peer influence and promotes peer conformity [58]. The motivational effect of these aspects is also explainable through Fogg's [126] principle of recognition, stating that public recognition increases the likelihood that subjects adopt a target behavior.

In contrast to **H1**, the mobile app alone did not lead to a significant increase in step counts compared to the baseline. Given previous literature, this is not surprising, as research about using step trackers to motivate people being physically active is inconclusive: While some investigations were able to show effects of pedometers on daily step counts [347], others did not find effects [204]. Reasons for not finding an effect might be related to observer effects in the baseline phase [232], i.e. the fact that participants knew that their step counts were recorded and thus increased them in the baseline [126]. As of our study duration, we also cannot rule out novelty effects (which might also have had an influence on the walking behavior in the baseline and/or in the intervention phases). Furthermore, we are aware that there might have been ordering effects in our within-subjects design, as we decided against counterbalancing the study phases. This was a conscious decision due to two reasons: First, installing the public display in a prominent area would have also affected participants in different study phases, as the public display would have been visible for them, too. Second, it was shown that removing feedback mechanisms (the public display) has detrimental effects to participants in the gamification context [146].

Another limitation is that we cannot determine with certainty whether knowing that the public display is installed, or its actual physical presence, led to the increase in step counts. However, qualitative data from the interviews and the increase in social relatedness indicate that the public display was a reason for communication, suggesting that its physical presence was the deciding cause for the effects. Nevertheless, an interesting research direction is to investigate whether just the knowledge that one's own performance is publicly exposed leads to similar effects. Additionally, we acknowledge that our sample size was rather low, which might be an explanation for why we could not find supporting evidence for **H1**. We computed the post-hoc power of our study and found it to be 0.56 for the step count tests, which is still acceptable according to [349], but also shows that validating our findings with more participants is important.

Since we focused on the effect of introducing a public display in addition to a mobile app, it is unclear whether a public display alone would have led to similar results. We decided to investigate the combination of a public display and a mobile app since having a public display alone would have resulted in removing feedback, which might have negatively affected participants, leading to skewed results [146]. As the public showing of the data was the only element that changed, and due to the increased perceived social relatedness, it seems that showing the data publicly led to the increased step counts.

3.2.4 Contribution to Research Questions

The findings of this study contribute to **RQ1**. In the specific context of this study, the need for relatedness seems to open the door for gamification to succeed. Satisfying this need seems to have increased participants' motivation to change their behavior, i.e. to walk more. Although relatedness seems to be particularly important, the gamification elements that were used were also aimed at supporting perceived competence. However, we could not show that gamification actually helped to satisfy competence needs in this case. Potentially, in this specific context, the needs for competence and autonomy were already satisfied inherently to a certain degree: Participants were actively engaging in sports and already had a high level of fitness. Their average step counts in the baseline were already almost twice as high as the step counts of an average adult in the U.S. [35]. Thus, the step goal promoted by our gamified system was not particularly challenging to reach. These two facts – that participants were already fit and that the goal promoted by our system was not hard to reach – might well explain why competence might not have been enhanced by our system and might already have been satisfied. Also, the need for autonomy might have been inherently satisfied: Participants were fit, which suggests that they had already done sports for some time and might already have internalized the importance of sports for themselves to a certain degree. Thus, their type of motivation to do sports or to go to the gym might be autonomous rather than controlled, potentially diminishing the need for further autonomy support through a gamified system. In contrast, as was revealed by the semi-structured interviews, the public display increased social interactions both with other people attending the gym as well as with the trainers working at the gym. This might have facilitated a stronger sense of social belonging and reciprocal care, which might have been previously unmet needs, potentially due to the anonymous nature of a gym.

However, we also need to consider that negative effects of the system were found. The absence of an effect of the gamified smartphone application suggests that the gamified mobile app did not support the motivation to increase step counts for all users. Indeed, when looking at the individual step counts in Figure 3.5, it can be seen that some participants even decreased their step counts in the app-only phase, compared to the baseline. This suggests that there were interpersonal differences in the perception of the gamification elements used in the smartphone application. Moreover, the fact that daily step counts were shown publicly was not perceived well by all participants. As we have seen in the semi-structured interviews, some participants felt pressured to keep up with others, potentially shifting their motivation towards less autonomous and more controlled types, ultimately undermining intrinsic motivation. Even worse, one participant reported feelings of incompetence, i.e. they said that they were not able to keep up with others due to their daily routine. Such feelings of desperation might not only shift motivation to more controlled forms but even lead to amotivation. Therefore, it is important to also consider the individual person in order to tailor gamified systems to their motivational orientations and preferences.

Since the performance in this context is limited by human capabilities, i.e. the fitness level of participants, we decided to investigate similar gamification elements in a context, in which such limitations are not as prevalent: encouraging people to wash their hands for a long enough time. In this context, most people have the same chance to make progress and reach the goals established by gamification elements, since they can freely decide to adhere to the instructions given by the gamified system. We will introduce this system and present findings of two user studies in the following section.

3.3 A Gamified System to Increase Hand Washing Duration

In this section, we focus on the *Germ Destroyer* system. This gamified system aims at increasing hand washing duration in (shared) bathrooms and thus can be seen as a behavior change support system in the (public) health context. Similar to physical activity, this context is among the most relevant in persuasive technology research [145]. Encouraging people to take enough time to wash their hands effectively is important, since people generally do not wash their hands for long enough [49], even though it has been shown to be one of the most effective ways of preventing the spread of diseases [178].

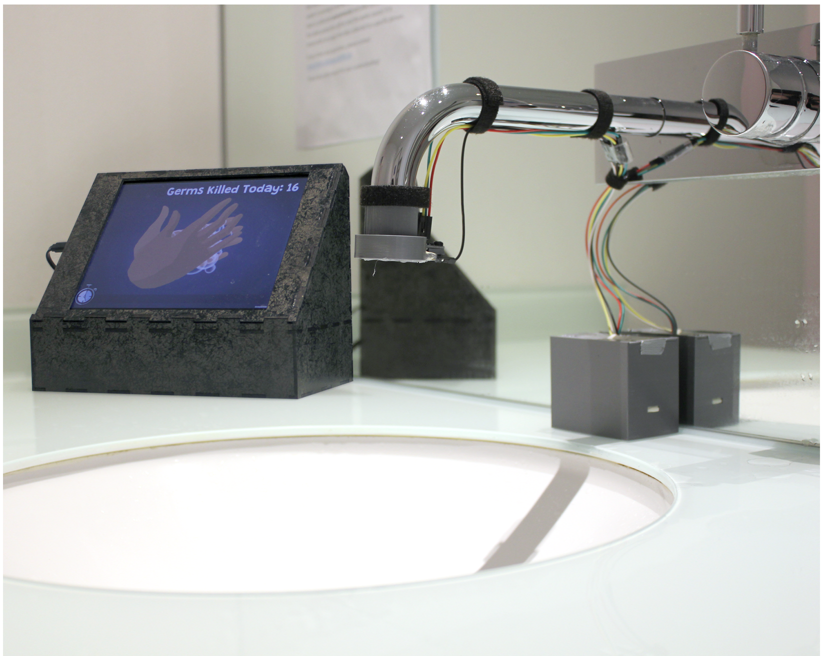


Figure 3.6: Our system during the in-the-wild study: a 3D-printed measuring unit which can be mounted on the tap and a gameful mobile application. The measuring unit detects whether water is running and whether hands are present, and sends this information to the mobile application.

We conducted two user studies: a laboratory study to investigate the perception and the effect of our system on motivation, affective experiences and enjoyment as well as an in-the-wild study to analyze the impact of *Germ Destroyer* on people's hand washing behavior. Since it may be hard for people to estimate the exact amount of time they spend washing their hands, and since they might not have enough knowledge about how long hand washing should take, we used gamification elements providing informative, positive feedback, which should educate users, increase their *perceived competence* and thus enhance their motivation to perform the target behavior.

3.3.1 Concept and System Design

We realized an open source, unattended system which is installed and evaluated in a shared bathroom. Encouraged by the success of gameful, persuasive systems in other domains, we hypothesize that our system affects people's behavior positively. Based on the related work presented in Section 2.3.2, we deduce the following design implications for our system:

D1: Focus on hand washing rather than hand disinfection [322]

D2: Encourage a washing duration of 20 seconds [72, 166]

D3: Avoid camera-based approaches and any sensors that can record sound or pictures [233]

D4: Ensure that the system integrates well with its surrounding and is self-attended [90]

D5: Use normative feedback for behavior change [97, 234]

D6: Promote disgust playfully to change behavior [97]

D7: Engage users by using gameful elements [14, 196, 306]

D8: Provide feedback about the washing duration [97, 306]

In this section, we describe the concept and implementation of our system. It consists of two parts – the sensing device and the gamified application – communicating via Bluetooth Low Energy ("BLE"). The sensing device can be mounted on the water tap such that water flows through it. The sensing device also has a base station which is responsible for sending hand washing states via BLE to the mobile application, running on an Android tablet device.

BLE Sensor Device

Since we focus on encouraging a sufficient hand washing duration (**D1**, **D2**), we conceptualized a sensing device which can be easily placed on the water tap,

Figure 3.7: Different screens of the gameful mobile application. a) The screen visualizing germs is shown whenever water is running. b) If users remove their hands, germs are shown as getting angry to encourage the user to keep on washing. c) During hand washing, germs are destroyed and generate points. d) After destroying all germs, the app shows an animation illustrating clean hands and all destroyed germs are added to the daily number of killed germs.

without interfering with the users' intended actions (**D4**, see Figure 3.8a). In addition, we have a base station which has a wired connection to the sensing device on the water tap. The base station holds the microcontroller as well as the battery (see Figure 3.8b). We designed the sensing device such that it fits on most water taps (see Figure 3.6) and can be further adjusted by tightening or loosening a screw located at its edge. All housing parts of our system are 3D printed. Figure 3.8c shows the 3D-model of the sensing device which is mounted on the tap.

To measure whether water is flowing or not, we used a capacitive sensing approach in the inside of the sensing device's tube (capacity changes as water flows through the tube, determined by a tensioned wire connected to a 1 MOhm

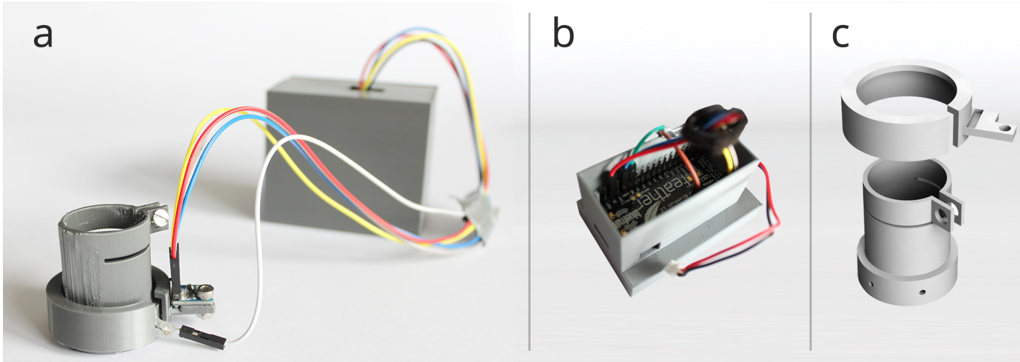


Figure 3.8: Overview of the measuring device. a) The 3D-printed measuring unit which can be mounted on the water tap. b) The microcontroller in its case. c) The exploded view of the 3D-model.

resistor). On the outer side of the sensing device, we installed a VL6180X time-of-flight distance ranging sensor to detect whether hands are present or not. The sensor uses a tiny laser source and is able to measure distances between 5mm and 100mm, which perfectly fits our needs. Since only these two sensors are needed, we are able to ensure a completely anonymous tracking process (**D3**), as no personal information can be detected. In addition to building the hardware, we also implemented a firmware to handle, interpret and transmit the measured sensor values to the gamified application (or any other BLE-enabled device). Our device transmits four different states using BLE characteristic notifications:

Idle: No water is flowing and no hands are present

Water Flow: Water is flowing and no hands are present

Hand Washing: Water is flowing and hands are present

Hands Only: No water is flowing and hands are present

Besides sending these notifications, the firmware also allows receiving commands in order to adjust the sensor thresholds (e.g. the threshold which is used to decide whether hands are present or not can be easily set by sending a corresponding message to the device). This firmware is installed on the Adafruit Feather M0 Bluefruit LE module[11]. To ensure a maintenance-free runtime (**D4**), our device also has a 2000 mAh battery, allowing it to run for 200+ hours. It can be easily recharged via a micro USB outlet. Besides sending the aforementioned states, the firmware can also be configured to send the raw sensor readings. To allow fellow researchers to use our system, we published the firmware, API documentation,

[11] Adafruit: *Adafruit Feather M0 Bluefruit LE*, `https://bit.ly/3gh4cIM` (last accessed: 2021-12-01)

Figure 3.9: The idle-screen shows a hand washing animation to indicate that the system can be interacted with by washing one's hands.

circuit scheme, a list of electronic components and the 3D models on GitHub[12]. With a price of less than $50, the measurement device can be seen as a low-cost solution.

Gamified Mobile Application

We conceptualized and implemented a gamified mobile application called "Germ Destroyer". Depending on the state transmitted by the sensing device, the app shows different screens and provides different feedback to the user. When water is flowing, the app visualizes nasty germs (see Figure 3.7a) to illustrate and represent the microbiological contamination. We used germs as virtual characters to promote disgust playfully since it was shown to lead to positive effects on hand washing behavior in public restrooms, encouraging people to reflect on proper hand washing (**D6**). Also, using virtual germs may highlight the impact of washing hands on their real-world contamination, which may enhance the persuasive power of our system [126]. Once users start washing their hands, i.e. put their hands under the tap while water is flowing, soap bubbles and washing animations appear in the app. Also, a progress bar next to a clock icon is shown, providing feedback on how long hands should be washed (**D8**). The progress bar fills at a constant rate until 20 seconds have been reached by the user. With increasing progress, the germs start to move and shake faster as well as change their facial expression from being nasty to being afraid to be destroyed. The closer the user is to the target duration, the more germs are being washed away and killed. Whenever a germ is killed, an auditory feedback is given and an animation adding a point to the total score is shown. If users stop washing their hands before destroying all germs, the remaining germs start looking angry (see Figure 3.7b) to motivate users to keep on washing hands. The number of germs

[12] Germ Destroyer: *m-altmeyer/GermDestroyer*, `https://bit.ly/3w2g3RK` (last accessed: 2021-12-01)

killed by the user through washing is counted and visualized in the app. During the 20-seconds-long hand washing phase, eight germs are being killed (one germ each 2.5 seconds). Figure 3.7c shows the screen when washing hands. To further encourage users to meet the recommended duration, the total number of killed germs by all users today is visualized in every screen of the app, thus providing normative feedback to the users (**D5**, **D7**). Once the user has finished washing their hands, the app visualizes how many germs have been killed by the user and adds these to the total number of killed germs per day. Also, an animation is shown visualizing clean hands indicated by sparkles and supported by a positive auditive feedback (see Figure 3.7d). The idle screen visualizes the total number of killed germs and repeatedly shows an animation to indicate that the system can be interacted with by hand washing, as can be seen in Figure 3.9. Independent of the current screen, an overlay visualizes whether water is running and whether hands are detected or not (indicated by an icon showing washing hands or a dripping water tap respectively in the upper left corner). We implemented the app using the Unity 3D engine, deployed it on a tablet device and laser-cut a stand holding the tablet for the user studies (see Figure 3.6).

3.3.2 Evaluation

We investigate the following hypotheses:

H1: Hand washing is more enjoyable with Germ Destroyer

H2: Hand washing for 20 seconds *seems* shorter to participants when using Germ Destroyer

H3: Germ Destroyer increases the hand washing duration

H4: Germ Destroyer increases the number of hand washing sessions meeting the recommended duration

H5: The amount of bacteria or fungal cells, estimated by the number of colony-forming units ("CFU") on the door handle of the bathroom is lower when using Germ Destroyer

H1 is a prerequisite for the system's success in changing the behavior of users. We expect that the use of game elements such as points, normative feedback, progression and the presence of virtual characters makes hand washing more enjoyable, since previous gamified systems have been shown to be successful in this regard [315]. To investigate **H1**, we performed a user study assessing the enjoyment of the system using validated questionnaires. **H2** builds on **H1** as we expect that the increased enjoyment makes hand washing less boring and thus decreases the perceived hand washing time. **H3** is based on related work showing that gamified systems have been successful in changing people's behavior positively [145]. Thus, we expect that the use of our gamified application leads to an increase in the hand washing duration of users. **H4** targets the number of hand washing sessions meeting the recommended duration of 20 seconds. As our gamified app takes 20 seconds of hand washing time to be completed, we expect that the number of hand washing sessions meeting this duration should be higher when using the system. **H5** builds on **H3** as related work has demonstrated that the duration of hand washing is one of the most important factors to remove bacteria and other microorganisms [49, 322]. While **H1** and **H2** were studied in the lab, **H3, H4** and **H5** were investigated as part of an in-the-wild study, in which we installed our system in a shared bathroom for ten days and analyzed the hand washing duration. Additionally, we monitored microbiological hygiene of the bathroom's door handle by using commercially available test slides. In the next section, we present the method, procedure and results of the two studies we have conducted.

Laboratory Study

To investigate whether our system provides an enjoyable experience (**H1, H2**), we performed a lab study, in which participants were instructed to wash their hands with and without our system using a within-subjects design.

Method After obtaining informed consent and answering demographic questions as well as questions concerning game experience (on 5-point Likert scales), participants were instructed to wash their hands twice – once without Germ Destroyer and once with Germ Destroyer. The order of the two conditions was counterbalanced using a Latin Square design. In the baseline condition, participants were instructed to wash their hands until they were told to stop by a researcher. In the test condition, participants were asked to wash their hands until all germs within the gamified app were destroyed. In both conditions, one researcher was present and the hand washing duration was the recommended 20 seconds. Participants had to fill out questionnaires after each hand washing session. More specifically, participants were asked to answer the validated short German version of the Intrinsic Motivation Inventory [361] ("IMI"), consisting of four sub-scales: Enjoyment, Competence, Choice and Pressure. Furthermore, participants were asked to fill out the validated German version of the Positive and Negative Affect Schedule [59] ("PANAS") in order to measure whether Germ Destroyer had any effects on positive or negative affect. Afterwards, participants had to estimate how long they had been washing their hands. This was done to investigate whether the presence of the gamified app had any influence on the perceived duration. We expected that the gamified application would entertain and engage users, which would make the long hand washing duration of 20 seconds appear shorter. Results were analyzed using paired t-tests.

Results We recruited 14 participants (6 male, 8 female; 50.0% were aged 18-24, 28.6% 25-31, 7.1% 32-38, 7.1% 46-52 and 7.1% 53-59 years). Participants

	Scale Range	Baseline	Intervention
IMI Competence	*1-15*	M=10.21 SD=2.01	M=11.43 SD=1.91
IMI Choice	*1-15*	M=12.86 SD=2.57	M=12.71 SD=1.98
IMI Pressure	*1-15*	M=5.57 SD=1.70	M=5.71 SD=3.00
IMI Enjoyment*	*1-15*	M=5.93 SD=2.20	M=11.64 SD=2.06
PANAS Pos. Affect*	*1-5*	M=2.19 SD=0.81	M=3.55 SD=0.51
PANAS Neg. Affect	*1-5*	M=1.13 SD=0.22	M=1.24 SD=0.36
Perceived Duration*	*[in seconds]*	M=27.64 SD=13.04	M=22.14 SD=7.41

Table 3.2: Mean ("M") and standard deviation ("SD") for each dependent variable. Significant differences (p<.05) between conditions are represented by an asterisk.

considered themselves gaming-affine (M = 3.36, SD = 1.15), claimed to frequently play video games (M = 3.36, SD = 1.45) and to have a passion for them (M = 3.07, SD = 1.49). Table 3.2 summarizes all results of this study. Results from the IMI show no significant differences concerning the sub-scales competence, choice and pressure. However, as expected, a strong significant effect was found for the enjoyment sub-scale (t(13)=-7.26, p<0.001, d=2.68). Here, the mean score roughly doubled from 5.93 without Germ Destroyer to 11.64 when using the system. Regarding the negative affect as measured by the PANAS, we could not find significant differences. However, complementing the findings for the IMI, we found a strong significant increase in positive affect, rising from a mean score of 2.19 in the baseline to 3.55 when using Germ Destroyer (t(13)=-6.42, p<0.001, d=2.01). Considering both the significant increases in the IMI enjoyment score and in the PANAS positive affect score, we conclude that Germ Destroyer positively influenced the user experience during hand washing, supporting **H1**: *Hand washing is more enjoyable with Germ Destroyer*. Since washing one's hands for the recommended 20 seconds takes much longer than people usually wash their hands [49], we aimed at decreasing the *perceived* hand washing duration (**H2**). Indeed, we found that participants estimated to have washed their hands for significantly longer in the baseline condition, i.e. without using Germ Destroyer (t(13)=2.44, p<0.05, d=0.52), even though they washed their hands for exactly the same amount of time in both conditions. This supports **H2**: *Hand washing for 20 seconds seems shorter to participants when using Germ Destroyer*.

Field Study

To investigate **H3, H4** and **H5**, we installed our system for ten days in a shared bathroom of a company. The bathroom was located on the first floor with approximately 30 employees having their offices nearby.

Method The first five days were used to establish a baseline using the measuring device only, while the gamified mobile app was installed additionally for the last five days. In both conditions, we stored information about whether water was running or not, whether hands were being washed and the microbial concentration of the restroom's door handle. The system was in place between 8am and 3pm, i.e. for 7 hours per day. Each morning, the door handle was disinfected to ensure comparability. The microbial concentration on the restroom's door handle was assessed at 3pm each day using mikrocount TPC slides[13]. The slides were incubated at 37° Celsius for 24 hours. Afterwards, the test slides were photographed and the number of CFUs was counted. The study was approved by our ethical review board[14] and thoroughly discussed with the company's data protection officer and its employee representatives. Since the data protection officer raised concerns about installing the system in the women's restroom (the low number of women having their offices nearby would potentially allow one to infer who was using the bathroom), we decided to test the system in the men's bathroom only. Given that literature has shown that men neglect hand washing much more than women [169], we see this as acceptable. Due to the anonymous data collection, all three parties involved approved the execution of the study.

Results Overall, 363 hand washing sessions were recorded throughout the study (36.30 per day on average, SD=7.17) with a mean duration of 7.64 seconds (SD=7.11 seconds). In the baseline phase (days 1–5), 161 hand washes were recorded (32.20 per day on average, SD=6.26) having a mean hand washing duration of 5.56 seconds (SD=4.99 seconds). In the intervention phase, i.e. after installing Germ Destroyer (days 6–10), the number of hand washing sessions increased to 202 (40.4 per day on average, SD=5.86). Also, the mean hand washing duration strongly increased to 9.30 seconds (SD=8.07 seconds). As revealed by a Welch's t-test (the assumption of homogeneity of variance was violated), this increase is significant ($t(341.24)=-5.43$, $p<0.001$, $d=0.54$). This provides strong evidence for **H3**: *Germ Destroyer increases the hand washing duration*. Figure 3.10 visualizes the average hand washing duration and the standard deviation for all days of the study separately.

In the baseline phase, only 1.86% (SD=13.57%) of hand washing sessions were at least 20 seconds long. When relaxing this to a duration of 15 seconds (which is

[13] Schülke mikrocount TPC: *For the determination of the total plate count*, `https://bit.ly/3z40hrn` (last accessed: 2021-12-01)

[14] Saarland University: *Ethical Review Board of the Faculty of Mathematics and Computer Science*, `https://bit.ly/2TK2Qii` (last accessed: 2021-12-01)

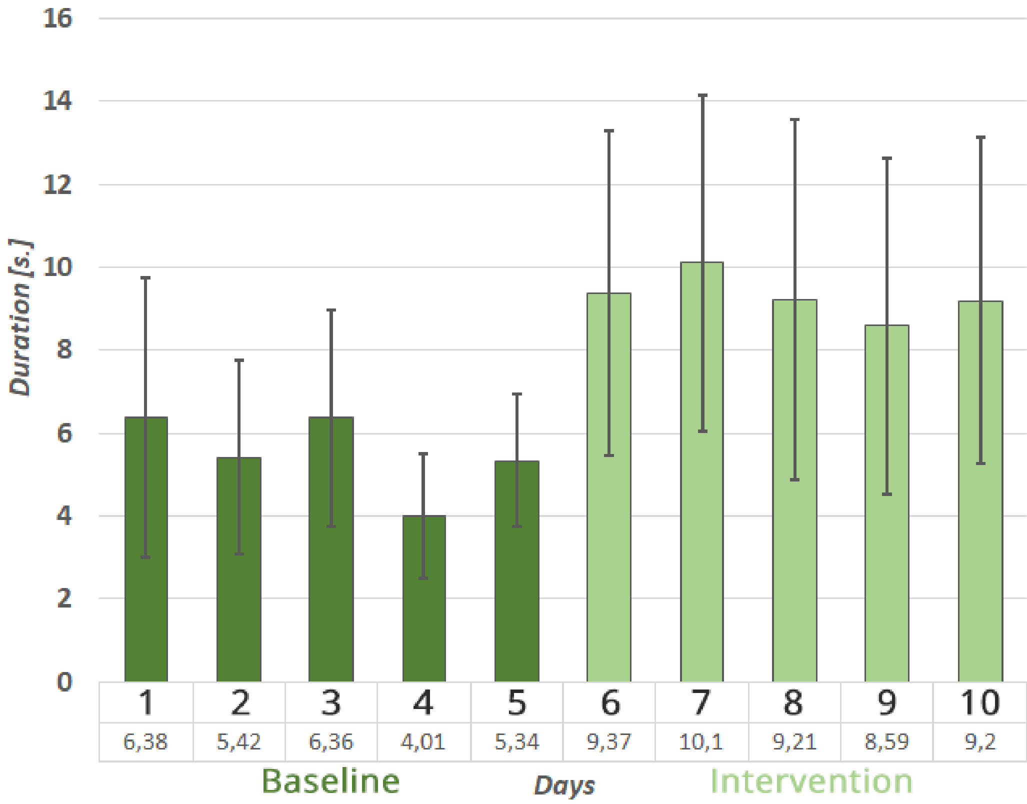


Figure 3.10: Average hand washing duration in seconds for each day of the study. The baseline phase was from day 1–5 (dark green), the intervention phase from day 6–10 (light green).

the lower bound of the recommended hand washing time [166]), the number of hand washing sessions meeting this criterion climbs up to 3.73% (SD=19.00%). In the intervention phase these results change substantially. The number of hand washing sessions lasting at least 20 seconds significantly increased to 17.82% (SD=38.37%; t(260.94)=-5.50, p<0.001, d=0.53). When relaxing this again to 15 seconds, the number of hand washing events having at least this length increases significantly to 26.73% (SD=44.37%; t(285.18)=-6.65, p<0.001, d=0.65). Therefore, **H4**: *Germ Destroyer increases the number of hand washing sessions meeting the recommended duration* is supported.

Since the door handle of the bathroom is most likely touched by all people using it and thus provides an increased risk of infection, we analyzed the bacteria count on it. The results were discussed with a microbiologist and a pharmacist. Since we only have five measurements (one measurement per day) per study phase and because the data was not normally distributed, we used the non-parametric Mann-Whitney-U test to compare the number of CFUs. In the baseline phase, we counted 90.40 CFUs per test slide on average (SD=67.67, Median=79.00, Min=27, Max=204). In the intervention phase, the number of CFUs declined significantly

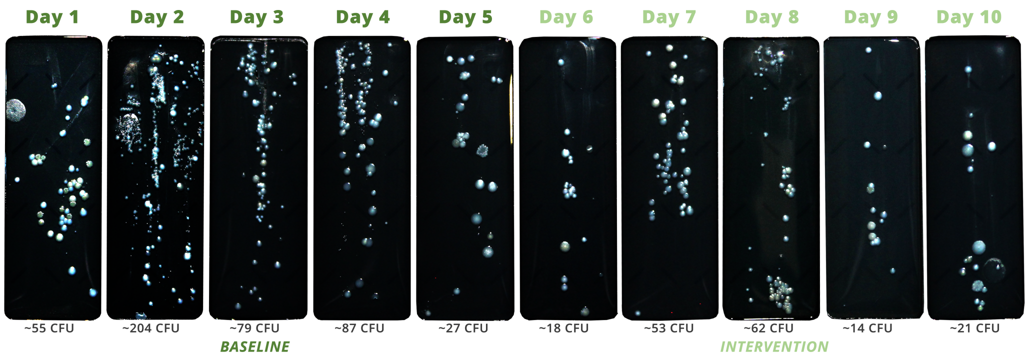


Figure 3.11: Pictures of the mikrocount TPC test slides after 24 hours incubation at 37° Celsius and the number of colony-forming units for each day. Days 1–5 belong to the baseline phase while Germ Destroyer was in place during days 6–10.

(U=3.00, Z=-1.98, $p<0.05$, r=0.63) to 33.60 on average (SD=22.19, Median=21.00, Min=14, Max=62). These results support evidence for **H5**: *The amount of colony-forming units ("CFU") on the door handle of the bathroom is lower when using Germ Destroyer*. Figure 3.11 shows the CFU counts and pictures of the test slides after incubation for each day.

3.3.3 Discussion and Limitations

In the course of a lab-based and an in-the-wild study we investigated the user experience and the effectiveness of our system. We found that Germ Destroyer makes hand washing more enjoyable and that people experience a more positive affect when using the system (**H1**). We assume that these positive effects are explainable by the gameful feedback provided by the system. More specifically, we suppose that gamification elements such as progression, points, virtual characters and praise lead to the increase in the IMI enjoyment as well as in the PANAS positive affect sub-scale since similar results have been reported in literature in different health-related contexts [146, 315]. The fact that participants perceived the hand washing time as shorter when using Germ Destroyer (**H2**) is most likely a direct consequence of the increased enjoyment [333] and thus supports **H1** further. Additionally, this finding shows that people tend to overestimate their time spent washing hands, which might explain the short hand washing duration found in the baseline phase of our in-the-wild study and in literature. The aforementioned evidence we found supporting both **H1** and **H2** forms the basis to find positive effects on hand washing behavior in the in-the-wild study.

Here, we found that the hand washing time in the baseline phase is far below the recommended duration of 20 seconds, with participants washing their hands for 5.56 seconds on average. This duration is in line with observational studies reporting that most people wash their hands for about 4-7 seconds [49, 322].

Additionally, the number of people washing hands for at least 15 seconds (3.73%) or 20 seconds (1.86%) is in line with previous research reporting that roughly 5% washed their hands longer than 15 seconds [49]. In view of these results, our measurement approach and our sample population seem appropriate. In the intervention phase (when Germ Destroyer was installed in the bathroom) both measures significantly increased (**H3, H4**). While the mean washing duration almost doubled, the number of people washing their hands for more than 20 seconds even approached a tenfold increase. These results show clearly that the presence of our system had a strong positive effect on the hand washing time. Potential reasons for these effects include an increased awareness caused by the application as was reported in [27], the higher enjoyment of hand washing (**H1, H2**), receiving gameful feedback and praise [126] or simulating the decrease of contamination using germs as virtual characters [97, 126]. The analysis of CFUs revealed that there were fewer viable bacteria or fungal cells when using Germ Destroyer (**H5**). One reason for this decrease could be the longer hand washing duration. Considering that the number of hand washing events was higher in the intervention phase, another reason could be that that the number of people who did not wash their hands at all was lower in the intervention phase. This would be in line with research showing that 33% usually do not wash their hands in shared bathrooms [363].

However, the studies presented in this section also have limitations. First, the in-the-wild study was conducted in the men's bathroom only. This was due to ethical concerns of the data protection officer, since the low number of women having their offices nearby would allow one to infer who was using the bathroom and when. Therefore, even though both men and women appreciated the system in the user experience lab study and no gender effects were found, it is not clear whether installing the system in the women's bathroom would lead to similar results. Given previous literature consistently reporting that men wash their hands less than women [49], we would expect that the effect might be smaller in the women's bathroom. It should also be noted that the system has been evaluated in a company – testing the system in different environments could lead to different effects. Although the results obtained as part of the in-the-wild study align well with previous literature, we would like to acknowledge that we cannot rule out measuring errors of our device. Due to the anonymous data collection, which was necessary so as to not violate the privacy of participants, we cannot give concrete information about the absolute number of discrete participants during the in-the-wild test. Considering that men use the bathroom 4.8 times during an 18-hour day on average [63] and thus assuming that people go to the bathroom one to two times between 8am and 3pm, we expect to have had 18–24 distinct users per day. This is supported by the number of offices nearby, as stated in the method section. It should also be noted that the duration of the in-the-wild study was not sufficient to make a statement about the long-term success of the system. Also, the potentially increased water consumption should be considered, especially when using our system in regions with water scarcity. Lastly, the measurement of the number of CFUs per day on the door handle has limitations. People possibly touching other things (like their face) between washing their hands and leaving the bathroom might confound a direct effect between the system and the bacterial counts. Also, the type of bacteria is unknown, i.e. it is unclear whether these bacteria are related to the use of the bathroom or are typical for human hands. Therefore, the results related to the bacterial counts on the door handle should not be overstated and need further validation.

3.3.4 Contribution to Research Questions

The results of the two studies contribute insights relevant to **RQ1**. We have seen that gamification affects psychological measures of motivation and (positive) affect. More specifically, we found that Germ Destroyer led to a significant increase on the interest/enjoyment subscale of the IMI. This subscale is considered the self-report measure of intrinsic motivation[15]. Consequently, we provide evidence that Germ Destroyer is capable of turning a rather boring activity into an enjoyable one and thus increases the intrinsic motivation of this activity. This is supported by an increased positive affect, since positive affective experiences were shown to foster intrinsic motivation [162]. Based on these results, it seems that the gamification elements used in Germ Destroyer elicited positive emotions, which in turn fostered intrinsically motivated behavior. In addition, besides positive affect, Germ Destroyer seems to have supported the satisfaction of participants' needs, as indicated by the increased intrinsic motivation. Although we did not find significant differences on the IMI competence nor choice subscales, the competence subscale descriptively increased when using Germ Destroyer. This may indicate that the context-inherent need for competence, which might result from people not knowing how long to wash their hands, could have been supported through Germ Destroyer, ultimately leading to higher levels of enjoyment and motivation.

That the positive effects on intrinsic motivation and positive affect actually translate into a behavioral change was shown as part of the in-the-wild study with the system. Here, not only did the duration and number of hand washing sessions increase, but the level of microbe contamination on the bathroom door handle also decreased. These results provide insights on how gamification works: Based on the results of the two studies, it seems that the gamification elements both elicited positive emotions and might have satisfied basic needs, translating into an increased intrinsic motivation and ultimately affecting participants' behavior. Besides contributing such insights, our findings also show the great potential gamification has for improving public health. Considering the ongoing Covid-19 pandemic, a system like Germ Destroyer could be helpful to increase hand washing adherence and thus decrease the risk of infection, especially in public spaces such as airports or train stations.

[15] Self-Determination Theory: *Intrinsic Motivation Inventory*, `https://bit.ly/3ko0RcJ` (last accessed: 2021-12-01)

Figure 3.12: Deactivate ads by clicking a "close" button or by playing a Paintball (b), Tetris (c) or Monster (d) game.

Now that we have considered a context in which social relatedness in Section 3.2 was particularly relevant and a context in which competence-supportive gamification elements such as points, goals or progression [370] were used to assist users in washing their hands in Section 3.3, we consider online advertising, a context in which autonomy needs are inherently thwarted.

3.4 Gamifying Online Advertisments

In this last section of this chapter, we shift the behavior change context from health to marketing and advertising. This context is also among those investigated regarding persuasive technology and behavior change [145]. Nowadays, most popular websites such as search engines, streaming services or news platforms are primarily monetized through online advertisements [231, 277]. This allows them to offer many of these services free of charge. However, users perceive ads increasingly as annoying (e.g. because of ad clutter) [83], intrusive (e.g. caused by ads that require explicit interaction) [277] and disruptive (e.g. because of animated content or autoplaying videos that cannot be deactivated) [23], leading to a negative user experience [53, 292]. As a result, users react with site abandonment [92] or make use of so called "ad blockers", preventing ads from loading on websites [23]. These both present a serious threat to the business model of commercial web services [277]. Therefore, we investigate to what extent gamification could be used to change the user's attitude towards and experience of online advertisements, to bridge the gap between the enjoyment and effectiveness of online ads.

Having previously focused on contexts in which the basic psychological needs for relatedness (see Section 3.2) and competence (see Section 3.3) played an important role in whether and how gamification works, to conclude this chapter, we now focus on a context in which the need for *autonomy* is inherently thwarted. Users usually have no choice about whether they will see or interact with online ads – they *have* to consume ads to use the service on which they appear. Even if they have installed ad blockers, many websites and services have started to detect them and exclude users who have them installed. Thus, we can see that the context of online ads is characterized by low autonomy, which may explain why we perceive ads rather negatively.

Based on two user studies – an online study and a laboratory study – we investigate whether playfully deactivating ads could help users regain autonomy and control over when they would like to consume ads, increase their intrinsic motivation and improve their overall user experience with ads. We also study whether gamification can at the same time increase the effectiveness of online advertisements.

3.4.1 Concepts for Gamified Advertisements

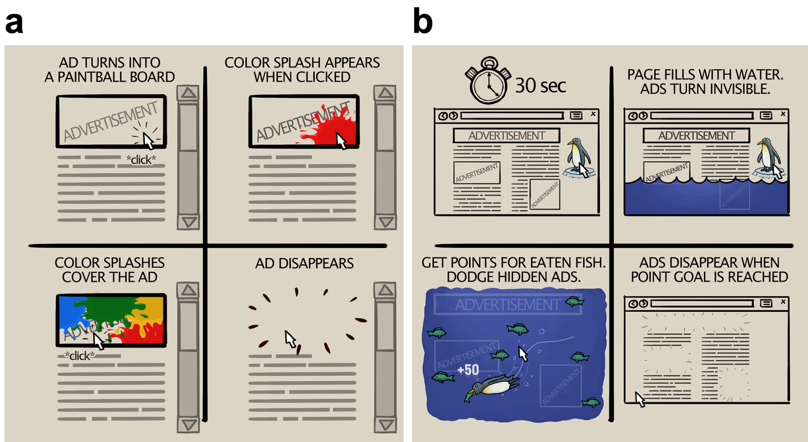


Figure 3.13: Storyboards for the Paintball (a) and the Underwater (b) concepts

Since we provide the first investigation of playfully deactivating ads (as far as we know), we created eight different gamification concepts based on simple and well-known games [51, 129]; these will be described in the following sections. For each gamification concept a storyboard was created, based on the guidelines by Truong et al. [345]. The storyboards were used to evaluate the perception of each gamification concept in order to elicit well perceived concepts for later implementation. All figures show the final versions of the storyboards, i.e. all changes that were derived through the validation study described later in this section are already integrated.

Paintball

In this gamification concept (see Figure 3.13a), ads can be shot at with a virtual paintball gun, covering parts of them with color splashes. Once the advertisement is completely covered, it disappears and will not be displayed again.

Underwater Website

On the sidebar of the web page, a virtual penguin is displayed. When clicked, the user has 30 seconds to memorize the layout of the web page. After that time, the web page begins to fill with water and all ads turn invisible. Once the web page is fully filled with water, the user can control the penguin through the water by dragging the mouse and catch fish. Hereby, the penguin has to evade the invisible advertisements. Catching fish gives points. Once a point goal is reached, all ads disappear (see Figure 3.13b).

Rival Teams Competing Over Website Control

Two rival teams compete over control of ads on web pages. Users can capture ads, which causes them to disappear and not be displayed again. The user's team gets rewarded points for the claimed ad. The number of claims is visualized and shows which team is winning (see Figure 3.14a).

Pacman

In this concept, ads can be turned into a Pacman game (see Figure 3.14b). The Pacman character follows the mouse and needs to evade the chasing ghosts. By moving, the Pacman character bites off parts of the ad. Once the ad is small enough, it disappears and will not be displayed again.

Tetris

The ad can be turned into a Tetris game (see Figure 3.15a). Falling blocks can be moved with the arrow keys on the keyboard. The ad shrinks every time a row is

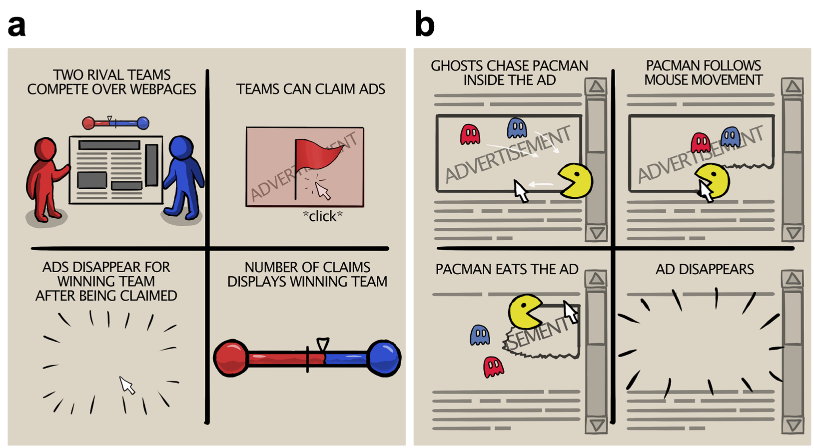


Figure 3.14: Storyboards for the Rival Teams (a) and Pacman (b) concepts

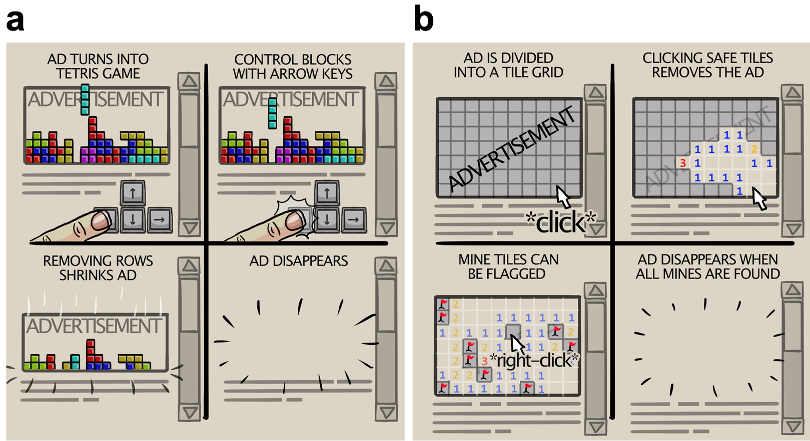


Figure 3.15: Storyboards for the Tetris (a) and Minesweeper (b) concepts

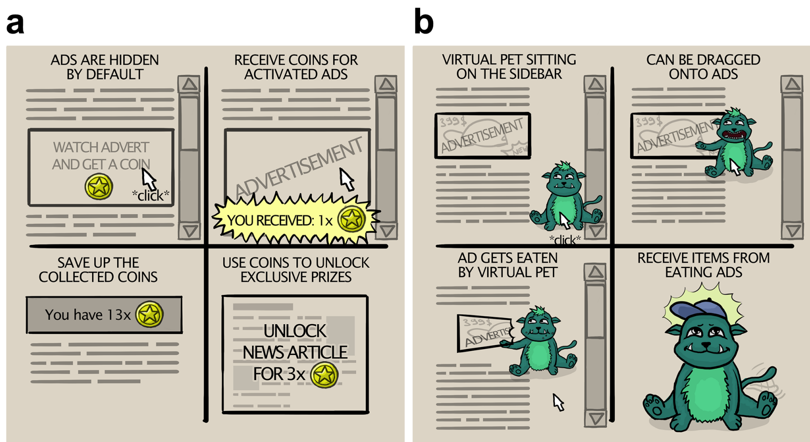


Figure 3.16: Storyboards for the Coins (a) and Monster (b) concepts

removed. Once the advertisement is small enough, it disappears and will not be displayed again.

Minesweeper

The ad can be turned into a Minesweeper game (see Figure 3.15b). When tiles are clicked, they give hint on the number of mines in their proximity. The mine tiles can be flagged by right clicking. Once every mine is found, the ad disappears and will not be displayed again.

Collecting Coins

In this concept, ads are hidden by default on the website. They can be manually clicked to display the hidden ad, granting virtual coins to the user. These coins can be used to unlock exclusive content or other prizes on the web page (see Figure 3.16a).

Ad-Eating Monster

On the sidebar of the web page a virtual character is displayed. It can be clicked and dragged onto ads. The virtual character then eats the ad and receives a virtual item that can be worn or equipped by the virtual character. The ad disappears and will not be displayed again (see Figure 3.16b).

Storyboard Validation

Following the guidelines by Truong et al. [345], we performed a study to ensure that the storyboards are comprehensible. We set up an online questionnaire, in which the storyboards were shown to participants (all storyboards were presented in random order). Participants were given the following task: *"In your own words, please describe the game idea that is illustrated in this storyboard"*. Answers could be provided in a free-text field. Afterwards, two independent raters read through all answers and rated whether participants understood the underlying game idea, using a scale of 1=*"the participant did not understand the concept at all"*, 2=*"there were minor comprehension issues"* and 3=*"the participant fully understood the concept"*. Moreover, each rater was asked to note aspects that were misunderstood.

We recruited 20 American participants (15 male) from Amazon Mechanical Turk ("AMT"), who were paid $1.50 each (the study took roughly 10-15 minutes to complete). To ensure that the ratings could be interpreted objectively, we calculated the inter-rater agreement and found it to be Cohen's Kappa κ=0.94, which is considered almost perfect [221]. The average rating across all storyboards was 2.78 (SD = 0.16, Mdn = 3), showing that there were no major comprehension issues and that the storyboards indeed explain the intended gamification concepts. This is supported by the fact that all storyboards were given a median rating of 3 by the raters and that their mean rating was higher than 2.50. Therefore, based on the transcribed comprehension issues, only minor adaptations had to be made: Before the validation study, we used the term "token", on the "Collecting Coins" storyboard, which led to false interpretations. Therefore, it was renamed to "coins" instead. Also, some participants did not realize that the ad was embedded in a web page, which is why scroll bars on the right side of each screen were added, indicating the use of a web browser.

3.4.2 Online-Study: Concept Evaluation

To inform which gamification concepts should be implemented and to get further insights about important requirements for later realization, we performed a study assessing the perception of each gamification concept. For this study, the final versions of the storyboards, i.e. those after the validation study, were used.

Method

After asking about demographic data, and whether participants have an ad blocker installed, participants were asked to develop a game idea, which has the goal to improve their perception of ads on a website. This was done to elicit requirements that should be considered when implementing well-perceived gamification concepts. To get unbiased concepts, storyboards were presented in random order afterwards and the following perception statements had to be rated (on 5-point Likert scales) for each concept:

Enh.Perc.: *"This game idea would enhance my perception of ads on a website"*

LikePlay: *"I would like to play this game to deactivate ads"*

FunPlay: *"This game would be fun to play"*

LikeIdea: *"I like this game idea"*

Furthermore, a comprehension question was asked for each storyboard, to be answered on a 5-point Likert scale (*"I think this game idea is easy to understand"*). American participants were recruited from AMT and paid $2.50 for participating (the study took roughly 15-20 minutes to complete).

Results

50 participants took part (31 male), of which 66% had an ad blocker installed. The written answers were analyzed by conducting an inductive content analysis [156] with two coders. Results were discussed and deviations solved to establish a final set of themes. Based on this, we derived the following requirements for the realization of the gamification concepts:

R1: Casual games: Most gamification concepts (35) were using simple rules without requiring substantial skill improvements.

R2: Entertainment over effectiveness: 20 participants emphasized that the main purpose should be entertainment and not promotion of a product or brand.

R3: Unobtrusiveness: 16 participants required that no ad-specific events (e.g. redirecting to another website) should be triggered during playing the game; the ad and the game should be separated regarding user interaction (e.g. by using a start button).

R4: Short play time: The games should not take long to complete (mentioned by 13 participants). Five participants even reported concrete time spans (15 to 60 seconds).

	Enhance Perception	Like Play	Fun Play	Like Idea
Paintball	M = 3.32 SD = 1.42 Mdn = 4.00	M = 3.24 SD = 1.45 Mdn = 4.00	M = 3.36 SD = 1.43 Mdn = 4.00	M = 3.56 SD = 1.37 Mdn = 4.00
Underwater	M = 3.24 SD = 1.36 Mdn = 3.50	M = 3.04 SD = 1.43 Mdn = 3.00	M = 2.94 SD = 1.42 Mdn = 3.00	M = 3.12 SD = 1.61 Mdn = 3.00
Rival Teams	M = 2.64 SD = 1.24 Mdn = 3.00	M = 2.38 SD = 1.14 Mdn = 2.00	M = 2.54 SD = 1.39 Mdn = 2.00	M = 2.62 SD = 1.23 Mdn = 3.00
Pacman	M = 3.04 SD = 1.43 Mdn = 4.00	M = 3.06 SD = 1.45 Mdn = 3.00	M = 3.22 SD = 1.43 Mdn = 4.00	M = 3.28 SD = 1.37 Mdn = 4.00
Tetris	M = 3.72 SD = 1.26 Mdn = 4.00	M = 3.76 SD = 1.38 Mdn = 4.00	M = 3.96 SD = 1.21 Mdn = 4.00	M = 3.86 SD = 1.18 Mdn = 4.00
Minesweeper	M = 3.74 SD = 1.21 Mdn = 4.00	M = 3.66 SD = 1.33 Mdn = 4.00	M = 3.82 SD = 1.34 Mdn = 4.00	M = 3.88 SD = 1.27 Mdn = 4.00
Collecting Coins	M = 2.92 SD = 1.31 Mdn = 3.00	M = 2.96 SD = 1.38 Mdn = 3.00	M = 3.28 SD = 1.42 Mdn = 4.00	M = 3.14 SD = 1.40 Mdn = 4.00
Monster	M = 3.24 SD = 1.38 Mdn = 4.00	M = 3.28 SD = 1.43 Mdn = 3.50	M = 3.36 SD = 1.34 Mdn = 4.00	M = 3.52 SD = 1.31 Mdn = 4.00

Table 3.3: Mean ("M"), standard deviation ("SD") and median ("Mdn") for each concept. Significant differences ($p<.05$) from the neutral choice are color-coded (green for positive deviations, red for negative ones).

The gamification concepts were easy to understand (M = 4.21, SD = 0.35, Mdn = 4.5), backing up the findings from the validation study. We performed one-sample t-tests against the value 3 ("*neither agree or disagree*") for all perception statements, to see which gamification concepts were perceived significantly better than the neutral choice. Based on the results (see Table 3.3), four gamification concepts showed positive effects: "Paintball", "Tetris", "Minesweeper" and "Monster". Since it is hard to reduce the playing time of the "Minesweeper" concept substantially (**R4**) because of the inherent strategic nature of the game, the Paintball, Tetris and Monster concepts were implemented. These three concepts differ in their level of interactivity (while "Tetris" requires the most interaction to deactivate an ad, "Paintball" requires less and the "Monster" the least interaction).

3.4.3 Implementation of the Gamification Concepts

The selected concepts were implemented within a fictitious news website. In line with previous advertising research [61, 92, 99, 368], we decided to use a news website, since they are relying on financial revenue from ads [23, 368], provide the opportunity for goal-oriented tasks [99], and because reading news articles is considered a typical activity users perform on the web [23,53,71], thus embodying a realistic setting.

Since there are only a few simple rules to be considered to play each of the three gamification concepts, we see **R1** as fulfilled. Also, since we aimed for gamification concepts that are independent of the actual ad, there was no connection between ad content and the game, fulfilling **R2**. Regarding **R3**, we implemented two buttons placed on the top left corner of an ad, to get information about how to play the game and how to start the game. Once the button to start the game is pressed, the interactivity of the actual ad (redirecting the user to the website of the brand or product being advertised) is deactivated. Considering **R4**, we adapted the game goals such that winning is possible within a short duration (all games can be completed in less than 30 seconds). In the following, each implementation is described in more detail.

Paintball

Once users start the game by clicking on the respective button, they can shoot color splashes on the ad by clicking on it (see Figure 3.12b). In the game, the ammunition is limited and the capacity of the magazine holds five shots, i.e. the gun needs to be reloaded by pressing the space key. Once enough of the ad space is covered with splashes, the game is won, a congratulatory message appears and the ad finally fades out progressively. To account for **R4**, color splashes adapt to the ad size. If the game is lost, i.e. there is no ammunition left, the game stops and the user is shown the ad without the game tools. However, the game can be started again at any time by pressing the “Start” button.

Tetris

After activating the game, the ad is overlaid with a semi-transparent playing field (see Figure 3.12c). A random sequence of geometric pieces falls down. With the “up” key, pieces can be rotated and with the “left” and “right” keys they can be moved. By pressing the down key, pieces fall down faster. The goal of the game is to place the pieces such that they create a horizontal line. When such a line is created, the game is won, a congratulatory message appears and the ad steadily shrinks until it is gone. To account for **R4**, we required users to only complete one row. If blocks touch the upper edge of the playing field, the game stops. It can be started again by pressing the “Start” button.

Monster

In contrast to before, there is no dedicated "Start" button, since the monster is always visible on the left side of the screen. However, to account for **R3**, the monster needs to be dragged onto an ad. Similar to the storyboard, we created a green, supposedly cute monster that likes eating ads. We implemented an idle animation, in which the monster waits at the left side of the screen, slowly moving its head and eyes. When the user starts dragging the monster, the monster starts to smile. Once the monster is dragged over an ad, it sticks out its tongue to indicate its pleasant anticipation (see Figure 3.12d). When released, an eating animation starts, showing the monster chewing while the ad progressively disappears. Afterwards, the ad turns into a virtual item (e.g. a cap), that can be placed on the monster.

3.4.4 Laboratory Study

We conducted a lab experiment, which was approved by our ethical review board[16], to investigate the effects of gamified ads on ad effectiveness and user enjoyment. We compared every gamified condition against the baseline (deactivating ads without gamification) and expected the following effects:

H1 Playfully deactivating ads is more enjoyable

H2 A lower number of participants prefers using an ad blocker for reading a news article in the gamified conditions

H3 Brand recall is higher in the gamified conditions

H4 Product recall is higher in the gamified conditions

H5 Brand recognition is higher in the gamified conditions

H6 Product recognition is higher in the gamified conditions

H7 Ad recognition is higher in the gamified conditions

H8 Implicit brand memory is higher in gamified conditions

H9 Gamification enhances the perception of the website

H10 Gamification decreases the perception of news articles

H11 Gamification enhances the perception of ads

[16] Saarland University: *Ethical Review Board of the Faculty of Mathematics and Computer Science*, `https://bit.ly/2TK2Qii` (last accessed: 2021-12-01)

We assume that the positive aspects of playful deactivation predominate over the higher effort needed, based on the inherent motivational power of games [303], affecting user motivation and thus enjoyment positively (**H1, H2**) [300]. **H3 - H7** are motivated by previous research showing that interaction with ads leads to positive effects on explicit memory [205, 288]. As all of our implemented gamification concepts require directly interacting with the ad, we expect to find similar results. Since embedding ads in digital games was found to be beneficial for implicit (brand) memory [319, 369], we expect that the playful approach augmenting ads with game elements should also lead to positive effects, motivating **H8**. Because online advertising was found to influence the perception of both the website hosting the ad [71, 218, 248] as well as of news articles [368] previously, we were curious whether using gamified ads also leads to effects on these aspects (**H9**, **H10**). While we expect positive effects of the attitude towards the site, we expect that game elements should have a negative effect on the perception of news articles. Since interactive ads and slogans indicating playful features of an advertised product were found to affect the attitude towards the ad positively [45, 71, 329], we expect similar effects (**H11**).

Method

The main task of the experiment was to visit a news website, read three different articles, answer a comprehension question for each article and deactivate an ad that was placed in each one. Participants were recruited via flyers on campus, social media and mailing lists. The study took approximately 35 minutes to complete, was available in English and German, and was compensated with a 7 € Amazon voucher. Following a between-subjects design, participants were randomly assigned to one of four conditions (Baseline, Paintball, Tetris, Monster). In the Baseline condition, they could deactivate ads by clicking on a button (labeled with an "X" mark) on the upper right corner of the ad (similar to [39]; see Figure 3.12a). By requiring participants to close the ad in the baseline condition, we ensured that they interacted with the ad, which was necessary for the comparability of the conditions (i.e. without requiring participants to deactivate an ad, we could not ensure that the ad was recognized at all). Also, since interactive ads were shown to have positive effects on ad effectiveness [205, 288], using an interactive baseline is necessary to investigate the effects of gamified ads.

Advertisements We created six different ads, promoting fictitious brands and products, since advertising effects can be better assessed if participants have no preexisting attitudes towards the brands [39, 77]. Of these six, three were chosen to be shown in each of the three articles on the news website whereas the other three ads were added in the recognition tests as "false choices", such that each included ad had a similar product as a false counterpart. Based on related work, we included an ad for mineral water with a soft drink as a counterpart (based on [288, 329]), an ad for a car having a motorcycle as a counterpart (based

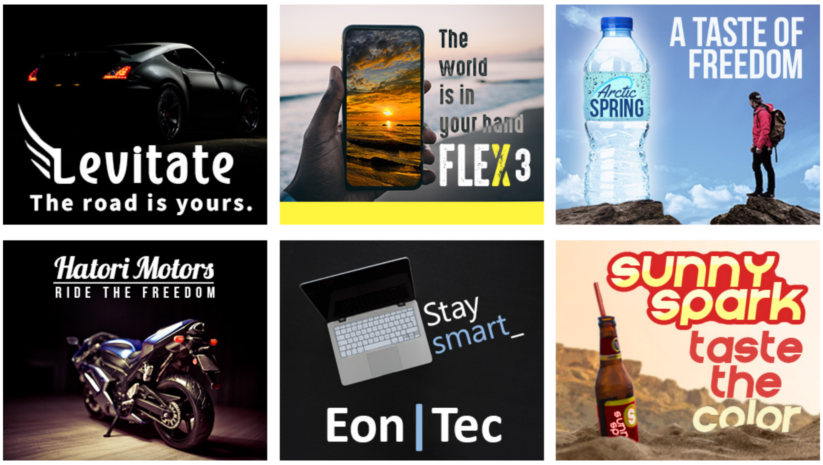


Figure 3.17: Banner ads of fictitious brands used in the study. All ads in the upper row were shown on the news web page; ads in the lower row were added in the questionnaire as false counterparts that were never shown on the news web page

on [77, 99, 211]) and an ad for a mobile phone and one promoting a notebook as its counterpart (based on [125, 227]). All ads (see Figure 3.17) were created as banner ads (since this type of ad is the most frequently used [176]), in Google Ad Words' 336x286 format[17] and placed approximately 2/3 of the way down the page, following recommendations from an ad placement study [38].

Measures and Procedure

The experiment started with the instructor explaining the task to the participant. Next, participants signed a consent form, took a seat in front of a desktop PC and started filling out a web questionnaire guiding participants through the whole study. It should be noted that participants were not informed that they would be asked to recall brands or products, nor were they told to pay special attention to the ads. The questionnaire started with questions about demographics, their game affinity and whether they have an ad blocker installed. Next, participants saw a description of their task, stating that they will have to read several articles (articles were taken from a website offering free news content[18]), answer a comprehension question for each article and deactivate an ad that is placed within each article. The description of how to deactivate an ad changed depending on the assigned condition (Baseline, Paintball, Tetris, Monster). These three tasks were then presented to the user (in randomized order). As an example, one

[17] Google: *About common sizes for responsive display ads*, `https://bit.ly/2NPN0OO` (last accessed: 2021-12-01)
[18] BrandpointContent: *Free Premium Content*, `https://bit.ly/3ps0HmU` (last accessed: 2021-12-01)

article was about group travel[19] and the task was "*Please state benefits of group travel*". Once participants found the article, deactivated the ad and answered the comprehension question, they were asked questions about the perception of the news article, and of the ad they had just deactivated.

For the news perception, we used Sundar's news evaluation items [328], consisting of three subscales (Value, Credibility, Entertainment). The perception of ads was measured using the scale by Yang et al. [368] consisting of two subscales: Professionalism and Appropriateness. Both news perception and ad perception items were measured using 10-point Likert scales; for all remaining questionnaires 7-point Likert scales were used. To investigate the enjoyment of deactivating ads, the short German version of the Intrinsic Motivation Inventory [361] ("IMI") followed after completing the three tasks, consisting of four subscales: Enjoyment, Competence, Choice and Pressure. The scale was translated to English by a professional, bilingual translator. Afterwards, the perception of the news site regarding the overall user experience and the attitude towards the website was measured using the short version of the User Experience Questionnaire (UEQ-S) [310] and the "Attitudes Towards Site" scale developed by Chen and Wells [78] (having the subscales "Entertainment", "Informativeness" and "Organization"). Next, participants were asked whether they would like to use an ad blocker or deactivate ads as they did before for an additional article, even though they were told afterwards that there would be no such additional article. This was done to get another indication of the perception of playfully deactivating ads. The aforementioned questionnaires also served as distraction tasks, clearing the short-term memory prior to completing the word stem completion and recall/recognition tests [369], which are explained next.

To measure implicit brand memory, word stem completion tests followed [369]. In these tests, participants had to fill in the missing letters of a word stem to make it into a meaningful word. We included six word stems, of which three were stems from brand names from the previously deactivated ads. To measure unaided product and brand recall, two free-text fields followed, asking the participants to name any product or brand they can remember. Textual recognition tests followed, in which participants had to choose any brand or product they recognize (again, six brands were shown in randomized order, of which three were actually advertised, and six products, of which again three were actually advertised). Lastly, visual ad recognition tests followed, in which participants were shown six ads (in randomized order), of which three were the same as on the news site, and were asked to choose any ad they recognized.

Results

Overall, 72 participants were recruited (42 male, 29 female, 1 not specified), 18 participants for each group. This number of participants was informed by

[19] BrandpointContent: *Consider a new way to see the world - the benefits of group travel*, `https://bit.ly/3gc1AvZ` (last accessed: 2021-12-01)

	Baseline	Tetris	Paintball	Monster
IMI Enjoyment Range: 3-21	9.00 / 4.38 / 8.50	14.28 / 3.75 / 14.50	13.89 / 4.34 / 12.50	14.50 / 3.49 / 15.00
IMI Competence Range: 3-21	14.94 / 4.48, / 15.00	15.44 / 4.34 / 15.50	14.28 / 4.11 / 14.50	18.06 / 4.05 / 19.00
IMI Choice Range: 3-21	15.06 / 4.11 / 15.00	15.00 / 3.58 / 15.50	13.72 / 4.23 / 14.50	13.56 / 3.75 / 14.00
IMI Pressure (neg.) Range: 3-21	17.56 / 4.19 / 19.00	16.17 / 4.54 / 16.50	15.67 / 5.21 / 16.00	19.17 / 3.90 / 21.00
UEQ-S Range: 7-49	36.33 / 4.19 / 35.50	36.50 / 7.00 / 36.50	38.56 / 7.33 / 39.50	34.67 / 7.43 / 34.50
Site Entertainment Range: 4-28	16.83 / 5.07 / 16.50	15.89 / 5.86 / 15.50	18.56 / 5.22 / 19.00	16.67 / 5.37 / 17.00
Site Informativeness Range: 4-28	20.17 / 4.00 / 20.50	18.17 / 5.10 / 17.50	21.00 / 3.90 / 20.50	19.06 / 4.09 / 19.00
Site Organization Range: 4-28	21.61 / 3.05 / 22.00	23.00 / 3.50 / 23.00	23.67 / 3.41 / 24.00	24.39 / 3.26 / 25.00
News Value Range: 1-10	7.25 / 1.44 / 7.38	6.46 / 1.35 / 6.17	7.27 / 1.38 / 7.17	5.96 / 1.59 / 6.21
News Credibility Range: 1-10	7.19 / 1.26 / 7.25	6.28 / 1.56 / 6.38	7.45 / 1.21 / 7.00	6.35 / 1.15 / 6.25
News Entertainment Range: 1-10	6.90 / 1.19 / 6.72	6.35 / 1.81 / 6.22	7.25 / 1.90 / 7.06	5.88 / 1.77 / 6.22
Ad Professionalism Range: 1-10	5.65 / 1.73 / 5.50	5.37 / 1.50 / 5.22	5.73 / 1.79 / 5.78	5.43 / 1.25 / 5.33
Ad Appropriateness Range: 1-10	6.41 / 2.59 / 6.50	6.49 / 1.70 / 6.67	5.88 / 2.36 / 6.33	6.69 / 2.27 / 7.39
Brand Recall Range: 0-3	0.00 / 0.00 / 0.00	0.28 / 0.46 / 0.00	0.06 / 0.24 / 0.00	0.11 / 0.32 / 0.00
Product Recall Range: 0-3	2.00 / 0.77 / 2.00	2.33 / 0.84 / 2.50	1.50 / 0.99 / 1.50	1.56 / 0.98 / 2.00
Brand Recognition Range: 0-3	1.28 / 0.83 /1.00	1.89 / 0.83 / 2.00	1.94 / 1.00 / 2.00	1.94 / 1.00 / 2.00
Product Recognition Range: 0-3	2.28 / 0.58 / 2.00	2.72 / 0.46 / 3.00	2.44 / 0.86 / 3.00	2.56 / 0.62 / 3.00
Ad Recognition Range: 0-3	2.56 / 0.62 / 3.00	2.83 / 0.38 / 3.00	2.72 / 0.75 / 3.00	2.78 / 0.43 / 3.00
Word Stems Range: 0-3	0.50 / 0.62 / 0.00	0.67 / 0.69 / 1.00	0.61 / 0.61 / 1.00	0.67 / 0.78 / 0.50
Prefer Ad Blocker Range: 0-100	83% / 38% / 100%	67% / 49% / 100%	89% / 32% / 100%	33% / 49% / 0%

Table 3.4: Possible range of values and mean / standard deviation / median of all dependent variables for each condition. Green cells indicate positive, red negative significant effects compared to the baseline.

an a-priori performed power analysis (effect size f=.41 [256] and a power of 80.75% [86]). Participants considered themselves gaming-affine (M = 3.61, SD = 1.10, Mdn = 4.00), claimed to frequently play video games (M = 3.21, SD = 1.32, Mdn = 4.00) and to have a passion for them (M = 3.31, SD = 1.37, Mdn = 4.00). 82% (SD = 38.7%, Mdn = 100%) have an ad blocker installed. 56.94%

were aged 25-31, 36.11% 18-24 and 6.95% were aged 32 and older. The following sections present results from independent t-tests, comparing each game concept against the baseline, to investigate our hypotheses. It should be noted that our goal was not investigating effects between the game concepts (see **H1-H11**). We argue that such comparisons would not lead to valid conclusions, as the game concepts are not comparable due to their different specifics. Therefore, no pairwise comparisons have been made. All reported p-values were corrected for multiple comparisons using Benjamini and Hochberg's False Discovery Rate, as described in [40]. Table 3.4 visualizes relevant descriptive data for each condition at a glance. Significant differences from the Baseline condition are colored.

Enjoyment and Acceptance All game conditions scored significantly higher than the baseline on the "enjoyment" subscale of the IMI. Deactivating ads using the "Tetris" concept was perceived as significantly more enjoyable than the baseline ($t(34) = -3.88$, $p = 0.000$); this was also true for the "Paintball" ($t(34) = -3.37$, $p = 0.002$) and the "Monster" ($t(34) = -4.17$, $p = 0.000$) concepts. This provides strong evidence for **H1**: *Playfully deactivating ads is more enjoyable*. Furthermore, for reading a news article, we found that participants prefer to use an ad blocker instead of the gamification concepts for the "Tetris" (67% prefer to use an ad blocker) and the "Paintball" (89% prefer to use an ad blocker) concepts. However, the "Monster" concept is an exception: Here, only 33% of the participants preferred to use an ad blocker; thus 67% prefer to use the "Monster" gamification concept. When comparing these values against the baseline, the "Monster" gamification concept is the only one showing a significant effect. Therefore, **H2**: *A lower number of participants prefers using an ad blocker for for reading a news article in the gamified conditions* is supported for the "Monster" concept, while no statement can be made for the other gamified concepts.

Product and Brand Recall Descriptively, the number of correctly recalled brands is higher in every gamified condition than in the baseline (see Table 3.4). Participants in the "Tetris" condition recalled a significantly higher number of brands correctly ($t(34) = -2.56$, $p = 0.045$). For both other conditions, no significant effect was found. These results support **H3**: *Brand recall is higher in the gamified conditions* for the "Tetris" concept, while no statement can be made for the "Paintball" and "Monster" concepts. Additionally, there were no significant differences between the baseline and any of the gamified conditions in the number of correctly recalled products; thus our data do not support **H4**: *Product recall is higher in the gamified conditions*.

Product, Brand and Ad Recognition We analyzed the number of correctly recognized textual representations of products and brands and of recognized ads. Our results show that participants recognized significantly more brands correctly in all gamified conditions than in the baseline (Tetris: $t(34) = -2.21$, p

= 0.036); Paintball: t(34) = -2.18, p = 0.036); Monster: t(34) = -2.18, p = 0.036). This provides strong evidence for **H5**: *Brand recognition is higher in the gamified conditions*. Regarding product recognition, we found a significant difference between the "Tetris" condition and the baseline (t(34) = -2.56, p = 0.045). However, no significant effects were found in the "Paintball" and "Monster" conditions. Thus, **H6**: *Product recognition is higher in the gamified conditions* is only supported for the "Tetris" concept. Lastly, regarding ad recognition, we did not find evidence supporting **H7**: *Ad recognition is higher in the gamified conditions*.

Word Stem Tests Although the number of correct brands is higher in all gamified conditions, no significant effects could be found supporting **H8**: *Implicit brand memory is higher in gamified conditions*.

Perception of the News Website, its Articles and the Ads Only the "Monster" condition significantly influences the attitude toward the site and the perception of news articles. We found a significant effect on the "organization" subscale of the Attitude Toward the Site scale (t(34) = -2.64, p = 0.036), showing that the "Monster" concept increased how organized participants considered the website to be. However, no effect was found for the UEQ. At the same time, participants perceived the news articles as significantly less valuable (t(34) = 2.56, p = 0.045) in the "Monster" condition. Overall these findings support evidence for **H9**: *Gamification enhances the perception of the website* for the "Monster" concept. Also, **H10**: *Gamification decreases the perception of news articles* is supported for the "Monster" concept. No evidence was found for the other gamified conditions. Also, no evidence for **H11**: *Gamification enhances the perception of ads* was found.

3.4.5 Discussion and Limitations

Our results demonstrate that deactivating ads playfully is enjoyable for users, as was shown by the increased IMI enjoyment score for all gamification conditions (**H1**). The fact that the "Monster" concept was even preferred over using an ad blocker for reading a news article further supports this finding (**H2**). The reason why only the "Monster" game concept was preferred over using an ad blocker might be because it is the only concept allowing to unlock virtual rewards, stimulating users' feeling of accomplishment [126]. Moreover, we found that each gamified condition led to at least one increased measure of explicit memory, thus indicating that deactivating ads playfully has the potential to increase the effectiveness of ads (**H3-H6**). Considering that the "Tetris" concept, requiring the most user interaction, showed the highest number of significant effects on explicit memory measures, an explanation could be that the level of interaction is a deciding cause. This would be in line with previous research [205, 288]. Also, the animations used in the "Monster" concept and the fact that ads were covered with color splashes in the "Paintball" concept could explain why these concepts showed less of an effect on explicit memory [329]. The reason for not finding any effects regarding visual ad recognition (**H7**) might be related to the picture superiority effect [265], stating that images are more likely to be remembered than words, which might lead to the high ad recognition in all groups. Regarding implicit memory, no effects were found (**H8**). This might be due to the comparatively conservative baseline condition, in which we required users to interact with the ad, which in itself affects implicit memory measures positively [319]. Given the descriptive data regarding the word stem tests, testing more participants might lead to finding an effect. We also found that the "Monster" condition led to a more positive perception of the website (**H9**). This might be explainable by the fact that the positive perception of playfully deactivating ads, was transferred to the website, similarly as was shown for certain types of ads [71, 248]. However, the value of news articles was seen as significantly lower in the "Monster" condition (**H10**). This hints that the playfulness of the "Monster" concept was perceived as inappropriate when reading news articles. Lastly, we could not find evidence that augmenting ads with game elements has a positive effect on the attitude towards the ad (**H11**). However, it should be noted that the perception of ads was measured using Professionalism and Appropriateness; thus effects might be found for other criteria, like product involvement or purchase intentions, as was shown in [45]. As a limitation, it should be noted that our results could be different outside lab conditions. In the wild, the current context of the user should be considered, because playfully deactivating ads might be disturbing.

3.4.6 Contribution to Research Questions

We found that gamification can enhance the user experience of online advertisements. More specifically, we found that all gamified concepts that were implemented significantly increased intrinsic motivation to interact with advertisements. We also found that the positive user experience seemed to positively affect the perception of the website hosting the advertisement, as the site organization rating increased for one gamification concept. However, using gamification elements on a news website should also be carefully implemented, since the value of news articles was partially found to be detrimentally affected by gamification. Besides motivation, we found that gamification has an effect on cognitive outcomes. Measures such as brand recall and recognition as well as product recall and recognition were significantly increased when gamification was used. Also, regarding users' intention to change their behavior, we found that one gamification concept ("Monster") was preferred over using an ad blocker. Although this is not a behavioral outcome per se, it suggests that the intention to use gamified approaches to interact with advertisements increased, which, in turn, might affect actual behavior. Overall, these findings contribute to **RQ1** since they show that gamification is capable of increasing enjoyment and thus intrinsic motivation in this specific behavior change context. The increased enjoyment also seems to positively affect other attributes of the website, which indicates that gamification affects not only a certain activity or behavior but also the attitude towards other entities in the context. Moreover, the increased enjoyment seems to favor increased cognitive engagement with advertisements and positively affect the intention to interact with gamified advertisements.

On a more abstract level, the findings could again be seen as supporting evidence for the theory that gamification elements which support the basic psychological needs that are inherently thwarted by the context are particularly successful in increasing motivation. In this context, the need for *autonomy* is thwarted, since users have no choice over consuming advertisements on websites. By allowing them to deactivate advertisements, this need for autonomy can be supported. Regarding the game concepts, both the Tetris and Paintball concepts mainly use gamification elements aiming at increasing perceived competence. In Tetris, users have to strategically rotate blocks falling from the top of the advertisement to the bottom. Thus, they have to apply puzzle-solving skills within a time limit to master this task. This kind of task is primarily focused on the competence of users. Similarly, in the Paintball concept, users have to shoot at the advertisement to cover it completely with color splashes. Since they only have a limited number of available shots, users have to come up with a certain strategy to solve this task. Thus, solving this kind of task is also primarily focused on the competence of users. In contrast, in the Monster concept, users are responsible for the well-being of a virtual monster. They can decide when to feed it with advertisements to improve its mood. Also, they can customize the virtual character as they see fit. They can unlock certain decorative virtual items, such as a hat they can put on the monster. The gamification elements used in this concept are focused on meeting autonomy needs and less focused on competence needs. Interestingly, this concept was the only one that participants preferred over using ad blockers, and it had the highest enjoyment ratings.

3.5 Summary

In this chapter, we sought to answer the question of how gamification affects motivation in behavior change contexts and which effects it may have on other psychological as well as behavioral outcomes (**RQ1**). Also, we wanted to gain a better understanding of how contextual factors may increase or diminish the importance of basic psychological needs. We investigated these questions in three different behavior change contexts – encouraging people to increase their physical activity, to adhere to hand washing practices and to strengthen their intention to consume advertisements.

We found that a significantly increased *relatedness* cleared the way for gamification to affect walking behavior in a context in which users already have a rather high fitness level and thus potentially a comparably high perceived competence (see Section 3.2). We learned that showing the gamified system publicly facilitated social interaction and enabled users to get in touch with others, even those who were not using the system (such as professional trainers in the gym). We found that such a gamified system increased step counts as well as self-reported physical activity levels significantly. However, we also learned that the inability to keep up with others, e.g. because of one's job or other occupations, detrimentally affects the perception of the system and may even lead to amotivation. Also, we saw that the gamification elements that were used did not affect all participants positively. These aspects show that the individual characteristics and preferences of a single user need to be considered to tailor gamified systems to that user and to increase the motivation of all users. These personalization aspects will be covered by **RQ2** and **RQ3**.

Moreover, when evaluating Germ Destroyer (see Section 3.3), we found that gamification elements focused on increasing the perceived *competence* of users (such as points, progression, or goals [370]) by providing informational, gameful feedback seem to affect intrinsic motivation positively in a context in which users have less knowledge about how to perform the target behavior correctly (and therefore potentially lack a feeling of competence). In addition, we learned that gamification evoked positively valenced emotional responses when washing one's hands, which may have been another reason why intrinsic motivation increased significantly, besides satisfying competence needs. In a field study we were able to show that the increased positive affect and intrinsic motivation ultimately led to users almost doubling the time they spent washing their hands. This supports the power of Germ Destroyer to change hand washing behavior and thus enhance public health. Also, the increased hand washing duration seemed to be the cause for a significantly decreased level of microbial contamination on the door handle of the bathroom, further adding to the positive effect on public health.

Lastly, we also considered the domain of online advertisements, which inherently thwarts the basic need for *autonomy*. In this context, we again were able to show that gamification increases intrinsic motivation. This increased motivation, and the increased interaction with online advertisements resulting from it, seemed to be the cause for the significantly enhanced cognitive outcomes such as increased brand or product recognition and recall. Notably, one gamification concept was preferred over the other proposed concepts and even over using a conventional ad blocker. In this concept, gamification elements were used which aimed at supporting the basic psychological need for autonomy by offering users the choice to decorate their virtual character with unlockable items and deciding for themselves when to care for it. Thus, meeting the need for autonomy might be particularly relevant in this autonomy-thwarting context.

To sum up, our findings demonstrate that gamification increased intrinsic motivation and led to changes in actual behavior, or the intention to change behavior, in all three contexts. These domain-specific insights contribute evidence for the power of gamification as a method to shape behavior and increase motivation in a gameful fashion. Moreover, we contributed insights on how gamification increases motivation and which SDT-related outcomes are affected by it. We showed that positively valenced emotional responses can be evoked by gamification and could be an explanation for why intrinsic motivation is increased. Also, we demonstrated the importance of basic psychological need satisfaction as a reason for an increased intrinsic motivation. We were able to show that meeting the need for relatedness can open the door for affecting behavioral outcomes of a gamified system positively. It also seemed as if gamification elements focused on increasing the perceived competence could be the cause for the increased intrinsic motivation in Germ Destroyer. The fact that, in a context in which their autonomy was inherently thwarted, participants preferred the game concept which utilized autonomy-focused gamification elements, adds to the relevance of basic psychological needs for the success of gamification. When recapitulating our findings across the contexts and considering the specific needs that these contexts supported and thwarted, it seems that gamification worked particularly well when the gamified system supported the needs that were otherwise inherently thwarted in the context. Therefore, to implement successful gamified systems, contextual factors should be considered to understand which basic psychological needs are diminished. Based on this, gamification elements can be selected supporting these diminished needs.

Chapter 4
Individual Factors Affecting the Perception of Gamification Elements

We have seen that gamification can affect motivation in different behavior change contexts, and that contextual factors could play a role in basic need satisfaction and the extent to which gamification elements support motivation. However, we have also seen that gamification elements which are perceived particularly well by some users may undermine the motivation of others in the same context. Thus, there must be factors beyond the context which affect the perception of gamification elements: factors that are tied to the person. Therefore, in this chapter, we will focus on the person, i.e. on individual factors which may differ between people. We will present a study investigating to what extent age plays a role in the perception of gamification elements in Section 4.2. Moreover, we will investigate behavior change intentions (operationalized by the stage of change construct of the transtheoretical model) as a potential mediator of how gamification elements are perceived in Section 4.3 and Section 4.5. Lastly, we will investigate the role of Hexad user types in explaining inter-personal differences in the perception of gamification elements in Section 4.4 and Section 4.5. Investigating these three individual factors mainly contributes to **RQ2**, i.e. the question of how personal factors affect the perception of gamification elements. In addition, we will gain knowledge on the effect of the context, since we will investigate the perception of gamification elements across two health contexts, physical activity and healthy eating, by using exactly the same approach for representing gamification elements. This allows us to better understand whether the preferences of users differ across these contexts, and provides insights into **RQ1**. Section 4.2 is based on [12], Section 4.3 on [16], Section 4.4 on [19] and Section 4.5 on [13].

4.1 Motivation

Both the findings in Chapter 3 and previous research have demonstrated that the perception of gamification elements differs substantially across users (for an overview see Klock et al. [182]). Therefore, gamification research focuses more and more on understanding why such inter-personal differences exist and how they can be explained [182, 239]. In this chapter, we contribute to this question and investigate the impact of several factors – age, behavior change intentions and Hexad user types – on the perception of gamification elements. These factors were selected since they are either tightly coupled to mini-theories of SDT or because they were empirically shown to have an effect on well-being and motivational orientations as defined by SDT.

Age was selected since previous research has demonstrated that the perceived importance of measures related to basic psychological needs differs in old age [283, 301, 302]. It was shown that the need to achieve new goals is less salient among older adults than among younger people [283], suggesting that satisfying competence needs might be less important in older age. In contrast, the importance of the basic psychological need for social relatedness increases in old age and plays a more important role [301, 302]. We expected that this shift in the importance of basic needs has an impact on the perception of gamification elements and on preferences in gamified systems.

Behavior change intentions, operationalized by the stages of change construct (see Section 2.2.1), play a major role in the internalization process as proposed by Organismic Integration Theory (see Section 2.1.2). As we have explained in Section 2.2.1, behavioral regulation becomes more self-determined in later stages of change. Since the type of motivation has an effect on which functional significance is given to a certain stimulus (or gamification element), we expect behavior change intentions to play an important role in how gamification elements are perceived. However, as far as we know, the extent to which the perception of gamification elements is affected by behavior change intentions has not been systematically investigated. We contribute toward closing this gap by investigating the impact of behavior change intentions on the perception of gamification elements in a fitness context. On a more general level, we also shed light on how behavior change intentions affect the perception of achievement goals.

Lastly, Hexad user types (see Section 2.2.3) were selected because they can be seen as personifications of basic psychological needs defined by SDT. The Hexad model is directly built upon organismic integration theory and basic psychological needs theory and can be seen as a proxy conceptualizing inter-personal motivational orientations and differences in the domain of gamification. Previous research has demonstrated its validity and reliability [337], and its capability for explaining user preferences in gamification [189, 261, 342].

4.2 Game and Gamification Preferences of Older Adults Aged 75+

In this first section of the chapter, we focus on age as a potential mediating variable in explaining preferences for gamification elements and individual differences in this regard. We start by investigating the preferences and attitudes towards playing games among older adults. In contrast to previous work, we consider people who are older than 75 in our study. The minimum age at which participants were considered as older adults in previous research ranges from 45 [100, 313] to 50 [172] or 65 years [55, 243, 293]. Thus, our population can be seen as underrepresented in previous research, although the benefits of gamified systems could be particularly relevant for this target group: Playing digital games has been shown to be associated with successful aging [222] and using gamification for behavior change could help older adults to remain physically, cognitively and socially active [133], which has positive effects on health and well-being [195]. The relevance of considering this target group is further supported by the demographic transition, which will lead to a substantial increase in the population of seniors in the near future [131]. Thus, a rapidly growing number of people can benefit from research in this domain.

After investigating general perceptions and attitudes towards gaming, we shift our focus to digital games and gamification. Understanding the preferences of older adults in gamified systems is important since they are mostly not used to game mechanics applied in digital games and might have different mental models [134]. Moreover, life goals and priorities differ between younger and older people [152] which might have an impact on what motivates or affects seniors and what they consider to be fun while playing. This is further supported by the fact that the importance of basic psychological needs changes in old age [283,301,302], with social relatedness becoming more important (as explained previously) [283]. We expect all these differences between younger and older adults to have an impact on how certain gamification elements are perceived. Understanding these differences enables us to tailor the gamification elements in gamified systems to older adults.

4.2.1 User Study

In this exploratory study, we did not have any a-priori hypotheses but instead investigated the following sub-research questions by combining qualitative and quantitative methods:

SRQ1: *Reasons for playing games:* What are underlying factors that motivate seniors to play games?

SRQ2: *Social play:* What social gamification elements are suitable for seniors? What do they fear or appreciate in competitive/cooperative settings?

SRQ3: *Perception of gamification elements:* How are frequently used gamification elements perceived by seniors?

SRQ4: *Differences to younger people* Do preferences for gamification elements differ between older and younger people?

We started by gathering information about the reasons for why older adults play games to implicitly infer conclusions about relevant gamification elements (**SRQ1**). Since previous research in the field of psychology suggests that social relatedness becomes more important in old age [283], we were particularly interested in this aspect and to what extent it has an influence on the preferences of older adults when playing games (**SRQ2**). After gathering insights on the overall perception of and preferences when playing games, we shift our focus to gamification elements specifically (**SRQ3**) and age-related changes in these preferences (**SRQ4**).

We recruited participants from three nursing homes and participants who are living on their own. To ensure that they do not suffer from severe mental diseases and are able to communicate without problems, we consulted the nursing management (having access to disease-specific diagnosis of all residents), who recommended participants. All participants agreed on taking part in this study voluntarily. The study started with a short questionnaire covering demographic data and gaming frequency (both for analog as well as digital games) with statements to be answered on 5-point Likert scales. A semi-structured interview followed to learn more about their gaming experience, their motivation for playing games and potential age-related changes in these aspects. The semi-structured interviews were directly transcribed and were conducted in face-to-face conversations alone with the participants in separate rooms (in the nursing homes) or in participants' apartments (for those living on their own). We followed a directed content analysis approach [156], i.e. we went through the transcripts to find themes related to each of the SRQs (e.g. themes for **SRQ1** included "socializing", "watch others play", and "cognitive benefits"). We then counted for each theme how many participants mentioned it and exemplary reported statements of participants related to the theme [96].

After the interview, we adapted the procedure of Orji et al. [262]: participants were shown seven storyboards, each explaining one gamification element in the scenario *"Motivate yourself to go for a walk"*. We decided to use a concrete scenario since context is crucial for a proper imagination of game concepts among older adults [293]. The scenario was chosen based on a literature review of persuasive systems [146] showing that a huge majority are conceptualized in the exercise domain. For the scenario to be relevant, we ensured that all participants were able to walk on their own. We also limited the amount of storyboards so as not to overwhelm them (cf. [321]) and chose the gamification elements points, badges, virtual characters, unlockables, competition, collaboration and progress bars, as they are frequently used [315]. For every storyboard, we asked questions like *"What is awarded to the person on the storyboard?"*, *"What does the person on the story-*

Figure 4.1: The storyboard illustrating the concept of the gamification element competition

board need to do to get the reward?" to make sure that the participants understood the underlying gamification elements. We then assessed for each gamification element whether it is considered to be motivational using 5-point Likert scales. To improve older participants' abilities to respond adequately [321] and to get insights about reasons why participants like or dislike different elements, we discussed their decisions in a semi-structured way. Exemplary, the storyboard for "competition" is shown in Figure 4.1.

Lastly, we determined the Hexad user type of participants to get further insights about motivating factors for this age group and to learn more about suitable gamification elements. To do so, we used the German Hexad questionnaire of Korbas [187]. We additionally classified a much younger sample using the same questionnaire. This was done to investigate whether there are age-related changes in game preferences. The survey consists of 24 statements that are divided into four blocks consisting of six elements, each representing one user type. We gave the printed statements in four blocks one after another to the participants to not overwhelm them (following the recommendations from Smeddinck et al. [321]).

The study took approximately one hour per participant and was approved by the Ethical Review Board of the Faculty of Mathematics and Computer Science at Saarland University[20]. During the study, drinks and snacks were provided to the participants and breaks could be taken at any time.

Results and Interpretation

18 German participants took part in the study (10 female, 8 male – 13 living in nursing homes and 5 living on their own) aged 84.61 years on average (Mdn=86, Min=75, Max=93). They reported not to be familiar with technology (M=1.5, SD=1.01, Mdn=1) but agreed to being interested in accumulating more experience with technology (M=3.83, SD=1.12, Mdn=4). In addition, participants reported

[20] Saarland University: *Ethical Review Board of the Faculty of Mathematics and Computer Science*, `https://bit.ly/2TK2Qii` (last accessed: 2021-12-01)

playing parlor games multiple times a week (M=3.72, SD=0.80, Mdn=4) but never play video games (M=1.39, SD=0.83, Mdn=1).

SRQ1: Reasons for Playing Games Participants reported to enjoy playing card games like Rummy or Skip-Bo, followed by the lottery game Bingo and board games like Ludo, Merels or Checkers. They emphasized to value the time spent playing: *"After playing games I have the feeling of accomplishment, that time was not wasted"*[21] **(P3)**. Even those that do not play regularly stated that this is mostly because they do not have people to play with: *"When my children were younger we used to have a whole cupboard full of games and played really a lot. Today I don't have people to play with. [...] I would definitely like to play more games again"* **(P16)**.

The main reason to play (mentioned by all), is socializing with others (*"In first place we meet to communicate. Playing games supports this by inducing a good mood"* **(P14)**). They also reported seeing games as a starter for conversations (*"The Bingo evenings here helped me to get in touch with other seniors living here"* **(P2)**). We furthermore found evidence that elderly people have fun watching others play and use the occasion to get in touch with them (reported by 5): *"Sometimes I just sit there and watch others playing. It is fun to see their reactions and it offers me the opportunity to talk to them"* **(P9)**. Despite socializing, 12 participants stated to value the perceived cognitive benefits that arise from playing games (*"A nice side effect of playing games is that I keep mentally fit"* **(P10)**).

SRQ2: Social Play Most participants (16) reported not to be driven by winning the game in first place but instead enjoy spending time with others: *"It is not about winning at all, it's about spending time together"* **(P3)**. While for 6 participants winning does not matter at all, 10 participants stated that they also like to win, but that is not most important:*"The main reason [to play] for me is to avoid being alone and enjoy time with others. However, winning a game is also nice sometimes"* **(P10)**. We also found that nearly all participants (17) prefer collaborating: *"We sometimes do teamwork when playing Skip-Bo. [...] Winning as a team makes me much happier than winning on my own"* **(P6)**. Since Skip-Bo is a competitive game, this statement underlines the strive for collaboration even more. In addition, the aspect of taking care of others was mentioned by almost all (17). They indicated to have a better experience when all players are satisfied: *"It is not too much about winning, it is more about ensuring everybody has a good time"* **(P9)**.

15 participants reported that they were more inclined to competition at younger ages: *"When I was young I was a swimmer and very ambitious [...]. Today I don't want to compete against others, those times are over"* **(P5)**. In addition, there is less pressure to win a game and a more relaxed atmosphere during play: *"I think what has changed is that we don't take things too seriously when playing"* **(P4)**. A majority (14) of participants stated that they value social contacts and communication with

[21] All statements were translated from German to English

others much more than in their younger years: *"Once you are old and live alone you realize that having people around you is the most valuable thing you can have"* **(P18)**.

SRQ3: Perception of Gamification Elements Considering the storyboards representing different gamification elements, participants rated both collaboration, i.e. working as a team to reach a goal together, and caring for a virtual character (we showed them a virtual dog), i.e. reaching a goal to make a virtual character happy, to be most motivational (see Table 4.1). When discussing what participants like about collaboration, reasons were related to statements mentioned in the interview. In addition, most of the participants (10) reported that collaboration includes all players when winning a game, regardless of their abilities, which is especially important at older ages: *"When collaborating, those that are not as fit as others also participate in winning the game, which is important as age brings disabilities that you cannot control. Being confronted with others reaching scores you will never be able to reach because of your disabilities is very demotivating"* **(P5)**. 12 participants reported that the virtual character provides a meaningful incentive for them, since they felt responsible for it: *"I would definitely do my best to make it [the dog] feel good"* **(P1)**. **P15** has some reservations and remarks that the character should not try to imitate a real pet too much, but should *"look like in a cartoon for example"* to be more authentic.

Customizing the virtual character, i.e. collecting/unlocking gifts for it, was considered to be slightly motivational. 8 participants liked this idea as they perceived it as a way to care for it, but on the other side four of them reported that the motivational impact strongly depends on the gift itself (*"Of course I would like to collect gifts or things for it, but I really need to have the feeling that the gift makes it happy and that the gift is suitable for it"* **(P16)**) and that they like to be able to decide what kind of gift is given to their virtual character (*"I want to be able to decide what gift is given to my pet [...] just like in real life I don't want to give generic presents, they need to be personal"* **(P6)**). Seeing the current progress towards a goal was considered neutral regarding its motivational effect. Asked for reasons, participants most of the time appreciated seeing their current progress towards a goal, but also reported that they are afraid not to be able to reach the goal, which would make them feel discouraged or sad.

Interestingly, points, badges and leaderboards were negatively perceived (see Table 4.1). Participants (14) stated that they do not see the benefit of earning points or emphasized that they miss the feeling of reaching something that is meaningful: *"I don't have the feeling of having reached something that has value. I don't see the benefit of collecting points"* **(P7)**. Concerning badges, they had similar concerns, with 8 participants that additionally disliked the visibility badges provide: *"I don't want to show these badges to anyone, I don't like putting myself in the foreground"* **(P3)**, *"Having those badges would put me under pressure. I would have the feeling that others expect me to perform even better"* **(P12)**.

The competitive aspect of leaderboards was criticized by all of our participants,

Gamification element	M	SD	Mdn
Collaboration	4.22	0.71	4
Virtual Characters	4.17	1.01	4
Unlockables/ Customization	3.39	0.89	3.5
Progress Bars	3.39	0.68	3
Badges	1.67	0.82	1.5
Points	1.56	0.83	1
Competition	1.44	0.60	1

Table 4.1: Mean ("M"), standard deviation ("SD") and median ("Mdn") of the perceived motivational effect of gamification elements, rated using 5-point Likert scales.

stating that they prefer playing for fun and leisure and do not want to make other players feel sad. Digging deeper into that, we also found that a reason to avoid competition comes from a fear of failure: *"I am afraid that people will think I am not fit anymore"* **(P16)**. **P15** adds that reasons or explanations for failures changed compared to when he was younger: *"When I lost a competition at a younger age, there were various reasons [...] you just had a bad day or something. When I fail today, people often attribute this to my age which is quite frustrating."* Participants additionally often stated being tired of competition: *"I don't want to compete against others, I had enough competition in my life"* **(P1)**.

SRQ4: Hexad User Types Considering the user types of seniors, the large majority of them (83.3%) were classified as Philanthropists, equally followed by Socializer (5.6%), Player (5.6%), Achiever (5.6%) and no participants being classified as Free-Spirits (see Figure 4.2a). To investigate age-related changes in game preferences, we compared this distribution to a much younger sample consisting of 31 participants that were recruited from our university (20 male, 11 female) having a mean age of 25.61 years (SD=4.64, Mdn=24). The user type distribution of the younger sample was much more diverse than in the older sample (supporting findings from Korbas [187] and from an online survey using a different user type instrument[22] both also considering a younger sample, see Figure 4.2c). 29.41% of the younger sample were classified as Philanthropists, followed by Free-Spirits (20.59%), Players (17.65%), Achievers (17.65%) and Socializers (14.71%) (see Figure 4.2b).

[22] Gamified UK: *Gamified UK User Type HEXAD Results*, `https://bit.ly/3x8paAw` (last accessed: 2021-12-01)

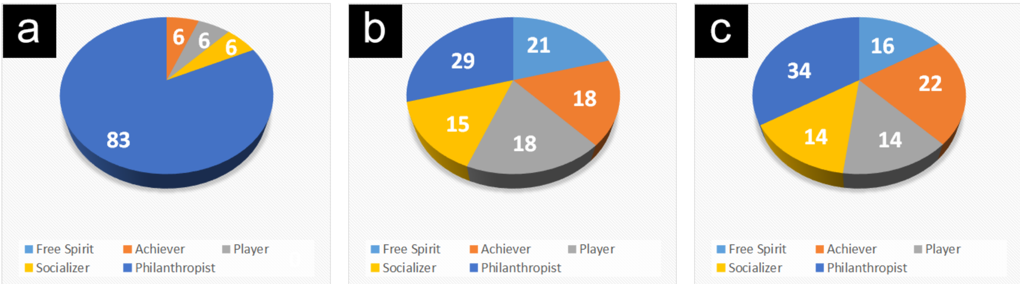


Figure 4.2: Distribution of primary user type classifications in different populations. a) User type distribution among elderly people (N=18). b) User type distribution in the younger age group (N=31). c) User type distribution reported by Korbas [187] (N=121)

4.2.2 Discussion and Limitations

We found that besides leisure and fun, socializing is a core motivator for participants to play games. This may come from the demand to socialize that was found to be higher among seniors [91]. We also found that collaboration and care-taking are motivating factors to reach certain goals, which may be explainable through research done by Cornwell et al. [91] showing that age is positively related with socializing and volunteering. We learned that socializing is a core motivator for seniors to play. More specifically, we found that seniors play to communicate and maintain social contacts and not to win the game in first place. In addition, they considered taking care of and collaborating with other players to be very important, as they do not want to make other players feel sad. These findings indicate that the increased importance of the basic need for relatedness in old age [283], might be transferable to the games- and gamification context. We therefore suggest using collaborative gamification elements over competitive ones and integrating communication capabilities to support the need for social exchange.

We moreover found indications that the importance of games as a catalyzer for social relationships increases in older ages and that winning the game is not as important as it is in younger ages. These changes are very related to the changes of the importance of basic psychological needs [283, 301, 302]. The increased importance of positive relationships relates well to the importance of games to build social relationships, to the aspect of care-taking and to the strive to ensure all players feel well. The changes are also reflected in the distribution of user types in our sample as a vast majority is classified as Philanthropists, a socially-minded, altruistic user type that loves to share knowledge and take care of others. The altruism and care-taking aspect underline the requirement to build positive relations with others even more.

Points, badges and leaderboards, which are widely used in gamified interventions [315], are negatively perceived by the participants. Concerning points, the main reason for the low perceived motivational influence on older adults was found to be the lack of value, which is explainable by findings from Wandke et al. [356] and Melenhorst et al [226], showing that older adults are more inclined to value meaningfulness. Besides that, badges were often reported to harm motivation as they provide a certain level of visibility, i.e. participants complained that badges put them in focus. A potential explanation for this might be lower self-esteem among elderly people: Robins et al. [291] investigated self-esteem during the life span and found that self-esteem decreases in old age. Keeping in mind that older adults like collaborating and care-taking, the low motivational effect of leaderboards is not very surprising. Reasons include the fear of failing, the fear of making other players feel sad and of compromising harmony or starting arguments.

Given that the younger sample we considered showed a much more heterogeneous user type distribution, it would be interesting to investigate reasons for this as we cannot reliably say whether people turn into Philanthropists with increasing age, whether the difference is attributable to the generation in which participants were born or whether the instrument we used to determine the user type is not suitable for older populations. Given our current results, we think that the perception of certain game aspects changes during life-span, as a majority of participants explicitly stated that they were more ambitious in games and that the aspect of social relationships in games was reasonably lower when they were younger.

Findings related to the gamification elements have two limitations: First, they were not based on interventions in which these gamification elements were applied but on subjective assessments of the participants. Second, we used only one specific context and did not investigate the gamification elements in other contexts as well. Addressing the first limitation, we decided to use storyboards over a software prototype due to similar reasons as mentioned in [262]: To avoid inducing confounding variables (visual attractiveness of the prototype, usability issues, issues related to the lack of experience with technology, concrete implementation details) as well as to provide a common visual language that is easier to understand and does not involve game- or technology-specific knowledge. Concerning the second limitation, we decided to use only one scenario so as not to overwhelm participants (which was shown to be especially important for seniors [321]). Participants' little to no experience with digital games is another limitation which may have an impact on the transferability of our findings to digital games. However, low experience with digital games is very common in this age group [135]. Moreover, the design of the storyboards themselves may have also had an effect on the perception of the gamification elements.

4.2.3 Contribution to Research Questions

Our findings show that age is an important factor to consider when conceptualizing and implementing gamified systems. We have seen that the changes in life goals and the shift in the importance of certain psychological needs for well-being and motivation in old age seems to affect the reasons why older adults are playing games and the perception of gamification elements. Instead of focusing on competitive aspects and performance when playing games, older adults in our sample valued the social aspects of playing games with others. They also leaned towards altruistic and supportive playing styles. This was also seen in the perception of commonly used gamification elements. The most frequently used gamification elements – points, badges, and leaderboards – were perceived especially unfavorably. Our findings can provide a good explanation for this, because all these gamification elements provide feedback on users' performance, are rather competitive, and may involve social pressure. Lastly, the fact that the Hexad user type distribution is completely different, compared to a younger sample, underlines the fact that there seems to be a shift in what motivates people when playing games and interacting with gamified systems with increasing age. A huge majority of older adults scored highest on the Philanthropist factors of the Hexad model, a social-oriented user type which is driven by meaning, purpose and altruism.

Overall, these findings contribute knowledge relevant to **RQ2**: They provide evidence for the importance of age as a personal factor affecting the perception of gamification elements and elucidate potential reasons for why the perception of gamification elements changes with increasing age.

4.3 The Impact of Behavior Change Intentions on the Perception of Achievement Goals in Fitness Systems

In this section, we shift our focus to behavior change intentions, i.e. the stage of users in the stages of change construct of the Transtheoretical Model (see Section 2.2.1). The intention to change behavior, operationalized by assessing the stage of change, can be seen as a proxy to assess the extent to which a behavior has been internalized and brought in congruence with one's own values and needs, as proposed by the Organismic Integration Theory of SDT (see Section 2.1.2). The shift from external and rather controlled types of motivation to more internal and autonomously regulated forms of motivation plays an important role in whether feedback and stimuli are perceived as informational, controlling or amotivating (see Section 2.1.2). Thus, we expect that changes in one's behavioral intention to perform a target behavior should have an impact on how certain gamification elements are perceived.

In this section, we contribute to this open question by investigating to what extent the perception of different types of achievement goals (see Section 2.2.2) is related to the behavioral intention of participants. We decided to investigate achievement goals, since gamification, being derived from games, is inherently a goal-oriented activity [105]. There are various ways for goals to be realized in gamified systems. For instance, they can be represented explicitly, e.g. as goals or quests, or they can be used implicitly, e.g. when unlocking achievements, earning badges, unlocking virtual items, collaborating to reach a team goal or competing on a leaderboard [340]. Ultimately, goals are at the heart of most gamification elements [340]. This means that understanding the perception of different types of goals and their potential impact on user behavior can likewise advance our knowledge about gamification elements.

We created visualizations for each of the three goal types of the 3x2 Achievement-Goal model. In a mixed-methods study, we investigated their general perception and perceived persuasiveness as well as potential correlations to behavioral intentions, using both quantitative and qualitative evaluation approaches. Again, we decided to use physical activity as the application domain, mainly due to the fact that research in this domain has the potential to improve public health and help to decrease the high prevalence of cardiovascular diseases. Also, it is one of the most commonly studied domains in persuasive technology research [145], increasing the impact of our research in academia and facilitating the replication of previous findings.

4.3.1 Goal Design Process

We use the Achievement-Goal Model (see Section 2.2.2) as a basis for the three types of goals. We started by a conducting a qualitative pre-study to elicit requirements for the realization of the goal types. After designing the three goal visualizations, we conducted another pre-study to investigate whether the visualizations are comprehensible.

Design Requirements Analysis

We conducted an online study on Prolific[23]. The study took approximately 8 minutes and participants were paid GBP 1. We asked participants *"Imagine you want to increase your daily step count. How would a step goal visualization have to look like in a fitness application to motivate you reaching this goal?"*. They could enter their response in a free-text field. 18 participants took part (11 female, 7 male; age: 18-24:4, 25-31:4, 32-38:5, 39-45:2, 46-52:2, 53-59:1), of which 50% reported to do sports on a regular basis. The written answers were analyzed by conducting an inductive content analysis [156] resulting in a set of overarching themes. Based on this, we derived the following requirements for the realization of the visualizations:

[23] Prolific: *Quickly find research participants you can trust*, `https://www.prolific.co/` (last accessed: 2021-12-01)

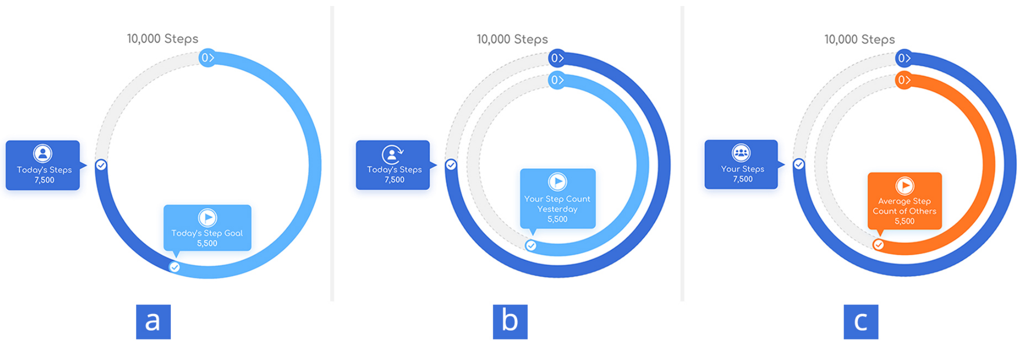


Figure 4.3: Goal visualizations for task-approach ("TAG") (a), self-approach ("SAG") (b) and other-approach goals ("OAG") (c).

RE1: Progress: 11 participants explicitly stated that real-time feedback is important to them.

RE2: Visualize progress through a circular chart: 55% of those participants that stated that real-time progress is important to them asked for visualizing step counts through graphs. Of those, 67% explicitly asked for charts using circular charts.

RE3: Show a concrete step count: 4 participants stated that showing a concrete step count is important to them.

RE4: Use bright colors: The graphs should be visualized using bright colors (mentioned by 2 participants).

Realization

The three types of goals to be designed are defined by the Achievement-Goal Model. In addition, the requirements **RE1** to **RE4** were considered. We decided to use circular charts to provide real-time feedback about the current step count of users (cf. **RE1, RE2**). To account for **RE3**, we decided to show a concrete step count in the visualizations. Since blue is in general a positively perceived color irrespective of demographic factors [29], we decided to use blue for visualizing progress in our designs (cf. **RE4**). To visualize the progress of other users in the other-approach goal, we decided to use the complementary of blue, orange. To realize the concept of the task-approach goal, a concrete objective is shown to the user, indicated by a respective icon and a label. To avoid confusion, the self-approach and the other-approach goal make use of a second circular chart in the inner circle of the graph to visualize one's own progress and the progress of other users respectively. To ensure the comparability of the visualizations, each one uses 5.500 steps as a goal. The final designs are shown in Figure 4.3.

Comprehensibility Analysis

Next, we wanted to ensure that participants understand each type of goal. We setup an online survey on Prolific showing participants each goal visualization one by one and asking them to describe them textually. Again, the study took approximately 8 minutes and participants were paid GBP 1. The textual descriptions were analyzed by two independent raters ("RA1", "RA2"). Their task was to rate how well each goal visualization was understood on a 3-point scale (1-very poor to 3-very well). Raters were told to assign the value "0" when there is not enough information to judge whether a participant understood a certain goal visualization. The neutral choice in the comprehensibility rating was used when the main concept was understood, but specific details were either not mentioned or misunderstood.

18 participants took part (11 female, 7 male; age: 18-24:6, 25-31:6, 32-38:1, 39-45:1, 46-52:1, 53-59:2, >59:1) of which 39% reported to do sports on a regular basis. When at least one rater rated a description as "0", the description was not considered for the analysis. This lead to the exclusion of 17 out of 54 descriptions. To ensure that the ratings can be interpreted objectively, we calculated the inter-rater agreement and found it to be Kappa=0.85, which is considered as almost perfect [221]. Analyzing the ratings of the two independent raters, we found that the participants understood the goal visualizations very well ($M_{RA1} = 2.90$, $Min_{RA1} = 2$; $M_{RA2} = 2.96$, $Min_{RA2} = 2$). Based on this, the three goal visualizations could be used to investigate their perception and perceived persuasiveness in the main study.

4.3.2 Evaluation

We investigate the persuasiveness of the goal types, reasons and the role of behavior change intentions.

Hypotheses

We expect to find evidence for the following hypotheses:

H1: The perceived persuasiveness and user preferences differ between the task-, self, and other-approach goals.

H2: Task-approach goals are perceived as more persuasive among people in high stages of change.

H3: Other-approach goals are perceived as more persuasive among people in high stages of change.

H4: Self-approach goals are perceived as more persuasive among people in low stages of change.

We expect that the visualizations are perceived differently (**H1**). This is supported by findings from [24], showing that self-based goals were more focused on improvement than on performance. **H2** is motivated by task-based goals affording a certain perceived competence to reach them and were shown to elicit responses related to performance [24]. **H2** is supported further by findings by Cham et al. [75] and Locke and Latham [206], showing that performance is associated with motivation and skill and that self-efficacy plays a major role in goal attainment. Similarly, **H3** is motivated by the importance of self-efficacy for the relevance of goals. Since comparing to other users might establish normative standards which seem to be out of reach for users in low stages of change, we expect that other-approach goals should be more relevant for users in high stages. This is supported by our findings presented in Section 4.5, showing that social gamification elements are more suitable for users in high stages of change. In contrast, establishing goals based on one's own performance should lead to reachable goals, which might be more suitable for users with lower self-esteem [75] in low stages of change (**H4**).

Method and Procedure

We conducted an online experiment on Prolific. It took approximately 12 minutes to complete and was approved by our Ethical Review Board[24]. Participants were paid GBP 1.50. After asking for demographic data, the stage of change was determined using a validated scale for the physical activity context [212]. Next, participants were shown each of the three goal visualizations individually in a random order and were asked to fill out the validated perceived persuasiveness scale by Thomas, Masthoff and Oren [163]. The scale consists of three factors (effectiveness, quality, capability) measured on 7-point scales. In addition to that, we also asked participants to describe what they like and dislike about the presented visualizations in a mandatory free-text field. The textual responses were qualitatively analyzed in order to understand *why* participants perceived the visualizations as persuasive or not. We analyzed the responses systematically by conducting an inductive content analysis [156] to identify patterns of meaning (themes). After being shown each of the three goal visualizations individually, participants were shown all visualizations at once (next to each other) and asked to select which of them they personally like the most to assess their overall preference. A Shapiro-Wilk test revealed that the persuasiveness scale items were not normally distributed, which is why we used non-parametric tests for our analysis. For correlation analysis, Kendall's τ was used, as it is well-suited for non-parametric data [155]. Kendall's τ is usually lower than Pearson's r for the same effect sizes. Therefore, we transformed interpretation thresholds for Pearson's r to Kendall's τ, according to Kendall's formula [355] (small: τ = 0.2; medium: τ = 0.3 ; large: τ = 0.5).

[24] Saarland University: *Ethical Review Board of the Faculty of Mathematics and Computer Science*, `https://bit.ly/2TK2Qii` (last accessed: 2021-12-01)

Results

		TAG Task-Approach	SAG Self-Approach	OAG Other-Approach
	Preference	M=0.25 SD=0.44 Mdn=0.00	M=0.44 SD=0.50 Mdn=0.00	M=0.31 SD=0.46 Mdn=0.00
Perceived Persuasiveness Scale	Persuasiveness	M=5.20 SD=0.94 Mdn=5.33	M=5.29 SD=0.94 Mdn=5.44	M=5.15 SD=0.98 Mdn=5.33
	Effectiveness	M=4.83 SD=1.37 Mdn=5.00	M=4.98 SD=1.35 Mdn=5.33	M=4.75 SD=1.45 Mdn=5.00
	Quality	M=5.26 SD=1.00 Mdn=5.33	M=5.31 SD=0.99 Mdn=5.33	M=5.05 SD=1.07 Mdn=5.00
	Capability	M=5.50 SD=1.07 Mdn=6.00	M=5.59 SD=0.98 Mdn=6.00	M=5.64 SD=1.08 Mdn=6.00

Table 4.2: Descriptive data of dependent variables for each goal visualization. TAG=Task-Approach Goal, SAG=Self-Approach Goal, OAG=Other-Approach Goal

We excluded participants who answered one of three test questions incorrectly, leading to a final answer set of 118 responses (64 female, 54 male; age: 18-24: 24, 25-31: 38, 32-38: 18, 39-45: 19, 46-52: 9, 53-59: 6, ¿59: 4). 12 participants were in the precontemplation, 18 in the contemplation, 28 in the preparation, 20 in the action and 40 in the maintenance stage of change. 49 participants were not doing any kind of sports whereas 69 did.

Differences Between Goal Visualizations The mean and median scores for all dependent variables can be found in Table 4.2. All goal visualizations were perceived as persuasive, as revealed by one-sample Wilcoxon signed rank tests against the neutral choice of four on the 7-point scale. The persuasiveness score is significantly higher than four for all goal types (each $p<0.001$). This leads to result **R1: All goal visualizations are perceived as persuasive**. Next, we analyzed whether there are differences between the goal visualizations. We calculated a Friedman's ANOVA for all dependent variables shown in Table 4.2 and used the Durbin-Conover method for post-hoc analysis. The Benjamini-Hochberg false discovery rate [40] was used to adjust significance values for multiple comparisons. We found a significant effect for the user preferences ("Preference") ($\chi^2(2)$=6.58, $p<0.05$) and for the "Quality" factor of the perceived persuasiveness scale ($\chi^2(2)$=15.20, p<0.01). For all other variables, no effects were found. The post-hoc analysis revealed **R2: Participants prefer self-approach goals over task-approach goals** (p=0.039). Regarding the "Quality" factor of the perceived persuasiveness scale, relating to the trustworthiness of the goal visualizations, pairwise comparisons revealed that **R3: Participants considered self-approach goals as more trustworthy than other-approach goals** (p=0.003) as well as **R4: Participants considered task-approach goals as more trustworthy than other-approach goals** (p=0.003). No further significant differences were found.

Effect of the Stage of Change To analyze the effect of the stage of change on the perceived persuasiveness of the three goal visualizations, we analyzed whether the a-priori formulated relationships (cf. **H2–H4**) exist by calculating one-tailed bivariate correlations between the stage of change and the items measuring perceived persuasiveness. We found that the overall persuasiveness of both the task-approach (τ=.18, p<.01) and other-approach goals (τ=.16, p<.05) are positively correlated with the stage of change. More specifically, we found that the "Effectiveness" (τ=.16, p<.05) and "Capability" (τ=.16, p<.05) factor of task-approach goals are positively correlated with the stage of change. In sum, we formulate **R5: Task-approach goals are perceived as more persuasive with increasing stages of change, mostly because of higher "Effectiveness" and "Capability" scores**. For other-approach goals, we found a positive correlation for the "Quality" factor (τ=.17, p<.01), leading to **R6: Other-approach goals are perceived as more persuasive with increasing stages of change, mostly because of a higher "Quality"**. For self-approach goals, no correlations were found.

Qualitative Analysis To better understand the underlying reasons for the quantitatively found effects, we analyzed the textual responses by participants qualitatively. For each participant and visualization, one textual response was recorded, resulting in 354 responses that have been analyzed. First, we analyzed whether participants perceived the goal visualization negatively, neutral or positively by assigning values from 1–3 to each textual summary (1=negative, 2=neutral, 3=positive). On average, participants were rather neutral about task-approach (M=2.08, SD=0.49), self-approach (M=2.14, SD=0.60) and other-approach goals (M=1.92, SD=0.69). A Friedman ANOVA revealed that there is a significant difference between these values (p=0.043). Pairwise comparisons using the Durbin-Conover method and the Benjamini-Hochberg false discovery rate [40] revealed that self-approach goals were coded to be perceived more positively than other-approach goals, which is similar to **R3**. No effects were found between task-approach goals and other approach goals.

When analyzing what participants like and dislike about the goal types, several themes emerged that might explain the results that have been found based on the quantitative analysis. Themes are written in bold italics. First, the fact that participants in general preferred self-approach goals over task-approach goals

(**R2**) seems to be related to participants considering self-approach goals as ***meaningful***. This theme was found consistently across the free text answers about aspects that participants liked about self-approach goals. Self-approach goals are considered as *"personally relevant"* (**P115**), and considered to *"give you a reason to push yourself against your own goals [...]"* (**P115**). In contrast, task-approach goals were considered as ***arbitrary*** or ***meaningless***. Participants reported that *"it lets you set an arbitrary goal that may not mean as much"* (**P105**). A main reason for why participants considered self-approach goals as more meaningful seems to be related to ***self-improvement***. Participants said that *"I like that it shows I am improving so I'd feel good about that"* (**P73**) or that *"seeing this makes me want to improve upon my previous record"* (**P104**). In addition, participants consider self-approach goals as more ***healthy***, i.e. they liked that the self-approach visualization establishes moderate, reachable goals. **P25** notes that *"it can work slowly towards achieving the goal"*. This is supported by **P47**, stating that *"it can make me more competitive with myself in a fun and healthy way"*.

Regarding **R3** and **R4**, trust seems to play a major role. This is supported by the thematic analysis, revealing that the major drawback of other-approach goals is missing ***trust***. We found that participants were afraid that other users might ***cheat*** to increase their step counts or that ***technology*** is not capable to reliably measure the steps taken. **P67** notes that other people *"may not be doing the right thing anyway"* and **P68** states that *"I would doubt the accuracy of the data [...] or want more information about where it comes from"*. In addition, our analysis revealed that ***comparability*** is a major concern. Participants frequently stated that they do not have enough information to judge whether others are comparable in terms of their ***fitness level***, their ***demographics*** or their ***circumstances***. A statement by **P77** summarizes this: *"Circumstances are different for everyone. The people who walk more could have more time on their hands, could be walking a lot in their work so it wouldn't influence me to exercise more"*. In addition, participants were concerned about ***over-training*** when using the other-approach goal. They noted that seeing other users' step counts might lead to peer-pressure which may result in people doing more than is good for them. **P79** states that *"My targets are based on my health needs and not on what others are doing"* and emphasizes that the other-based goal *"would encourage me to do more than my limbs may be ready for by tapping into my competitive spirit"*. However, in line with **R1**, participants also reported positive aspects about task-approach and other-approach goals. They like that task-approach goals are ***objective***, ***simplistic*** and consider them as ***reliable***. **P99** states that *"I like how accurate it is"* and **P78** supports this by stating that *"It is motivating and something concrete to base exercise on"*. Also, participants like that task-approach goals are not related to one's own performance and thus are perceived as rather ***non-binding***: **P81** states that *"It encourages the reaching of goals without worrying about shortfalls prior to the current day"* which is supported by **P71** stating that *"I like the simplistic approach and that there isn't a comparison to anything, it's just your step count and whether you have beaten your daily goal"*. For other-approach goals, participants liked that it may push ***self-efficacy*** when one's own performance is better than those of others: *"it leads me to believe that I am achieving above average results which makes me feel good about myself"* (**P64**). Participants also reported that ***competitiveness*** is a strong motivator for them: *"I like that it keeps people competitive"* (**P87**).

4.3.3 Discussion and Limitations

Our results show that in general, all three goal types are perceived as persuasive (**R1**). This might be related to the fact that all visualizations establish goals, which has been shown to affect action [206]. However, differences between the three goal visualizations were found. First, we found that participants preferred self-approach goals over task-approach goals (**R2**). Based on the qualitative analysis, it seems that participants appreciated that self-approach goals support self-improvement and thus considered these goals as more meaningful. This is in line with findings from Ansems et al. [24] who compared task-based against self-based goals in a dance game and found that participants responded more in terms of self-improvement in the self-based condition. Second, we found that both self-approach and task-approach goals were perceived as more trustworthy than other-approach goals, as revealed by significantly higher scores on the "Quality" factor of the perceived persuasiveness scale (**R3, R4**). When analyzing reasons for what participants did not like about other-approach goals qualitatively, we found supporting evidence for this effect, since trust emerged as a main theme. When further unfolding this, we learned that participants did not trust the data of other users mainly because they expected them to cheat and because they are concerned about measuring errors of step counters. These findings are in line with results by Niess and Woźniak [254] who found that building trust in the goal and in the fitness tracker is important for the goal to be meaningful. Thus, taking **R2–R4** together, **H1: The perceived persuasiveness and user preferences differ between the task-, self, and other-approach goals** is partially supported.

We furthermore learned that the stage of change is positively correlated with the overall perceived persuasiveness of task-approach goals. The "Effectiveness" and the "Capability" factors are positively correlated with the perceived persuasiveness of task-approach goals (**R5**). Given that objectiveness emerged as a main theme when analyzing what participants liked about task-approach goals, it seems that participants appreciated to have a clear goal which allows attaining task-based competence. This seems to be a reasonable explanation for why there is a positive relationship between the stage of change and the perceived persuasiveness of task-approach goals when considering the findings presented in Section 4.5. Here, we show that gamification elements such as challenges and badges, establishing clear goals allowing to evaluate how well or badly a task was solved, were perceived as significantly more motivating by participants in high stages of change. As such, we consider our results as supporting evidence for **H2: Task-approach goals are perceived as more persuasive among people in high stages of change**. We found that the perceived persuasiveness of other-approach goals is positively correlated with the stage of change of participants. Besides finding a positive correlation between the overall perceived persuasiveness and the stage of change, we also found a positive correlation between the ”Quality” factor of the perceived persuasiveness scale and the stage of change (**R6**). Thus, it seems like the perceived trustworthiness, which is measured by the ”Quality” factor [163], is the deciding cause for this positive correlation. Again, this is supported by the qualitative analysis, revealing that missing trust (mostly because of expecting other users to cheat and a low perceived accuracy of technology) is a main theme that emerged. Given that low ability to perform a task has been shown to be a key factor for cheating behavior [314] the positive correlation to the stage of change seems reasonable. Additionally, the positive correlation between the stage of change and the perceived persuasiveness of the other-approach goal is in line with [13], showing that leaderboards are perceived as significantly more persuasive by users in high stages of change. These results support **H3: Other-approach goals are perceived as more persuasive among people in high stages of change**. When analyzing a potentially negative correlation between the stage of change and the perceived persuasiveness of self-approach goals, we could not find significant effects. It seems like self-approach goals are perceived positively across all stages of change. This is in line with findings from the thematic analysis, since no themes emerged in this regard. The fact that self-approach goals adapt to the personal performance of participants, which in turn encourages intrinsic motives such as self-improvement, seems to stimulate both participants in high and low stages of change. Again, this seems to relate to our results presented in Section 4.5, where we found that both people in low and high stages of change perceived personalization, i.e. a system adapting the step goal to individuals, positively. Thus, based on our results, we did not find evidence for **H4: Self-approach goals are perceived as more persuasive among people in high stages of change**.

Design Guidelines

Based on both the quantitative and qualitative results, we establish the following set of design guidelines:

Use self-approach goals when having no information about a user’s stage of change

Our results show that self-approach goals are perceived as persuasive and are preferred over task-based and other-based goals, independent of the stage of change of users (cf. **R1, R2, R3**). Therefore, we generally recommend to use self-approach goals in systems encouraging physical activity in order to support self-improvement, which was shown to be perceived as meaningful by participants. Also, this type of goal was considered as healthy, since it is based on one’s own performance.

Use task-approach or other-approach goals for users in high stages of change

The findings (**R5**) show that task-approach goals are more relevant for people in higher stages of change. The fact that task-based goals establish a clear objective and allow to easily evaluate whether or not it has been met seems to be the deciding factor for the positive perception. Instead of focusing on self-improvement, task-based goals focus on mastery [24], which has been found to be more relevant for users that internalized their behavior [13].

Regarding other-approach goals, our results show that they should be more relevant for users in high stages of change (**R6**) and that users like other-approach goals mainly due to the inherent competitiveness of this type of goals. That competition is more relevant for users in high stages of change has also been shown in previous research [13], supporting our findings.

For self-approach goals, support self-improvement

Through the thematic analysis, we found that self-improvement is a strong motivator when using self-approach goals as it is perceived as meaningful. Therefore, we recommend to focus on supporting self-improvement when using self-approach goals. This could be achieved by highlighting personal growth, e.g. by visualizing trends of physical activity over time or by introducing metrics making self-improvements more graspable such as showing the relative improvement over a certain timespan.

For task-approach goals, avoid arbitrariness

Participants appreciated that task-based goals introduce a clear, reliable target which can be objectively measured. However, there is as risk that this might seem arbitrary and thus meaningless to users. Therefore, we suggest to make these goals more meaningful, by adding personal relevance through comparisons to the real world that make an arbitrary number more graspable. This could be achieved by comparing the step goal to a distance that people might be able to relate to (such as the distance between two cities that are known to users).

For other-approach goals, focus on transparency, comparability and avoiding over-training

Missing trust emerged as a main theme. This is in line with our quantitative results (**R3, R4**). Therefore, we recommend to communicate and explain transparently how the data of others has been measured and aggregated. Moreover, since comparability (in terms of demographics or fitness level) has been raised as another major concern, we recommend to provide information about the sample that the individual is compared to or even select a subset of other users which is comparable to the individual in terms of fitness level and demographics. Moreover, as revealed by the qualitative analysis, measures to prevent over-training should be incorporated.

Limitations

We used static visualizations to assess the perceived persuasiveness of each goal type; we did not implement them. Although this has several advantages such as reaching a higher number of participants and abstracting from specific implementation choices, which could bias the results in a way that is hard to control [262], validating our findings using real implementations is an important next step. Also, the fact that we decided to show visualizations in which participants had already reached their goal, might affect the perceived persuasiveness of the goal visualizations. Although we tried to inform and validate the realization of the three goal types by two pre-studies, it should be noted that there might be other realizations of the types of goals leading to different results.

4.3.4 Contribution to Research Questions

In general, without considering the individual person, the self-approach goal was preferred over task- and other-approach goals. Defining competence based on one's own performance stimulates self-growth and was able to elicit feelings of meaningfulness and trustworthiness among our participants. Also, task-based goals were seen as more trustworthy than other-approach goals. Participants were concerned about other users who might cheat or have a different fitness level, but they also raised concerns about the accuracy of technology, which might explain these results. Overall, the insights about the perception of the goal types mainly contribute not to understanding individual differences, but rather to understanding how gamification affects motivation in behavior change contexts (**RQ1**).

Regarding individual differences, we found that task-approach goals are perceived as more persuasive among people in higher stages of change. The effectiveness and capability factors of the perceived persuasion scale were the deciding cause for this. This indicates that perceived competence, i.e. the ability to reach a certain goal, plays a role in how effective a certain stimulus is in increasing motivation. Since the perceived competence might be rather low in low stages of change, the positive correlation between task-approach goals and behavior change intentions seems reasonable. We also found a positive correlation between the perceived persuasiveness of other-approach goals and the stage of change of participants. Since other-approach goals focus on performance and because perceived competence plays a major role in the effect of performance goals on intrinsic motivation (see Section 2.2.2), this effect seems plausible: Intrinsic motivation is more likely to be undermined when perceived competence to reach a performance goal is low. Overall, these findings show that behavioral intentions should be considered when tailoring gamified systems to the person, and contribute toward answering **RQ2**.

4.4 The Effect of Hexad User Types on the Perception of Gamification Elements in Healthy Eating

In this section, we focus on the role of Hexad user types in explaining preferences for gamification elements. More specifically, we are interested in understanding whether the score on a certain Hexad user type is correlated to how persuasive a certain gamification element is perceived to be. We assume that the inter-personal differences in motivational orientations, laid out by the Causality Orientations Theory of SDT (see Section 2.1.2), should have an impact on the user types with which users most closely identify themselves. Since these user types are tied to the basic psychological needs and extrinsic motivations, the perception of gamification elements should differ among these user types.

To investigate this relationship, we created storyboards explaining commonly used gamification elements. These storyboards are validated in a qualitative pre-study to make sure that they are easy to understand and describe the gamification elements as intended. Instead of considering physical activity, we contextualized our study in the domain of healthy eating, which is, next to physical activity, considered a core determinant of a healthy lifestyle [139]. Given that overweight because of poor eating habits is among the most important health issues today [103] and has been associated with chronic diseases such as cancer, cardiovascular diseases and diabetes [103], using technology to cultivate healthy eating habits is an important research topic.

4.4.1 Storyboards and Gamification Elements

We decided to use storyboards to explain the gamification elements in order to give participants a better idea of how those elements work. For the storyboards, we ensured to have at least one gamification element for each user type (based on [213, 342]), resulting in twelve different storyboards. A list of gamification elements and expected correlations can be found in Table 4.3. Figure 4.4 shows an exemplary storyboard. All storyboards are licensed under a CC BY 4.0 license and can be found in full resolution on figshare[25].

Storyboard Validation

We conducted a study to make sure that the storyboards are comprehensible (similar to our previous work, in which we created storyboards for the physical activity domain [13]). In this study, a semi-structured interview was conducted, in which participants were shown each printed storyboard in random order. First, participants were asked to describe the storyboard in their own words. When necessary, the interviewer asked questions to prompt participants to identify which

[25] figshare: *Storyboards illustrating Gamification Elements to encourage Healthy Eating*, `https://bit.ly/3pLndaH` (last accessed: 2021-12-01)

Gamification Element	Short Storyboard Description	Expected UT
Virtual Character	*The appearance of a virtual character is linked to calorie intake of the user.*	AC, PL
Custom Goal	*The user set themselves a custom calorie intake goal.*	AC, FS
Personalized Goal	*The system personalizes the users' calorie intake goal.*	AC
Challenge	*The user manages to reach a demanding goal.*	AC
Badges	*The user reaches their goal three times, unlocking a new badge.*	AC, PL
Points	*The system rewards the user with points for eating healthy.*	PL, AC
Rewards	*The user receives a coupon code for staying blow the daily calorie intake limit.*	PL
Knowledge Sharing	*The user helps others in a forum by answering questions.*	PH
Unlockable Content	*Staying belo the daily calorie intake limit three times in a row unlocks a new feature.*	FS
Cheating	*The user decides to cheat by entering wrong data about the food intake into the app.*	DI
Social Collaboration	*A group of users have to collaborate, to stay below their shared calorie intake limit.*	SO
Social Competition	*A group of users are shown on a leaderboard, competing for the top position.*	SO, PL

Table 4.3: Gamification elements included in the main study, a short textual description explaining what is depicted in the corresponding storyboard and the user types ("UT") we expect to be positively affected by them based on [213, 342].

activities are depicted by the storyboard. Next, participants were given a short textual summary of each gamification element. They were asked to assign each of the storyboards its respective gamification element by using the short textual summary. This was done to get an additional indication of whether participants understood the storyboards. Finally, interviews were transcribed and analyzed by two independent raters ("R1", "R2"). They received the transcriptions for each storyboard, without revealing which gamification element was described by the participants. Their tasks were to evaluate which element was being described and to rate how well the element was understood on a 5-point scale (1-very poor to 5-very well).

8 German participants took part (4 female, 4 male, average age 21.75). They

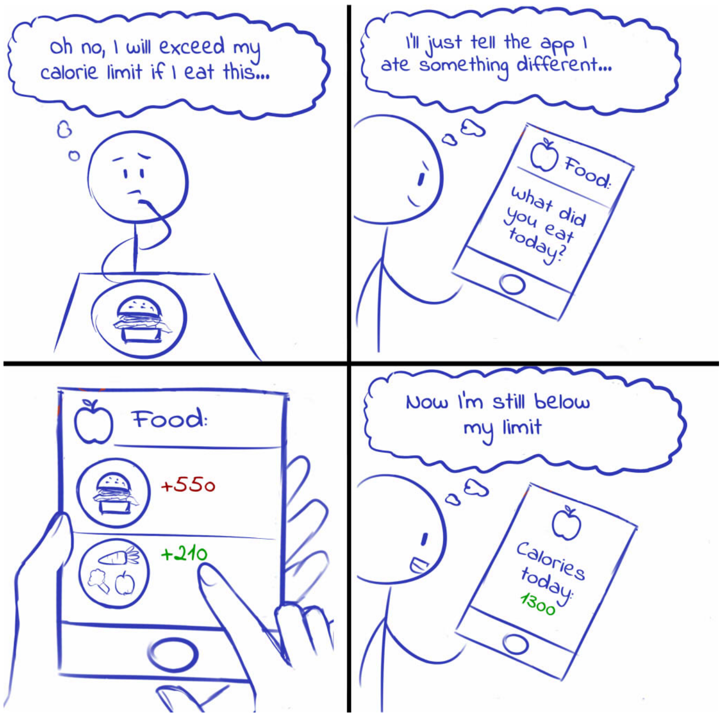


Figure 4.4: Storyboard of the gamification element "Cheating"

claimed to be gaming-affine (M=5.75, SD=1.40, Mdn=6), to frequently play video games (M=6.00, SD=1.41, Mdn=7) and to have a passion for video games (M=5.63, SD=1.85, Mdn=6.5). To ensure that the ratings can be interpreted objectively, we calculated the inter-rater agreement and found it to be Kappa=0.63, which is considered as substantial [221]. Analyzing the ratings of the two independent raters, we found that the participants understood the storyboards very well (M_{R1} = 4.89, Min_{R1} = 4; M_{R2} = 4.92, Min_{R2} = 4). This was supported by the fact that both raters successfully assigned all gamification elements correctly based on participants' storyboard descriptions.

4.4.2 Evaluation

We conducted an online survey, which was available in English and German. Participants were recruited via social media and Academic Prolific (participants were paid 1.50 pounds). The study took 10-15 minutes to complete and was approved by our Ethical Review Board[26]. After collecting demographic data, participants were asked to rate their gaming affinity on a 5-point scale. Next, participants' Hexad user type was determined using the 24 item Hexad User Types scale developed by Tondello et al. [342]. Finally, as the main part of the

[26] Saarland University: *Ethical Review Board of the Faculty of Mathematics and Computer Science*, `https://bit.ly/2TK2Qii` (last accessed: 2021-12-01)

	Mdn. PP	AC	DI	FS	PH	PL	SO
Virtual Character	4.25	**.146****	-	.152**	.134**	**.147****	.186**
Custom Goal	**5.00**	-	-	-	.102*	.036*	.112*
Personalized Goal	**5.25**	**.095***	-	-	.097*	-	.095*
Challenge	**5.00**	**.114***	-	.139**	.191**	.125**	.180**
Badges	**4.50**	-	-	-	.164**	**.113***	.200**
Points	**4.50**	**.120****	-	-	.106*	**.152****	.124**
Rewards	**5.75**	.135**	-	.100*	.142**	**.248****	.143**
Knowledge Sharing	**4.25**	-	-	.132**	**.260****	.121*	**.248****
Unlockable Content	**5.00**	.097*	-	**.115***	.196**	.109*	.234**
Cheating	**3.00**	-	-	-	-	-	-
Social Collaboration	**4.75**	-	-	-	.255**	-	**.268****
Social Competition	4.50	-	-	-	-	**.223****	**.143****

Table 4.4: Median perceived persuasiveness of all gamification elements ("Mdn. PP", colored cells indicate a significant ($p<.05$) deviation from the neutral choice (red=less, green=more persuasive)) and bivariate correlation coefficients (Kendall's τ) between the Hexad user types and the gamification elements (bold entries represent correlations that we expected (see Table 4.3)). *$p<.05$, **$p<.01$

questionnaire, participants were shown the 12 storyboards in a randomized order. To measure the persuasiveness of each gamification element depicted in the storyboards, we adapted the perceived persuasiveness scale by Drozd et al. [111] in the same way as was done by Orji et al. [259]. The scale consists of four items to be answered on 7-point Likert scales. A Shapiro-Wilk test revealed that the persuasiveness items were not normally distributed, which is why we used non-parametric tests for our analysis. For correlation analysis, Kendall's τ was used, as it is well-suited for non-parametric data [155]. It should be noted that Kendall's τ is usually lower than Pearson's r for the same effect sizes. Therefore, we transformed interpretation thresholds for Pearson's r to Kendall's τ, according to Kendall's formula [355] (small effect: $\tau = 0.2$; medium effect: $\tau = 0.3$; large effect: $\tau = 0.5$).

Demographics

237 participants completed the online survey. 38.4% were male, 60.8% female and 0.8% identified themselves as "non-binary" or "genderqueer". Most participants (35%) were aged 18-24 years, followed by 25-31 (31.6%), 32-38 (14.3%), 39-45 (8%), 46-52 (4.2%) and younger than 18 (4.2%). The remaining participants were

aged 53 and older (2.5%). They considered themselves as gaming-affine (M=3.58, SD=1.11, Mdn=4.00), claimed to have a passion for video games (M=3.58, SD=1.03, Mdn=4.00) and to frequently play video games (M=3.44, SD=1.16, Mdn=4.00).

Perceived Persuasiveness of Gamification Elements

The median scores of perceived persuasiveness can be found in Table 4.4. It can be seen that all but the ”Cheating” gamification element scored higher than the neutral choice of 4 on the 7-point scale. To analyze whether the median scores significantly differ from the neutral choice, we calculated one-sample Wilcoxon signed rank tests for each gamification element. All gamification elements but ”Virtual Character” and ”Social Competition” differed significantly ($p<.05$) from the neutral choice. Of those elements, ”Cheating” is the only one where the perceived persuasiveness is significantly lower than 4, whereas the rest scores significantly higher. These results suggest that most gamification elements should have positive effects on user behavior when being implemented and may help to select gamification elements to encourage healthy eating, when no information about the target audience or their user type distribution is known.

Figure 4.5: Distribution of participants’ Hexad scores

Hexad User Types and Gamification Elements

The distribution of participants’ scores across all Hexad subscales is shown in Figure 4.5. Regarding the correlations between Hexad user types and the perceived persuasiveness of gamification elements depicted in the storyboards, we found 13 correlations between user types and gamification elements out of 17 expected ones (see Table 4.4). Besides replicating previous research in the context of healthy eating [261, 342], we contribute correlations between gamification elements and the Philanthropist, which have been hypothesized, but not yet shown. These results may help to further personalize the set of gamification elements when implementing or designing behavior change applications to encourage healthy eating.

4.4.3 Discussion and Limitations

We investigated the perceived persuasiveness of twelve commonly used gamification elements by creating storyboards explaining each element, ensuring their comprehensibility (N=8) and presenting them to users in an online study (N=237). Our results show that most gamification elements scored significantly higher than the neutral choice on the perceived persuasiveness scale. Thus, these results may help researchers and practitioners to inform the design of gamified behavior change support systems encouraging healthy eating.

Confirming previous findings [261, 342], we also found that the Hexad user type is a useful factor for personalization of gamified systems. Besides replicating previously found correlations between gamification elements and Hexad user types in the context of healthy eating, we contribute a set of new correlations, which were expected in previous works [213, 342], but have not been shown before. This might be due to using storyboards rather than textual descriptions as in [342] and because of using a concrete context rather than a general context, also as in [342], potentially leading to a more concrete idea of how the elements work. Taking our results together, we show that certain gamification elements seem promising to encourage healthy eating and that Hexad user types are worthwhile to consider as a factor for personalization of such systems, extending previous work using the Hexad model.

This study has several limitations that should be considered. First, we used storyboards to assess the perceived persuasiveness of each gamification element; we did not implement them, i.e. we investigated perceived persuasiveness, not actual persuasiveness. Therefore, validating our findings using real implementations is an important next step that should be followed. Second, even though we investigated atomic gamification elements using storyboards, some aspects of the realization of these gamification elements are inherently a matter of interpretation, affecting the external validity of our results when implementing gamification elements differently. Additionally, it should be noted that combining gamification elements may create different experiences for the user, which has not been investigated and should be analyzed in future work. Also, we cannot say whether our findings generalize to different contexts besides Healthy Eating. Therefore, further research should be conducted about the Hexad user model as a factor for personalization in different contexts. Last, we would like to acknowledge that calorie intake is not the only factor of a healthy diet and that the use of this metric should be seen as a design decision to conceptualize and simplify healthy eating for the purpose of the study.

4.4.4 Contribution to Research Questions

Besides providing insights on the general perception of commonly used gamification elements in the healthy eating domain, we found that there are correlations between the perceived persuasiveness of the gamification elements and Hexad user types. This shows that Hexad user types should be considered to personalize gamified systems to the user. More specifically, we found that users who score particularly high on a certain user type have particularly positive perceptions of gamification elements that mainly address the basic psychological need underlying that respective user type. Also, we were able to show that the correlations established by previous works in different contexts could be replicated, supporting the validity of the Hexad model.

Overall, these findings contribute to **RQ2** since they show that Hexad user types play a role in explaining inter-personal differences in how gamification elements are perceived. The set of significant correlations can be seen as a set of guidelines to inform the design of gamified systems promoting healthy eating, and thus contribute answers to the question of how personal factors affect the perception of gamification elements.

4.5 The Impact of Behavior Change Intentions and Hexad User Types on the Perception of Gamification Elements in Fitness Systems

In this last section of the chapter, we will investigate the impact of both behavior change intentions and Hexad user types on the perception of gamification elements in the physical activity domain. This allows us not only to replicate and extend our previous results but also to analyze whether both these factors should be combined. We will compare the correlations between the perception of gamification elements and Hexad user types of users in low and high stages of change, to analyze whether the correlation strength and direction to Hexad user types is affected by users' behavior change intentions. Since Hexad user types represent general, rather static motivational orientations, and behavior change intentions represent the more dynamically changing extent to which a behavior has been internalized, we expect these two factors to complement each other.

In line with the study presented in Section 4.4, we created storyboards illustrating the same commonly used gamification elements. In contrast to that previous study, we situated the gamification elements in the domain of physical activity. This allows us to provide insights about potential correlations between Hexad user types and the perception of gamification elements in the physical activity domain and to compare these correlations to the healthy eating domain. Since both studies used the same methodology and because the storyboards were shown to be comparable (as we will elaborated upon in the following), we can provide further insights on the role of the application context by comparing to what extent the perception of the same gamification elements and correlations to Hexad user types changes.

Moreover, we provide a complement to the findings presented in Section 4.3, in which the role of behavior change intentions on the perception of different types of achievement goals was investigated. Instead of investigating high-level goal types, we narrow our focus down to specific gamification elements and potential differences between different stages of change in this study.

Lastly, as already stated, we contribute novel insights on the interplay between Hexad user types (representing rather static motivational orientations) and stages of change (representing more dynamic changes in the extent to which a behavior has been internalized) in the perception of gamification elements. Since both motivational orientations (see Section 2.1.2) and the extent to which a behavior has been internalized (see Section 2.1.2) have been shown to affect the functional significance of external stimuli, we expect that both these factors play a major role in the perception of gamification elements.

4.5.1 Storyboards and Gamification Elements

For the storyboards, we ensured to have at least one gamification element for each user type, based on [213, 342]. This resulted in twelve different storyboards (showing the gamification elements as stated in Table 4.5). These were created using the guidelines by Truong et al. [345]. We decided to use storyboards since they provide a common visual language that is easy to understand and do not involve game- or technology-specific knowledge [262]. Two storyboards are shown in Figure 4.6. All created storyboards can be found on figshare[27].

Storyboard Validation

To ensure that participants understand the storyboards, we conducted a qualitative pre-study in the lab. Moreover, we wanted to make sure that the storyboards for the physical activity domain are comparable to the ones used for the healthy eating domain (see Section 4.4).

Method After answering demographic questions, the printed storyboards were shown to each participant in random order. A semi-structured interview followed in which all sessions were conducted by one researcher and audio recordings were made. First, participants were asked to describe the storyboards in their own words. When necessary, the interviewer asked questions to prompt participants to identify which activities are depicted by the storyboards. Questions included: *"What is the character's goal?"* and *"What means does the character use to*

[27] figshare: *All storyboards that were used in the user study*, `https://bit.ly/3xctsXs` (last accessed: 2021-12-01)

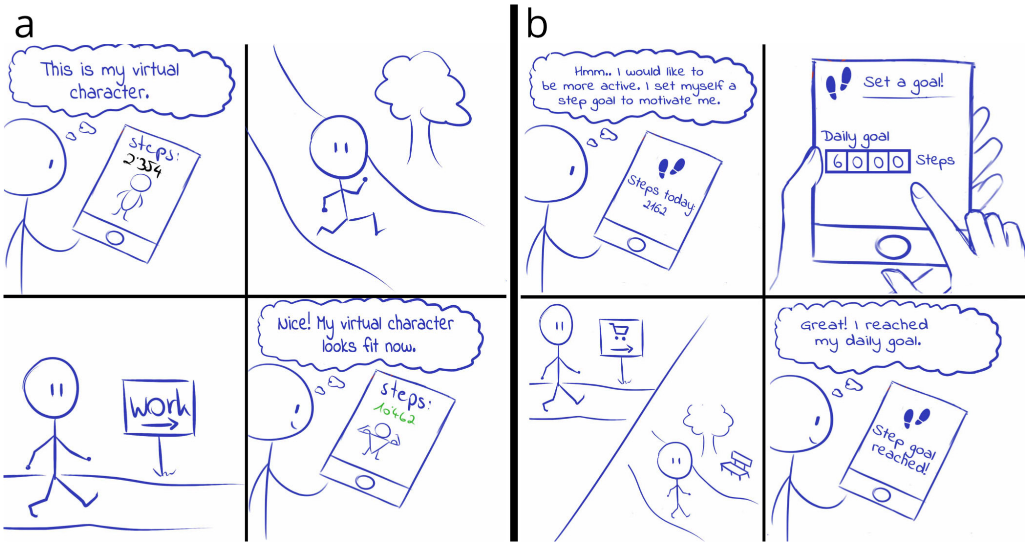


Figure 4.6: Virtual Character (a) and Custom Goal (b) storyboards

achieve her goal?". Afterwards, participants were given a short textual summary of each gamification element. They were asked to assign each of the storyboards its respective element by placing the aforementioned pieces of paper (holding the textual summaries) next to the respective storyboard. Also, they had to assign each storyboard from the physical activity context to the corresponding storyboard of the healthy eating context, to ensure the comparability of the storyboards/gamification elements. Next, interviews were transcribed and analyzed by two independent raters ("R1", "R2"). They received the transcriptions for each storyboard, without revealing which gamification element was described by the participants. Their tasks were to evaluate which element was being described and to rate how well the element was understood on a 5-point scale.

Results 8 German participants took part (4 female, average age 21.75). To ensure that the ratings can be interpreted objectively, we calculated the inter-rater agreement and found it to be Kappa=0.75, which is considered as substantial [221]. Analyzing the ratings of the two independent raters, we found that the participants understood the storyboards very well (M_{R1} = 4.90, Min_{R1} = 4; M_{R2} = 4.86, Min_{R2} = 4). This was supported by the fact that both raters successfully assigned the correct gamification element based on participants' storyboard descriptions. Regarding users assigning the textual summaries to the respective storyboard, only one assignment was incorrect. However, this wrong assignment was due to the participant misreading the descriptions of one of the gamification elements. The participant assured us that the storyboard and respective gamification element were clear. In addition, no errors were made matching corresponding storyboards from the two contexts, suggesting that comparability holds for both healthy eating and physical activity storyboards.

Gamification Element	Short Storyboard Description	Expected UT
Virtual Character	*The appearance of a virtual character is linked to the amount of steps walked.*	AC, PL
Custom Goal	*The user sets themself a custom step goal.*	AC, FS
Personalized Goal	*The system personalizes the users' step goal.*	AC
Challenge	*The user manages to reach a demanding goal.*	AC
Badges	*The user reaches their goal three times, unlocking a new badge.*	AC, PL
Points	*The system rewards the user with points for walking steps.*	PL, AC
Rewards	*After reaching the step goal three times, the user receives a coupon code.*	PL
Knowledge Sharing	*The user helps another user in a forum by answering a question.*	PH
Unlockable Content	*After reaching the step goal three times, the app unlocks a new feature .*	FS
Cheating	*The user decides to cheat by driving a car to reach their step goal.*	DI
Social Collaboration	*A group of users have to collaborate, to reach their shared step goal.*	SO
Social Competition	*A group of users are shown on a leaderboard, competing for the top position.*	SO, PL

Table 4.5: Gamification elements included in the main study, a short textual description explaining what is depicted in the corresponding storyboard and the user types ("UT") we expect to be positively affected by them based on [213, 342].

4.5.2 Evaluation

We conducted an online survey, which was available in English and German. Participants were recruited via social media and Academic Prolific (paid 1.50 pounds). The study took 10-15 minutes to complete and was approved by our Ethical Review Board[28]. After asking for demographic data and gaming behavior, the stage of change ("SoC") of the Transtheoretical Model of Behavior Change ("TTM") was determined using a validated scale for the Physical Activity

[28] Saarland University: *Ethical Review Board of the Faculty of Mathematics and Computer Science*, `https://bit.ly/2TK2Qii` (last accessed: 2021-12-01)

context [212]. For later analysis, participants were split into two groups: "Low-TTM" (participants who did not take action so far, having a SoC $<=$ 3 [366]) and "High-TTM" (participants who did take action, having a SoC $>=$ 4 [366]), according to the suggestions of Xiao et al. [366] on how to analyze the different TTM stages. Afterwards, participants' user type was determined using the Hexad User Types scale [342]. Finally, as the main part of the questionnaire, participants were shown the 12 storyboards in a randomized order. To measure the persuasiveness of each gamification element depicted in the storyboards, we adapted the perceived persuasiveness scale by Drozd et al. [111] in the same way as was done by Orji et al. [259]. The scale consists of four items to be answered on 7-point Likert scales. A Shapiro-Wilk test revealed that the persuasiveness items were not normally distributed, which is why we used non-parametric tests for our analysis. For correlation analysis, Kendall's τ was used, as it is well-suited for non-parametric data [155]. It should be noted that Kendall's τ is usually lower than Pearson's r for the same effect sizes. Therefore, we transformed interpretation thresholds for Pearson's r to Kendall's τ, according to Kendall's formula [355] (small effect: $\tau = 0.2$; medium effect: $\tau = 0.3$; large effect: $\tau = 0.5$).

Results

We excluded three participants who are unable to exercise or answered all gaming related questions with "Strongly disagree", leading to 179 valid responses. Of these participants, 44.1% were male, 55.3% were female and 0.6% identified themselves as "nonbinary". Most participants (38%) were aged 18-24 years, followed by 25-31 (34.1%), 32-38 (17.3%), 39-45 (6.7%) and younger than 18 (1.7%). The remaining participants were aged 45 and older (1.7%). Participants claimed to have a passion for video games (M = 3.70, SD = 1.11, Mdn = 4.00) and to frequently play video games (M = 3.58, SD = 1.24, Mdn = 4.00).

SoC and Gamification Elements After splitting participants into two TTM groups (as suggested in [366]), 72 participants were in the Low-TTM and 107 participants in the High-TTM group. To investigate whether the perceived persuasiveness changes between these groups, we performed a two-sided Mann-Whitney-U test for each gamification element. Also, a one-sample Wilcoxon signed rank test was performed against the value 4 on the 7-point scale to investigate which gamification elements were perceived as significantly better or worse than the neutral choice. Table 4.6 shows an overview of these tests and the means and medians of the perceived persuasiveness for each gamification element. Overall, we found that some gamification elements were perceived significantly different from the neutral choice in the High-TTM group but not in the Low-TTM group. Also, significant differences for four gamification elements were found. Badges and Challenges, both building on the need for mastery or competence [213], were shown to be significantly more persuasive for users at high stages of change than for users at low stages. This is explainable by goal-

	Low-TTM	High-TTM	Diff. sig.
Virtual Character	M = 4.05, SD = 1.77, Mdn = 4.50	M = 3.94, SD = 1.81, Mdn = 4.25	-
Custom Goal	M = 4.34, SD = 1.49, Mdn = 4.63	M = 4.70, SD = 1.55, Mdn = 5.25	-
Personalized Goal	M = 4.88, SD = 1.44, Mdn = 5.00	M = 4.93, SD = 1.38, Mdn = 5.25	.
Challenge	M = 4.32, SD = 1.65, Mdn = 4.75	M = 4.88, SD = 1.27, Mdn = 5.00	p = 0.045, Z = -2.00, U = 3173.50
Badges	M = 3.95, SD = 1.57, Mdn = 4.00	M = 4.46, SD = 1.40, Mdn = 4.75	p = 0.028, Z = -2.19, U = 3108.50
Points	M = 4.39, SD = 1.46, Mdn = 5.00	M = 4.52, SD = 1.43, Mdn = 4.50	-
Rewards	M = 5.16, SD = 1.48, Mdn = 5.25	M = 5.50, SD = 1.39, Mdn = 5.75	-
Knowledge Sharing	M = 4.06, SD = 1.52, Mdn = 4.25	M = 4.26, SD = 1.51, Mdn = 4.50	-
Unlockable Content	M = 4.70, SD = 1.49, Mdn = 5.00	M = 4.84, SD = 1.53, Mdn = 5.00	-
Cheating	M = 2.12, SD = 1.16, Mdn = 2.00	M = 2.35, SD = 1.44, Mdn = 2.00	-
Social Collaboration	M = 4.23, SD = 1.56, Mdn = 4.88	M = 4.81, SD = 1.61, Mdn = 5.25	p = 0.009, Z = -2.62, U = 2963.50
Social Competition	M = 4.09, SD = 1.74, Mdn = 4.50	M = 4.61, SD = 1.76, Mdn = 4.75	p = 0.048, Z = -1.98, U = 3180.50

Table 4.6: Persuasiveness of gamification elements in the Low- and the High-TTM group and results of Mann-Whitney-U tests comparing them ("Diff. sig."). Significant differences from the neutral choice are colored (green = positive, red = negative deviations)

setting theory (as both elements require reaching a goal), stating that goals are most effective when users are committed to them [340], which is unlikely for users in the Low-TTM group. Another reason could be that participants in Low-TTM considered themselves not to be able to reach those goals [126]. Moreover, Social Competition and Social Collaboration, both building on the relatedness motive [213] were perceived as significantly more persuasive in the High-TTM group. A potential reason for this includes the fear to not be able to keep up with other users [126], detrimentally affecting users' motivation. These findings show that the SoC on its own is a relevant factor that should be considered in tailoring persuasive, gamified interventions in the physical activity context.

	AC	DI	FS	PH	PL	SO
Virtual Character	-	-	-	-	**.237****	.114*
Custom Goal	**.205****	-	**.132***	.119*	-	.106*
Personalized Goal	**.211****	-	-	.145**	-	-
Challenge	**.200****	-	.145**	-	.177**	-
Badges	**.122***	-	-	-	**.223****	-
Points	**.201****	-	.110*	.192**	**.169****	.105*
Rewards	.114*	-	-	.152**	**.250****	.109*
Knowledge Sharing	.123*	-	-	**.234****	-	.175**
Unlockable Content	.140**	-	**.143****	-	.163**	-
Cheating	-	**.157****	-	-	-	-
Social Collaboration	.147**	-	.153**	.145**	.216**	**.314****
Social Competition	.105*	-	-	-	**.370****	**.204****

Table 4.7: Kendall's τ and significance between the Hexad user types and the gamification elements. Bold entries represent expected correlations (Table 4.5). *p$<$.05, **p$<$.01

Hexad User Types and Gamification Elements Table 4.7 presents the significant correlations of gamification elements to each user type. We found 16 positive correlations between user types and gamification elements out of 17 expected correlations (see Table 4.5). The positive correlation between the gamification element "Virtual Character" and the "Achiever" user type is the only correlation that was expected but not found. Given these results, we extend and replicate previous work [261,342]: We show the applicability of previous findings in the Physical Activity context and contribute evidence for previously hypothesized, but not yet shown correlations, i.e. between the Philanthropist and the gamification element "Knowledge Sharing" and between the Disruptor and the gamification element "Cheating" [342]. In addition to expected correlations, some unexpected correlations were found. However, this is in line with previous research about the Hexad user types [261, 342]. Also, all but one unexpected correlations are weak ($\tau < 0.2$), which suggests that their actual effect is negligible.

SoC, Hexad User Types and Gamification Elements To investigate potential effects of the SoC on the set of suitable gamification elements for each user type, we compared correlations of gamification elements to user types between the Low- and the High-TTM group. Table 4.8 shows these correlations for both groups. The analysis revealed that the set of significantly correlating gamification elements is different in both groups, suggesting that taking the SoC into account when tailoring persuasive systems for user types should improve personalization. To emphasize this, we also investigated whether the strength of correlations differs significantly between the Low- and the High-TTM groups. For this, we

	Low-TTM							*High-TTM*					
	AC	**DI**	**FS**	**PH**	**PL**	**SO**	-	**AC**	**DI**	**FS**	**PH**	**PL**	**SO**
Vir. Cha.	.218*	-	-	-	-	-		-	-	-	-	.304**	.183**
Cus. Goal	.192*	-	-	-	.171*	-		.215**	-	.178*	.194**	-	-
Per. Goal	-	-	-	-	-	-		.253**	-	.178*	-	-	-
Challenge	.182*	-	-	-	-	-		.214**	-	-	-	.249**	-
Badges	-	-	-	-	.215*	-		.161*	-	-	.141*	.276**	-
Points	-	-	-	.213*	-	.191*		.250**	-	.200**	.170*	.195**	-
Rewards	-	-	-	-	.182*	-		-	-	-	.144*	.303**	-
Kno. Sha.	-	-	-	-	-	-		.191**	-	-	.327**	-	.248**
Unl. Con.	-	-	.222*	-	-	-		.154*	-	-	-	.230**	-
Cheating	-	.222*	-	-	-	-		-	-	-	-	-	-
Soc. Col.	-	-	.191*	-	-	-		.153*	-	-	.185**	.285**	.343**
Soc. Com.	-	-	-	-	.316*	-		-	-	-	-	.422**	.206**

Table 4.8: Kendall's τ and significance between the Hexad user types and gamification elements for the Low- and the High-TTM group. Colored cells indicate that a correlation is significantly stronger in one group than in the other group. *p<.05, **p<.01

converted Kendall's τ to Pearson's r according to Kendall's formula described in [355]. Afterwards, we applied Fisher's z-transformation to these coefficients to check for effects. Here, we found multiple significant differences between the groups for all user types but the Disruptor. Gamification elements for which the correlation coefficient significantly increased on a user type level are colored green in Table 4.8. For example, we found that the correlation between the "Virtual Character" gamification element and the "Achiever" user type is significantly stronger in the Low-TTM than in the High-TTM group. Similarly, we found that social competition is positively affecting for Socializers only when being in a High-TTM stage. Besides the Disruptor, we found similar findings for all other user types. Therefore, these results should be considered when making decisions about which gamification elements should be included in a system, in order to enhance its persuasiveness.

Differences from the Healthy Eating Context

Since the storyboards that we used in this study and in the study presented in Section 4.4 were shown to be comparable and the methodology was the same in both studies, we also analyzed differences in the general perception of gamification elements and regarding correlations to Hexad user types. When comparing the significant correlations between Hexad user types and the perception of gamification elements in both contexts, it can be seen that we mostly found consistent results, supporting the validity of the Hexad types across different contexts. However, we also found differences in the set of significant correlations.

For example, while virtual characters were significantly correlated to the achiever user type in the healthy eating domain (as expected), this expected correlation could not be found in the physical activity domain. In contrast, the expected correlation between badges and achievers was found in the physical activity domain, but not in healthy eating. Regarding the perception of gamification elements independent of the Hexad user type, we compared the perceived persuasiveness of gamification elements by conducting Mann-Whitney U tests across the two domains. We found that the perceived persuasiveness of social collaboration ($U=18990.50$, $p=.067$, $d=.10$) and social competition ($U=18989.50$, $p=.067$, $d=.10$) was descriptively higher in the physical activity domain. In contrast, the gamification element cheating was perceived as significantly more persuasive in the healthy eating domain ($U=14691.00$, $p<.001$, $d=.31$). These findings suggest that the context itself plays a role in how persuasive certain gamification elements are perceived to be.

4.5.3 Discussion and Limitations

We investigated the effect of behavior change intentions on the perception of gamification elements in the physical activity domain. We contribute three main findings: First, we presented results about the individual impact of the SoC on the perception of each gamification element, leading to a set of well- and poorly perceived elements for each TTM group. We found that there are differences in this set, as many gamification elements are not perceived similarly across groups, showing that the SoC impacts their perception. This is supported by finding multiple significant differences between both groups, showing that considering the SoC for tailoring gamified, persuasive systems in the physical activity domain is important. Second, confirming previous findings [261, 342], we found 16 out of 17 expected correlations between gamification elements and Hexad user types. Besides validating previous findings in the physical activity context, we contribute a set of new correlations, which were expected in previous works [213, 342], but have not been shown. This might be due to using storyboards rather than textual descriptions as in [342] and because of using a concrete context rather than a general context, also as in [342], potentially leading to a more concrete idea of how the elements work. Additionally, we examined the "persuasiveness" of gamification elements, whereas past work by Tondello et al. [342] investigated "enjoyment".

Third, by analyzing the effect of the SoC on the set of relevant gamification elements for each user type, we show that even though the user type itself may remain stable [342], the set of relevant gamification elements does not. This is important, as so far a static set of elements has been suggested for each user type [342], not taking into account the dynamic process of behavior change intentions [276]. However, our work has several limitations that should be considered. First, we used storyboards to assess the persuasiveness of each gamification element. Therefore, validating our findings using real implementations is an important next step. Second, even though we investigated atomic gamification elements, some aspects of the realization of these gamification elements are inherently a matter of interpretation, affecting the external validity of our results when implementing gamification elements differently. Third, it should be noted that combining gamification elements may create different experiences for the user, which should be analyzed in future work. Fourth, our participants reported to have experience in games, which should be considered. Last, we cannot say whether our findings generalize to different contexts besides physical activity. Therefore, further research should be conducted about the SoC as a factor for personalization in different contexts.

When comparing the perception of gamification elements between the healthy eating and physical activity domains, we found that the correlations between Hexad user types and gamification elements were largely the same. However, some differences could be observed. While this could have been the result of different samples and sample sizes, it could also be attributed to differences in the contexts. The latter is supported by the fact that the general perception of gamification elements differed across both contexts: Social competition and social collaboration were descriptively perceived as more persuasive in the physical activity domain (almost reaching significance) while cheating was perceived as significantly more persuasive in the physical activity domain. These differences could potentially be explained by context-inherent factors. The physical activity context focuses on increasing the performance of users and thus might inherently afford the satisfaction of competence needs and demand for performance goals, as established by the achievement-goal model (see Section 2.2.2). In contrast, the healthy eating context does not focus on increasing performance but rather to change the eating habit of users, mainly by controlling their calorie intake. Thus, the autonomy of users might be undermined by restricting the amount and type of food users should eat. This might explain why cheating, a highly autonomous gamification element which allows for a certain amount of rebellion, is more important to users in this domain.

4.5.4 Contribution to Research Questions

In this study, we investigated the impact of Hexad user types and behavior change intentions on the perceived persuasiveness of gamification elements. We first had a look at each factor separately. Here, we found that the Hexad user type is significantly correlated with several gamification elements. Most of these correlations were expected based on previous work in a general context. Thus, in line with the study in the domain of healthy eating presented in Section 4.4, our findings show that Hexad user types play an important role in explaining user preferences in gamified systems. In addition, complementing the findings from the study presented in Section 4.3, we found evidence that the perception of gamification elements differs significantly between users who have more internalized motivations to change their behavior, as compared to users whose motivation is more extrinsically regulated, or who are amotivated. Lastly, we investigated whether the intention to change behavior has an impact on the strength of correlations between Hexad user types and the perception of gamification elements. Thus, we investigated whether combining both these factors changes which gamification elements should be considered for a user. Indeed, we found that the strength of correlations between Hexad user types and the perception of gamification elements differs significantly between users having low behavior change intentions, compared to those having high behavior change intentions. These findings suggest that both factors should be considered to tailor gamified systems to the user.

Overall, these findings add insights to **RQ2** since they provide answers to the question of how personal factors may change the perception of gamification elements. Although the set of significant correlations between Hexad user types and the perception of gamification elements was largely the same between the healthy eating and physical activity contexts, we found indications that the context might affect how gamification elements are perceived. First, some expected correlations between Hexad user types and gamification elements were found in one context but not in the other. Second, certain gamification elements were perceived as more persuasive in one context than in the other. As already discussed, it seems that the physical activity context focuses on increasing performance, which might demand gamification elements supporting the need for competence, while healthy eating might thwart autonomy needs to a certain extent and thus might demand gamification elements supporting the need for autonomy. These findings support the importance of contextual factors and contribute to **RQ1**.

4.6 Summary

In this chapter, we were mainly interested in understanding to what extent personal factors play a role in explaining differences in the perception of gamification elements (**RQ2**). We contributed to this question by investigating three different personal factors. First, we considered age, since the importance of basic psychological needs, their impact on what motivates people and the life goals people have change in older age [283, 301, 302]. More specifically, it was shown that the importance of performance and achieving new goals – and thus potentially the importance of satisfying *competence* needs – declines with increasing age [283]. In contrast, the relevance of *relatedness* needs has been shown to increase in old age [301, 302]. We found that this shift seems to be reflected in the motivations of older adults to engage in games and gameful systems, since social interaction and care taking – and thus *social relatedness* – were important drivers for older adults. Also, performance motives were less salient and winning was found to play a less important role for them. This was also reflected in the perception of gamification elements. Frequently used elements such as points, badges and competition were poorly perceived, while social collaboration and virtual characters, fostering reciprocal care, were particularly positively perceived. By comparing the Hexad user type distributions of younger and older adults, we were able to provide a third perspective (alongside the findings of the semi-structured interview and the rating of gamification elements) to these findings. We found that a remarkable number of participants scored highest on the Philanthropist trait in the older sample, while previously reported score distributions and our own reference sample consisting of younger people were much more heterogeneous. Since the Philanthropist is a socially-minded and altruistic user type, further characterized by care-taking, these results align well to the aforementioned findings. Overall, our findings provide a consistent picture of what one should consider when designing gamified systems for older adults, which can be readily explained by previous research in the domain of motivation psychology.

Second, we investigated behavior change intentions and their impact on how gamification elements are perceived. We found that the perception of different types of achievement goals, which can be seen as inherently implemented by most gamification elements, differs depending on the stage of change of users. While self-approach goals were well perceived across all stages, task-approach and other-approach goals were positively correlated with the stage of change of participants. More specifically, our findings showed that perceived *competence* might play an important role in how persuasive task-approach goals are. If the goal seems unreachable, people may be less motivated or even amotivated to try to reach it. Considering that people in low stages of change are likely affected by a lack of perceived competence, a static goal imposed by the system (a task-approach goal) might seem hard to reach for them. Similarly, if perceived competence is low, other-approach goals could seem hard to reach. This might undermine intrinsic motivation and could explain these findings. In contrast, self-approach goals define competence based on one's own past performance and thus are less likely to lead to feelings of perceived incompetence. These results and the importance of perceived competence, which might change when moving through the stages of change and internalizing certain behaviors, can be transferred to gamification elements. We found that social gamification elements (competition, collaboration) and gamification elements establishing demanding goals (challenges, badges) are perceived as significantly more persuasive in higher stages of change. Overall, these findings underline the importance of considering the extent to which a target behavior has been internalized to tailor gamified systems to the user.

Third, we investigated Hexad user types as mediating factors regarding the perception of gamification elements in two contexts: healthy eating and physical activity. In both these contexts, we found significant correlations between Hexad user types and the perception of gamification elements, which were expected based on previous research and were mostly consistent across the contexts. When looking at these correlations, it becomes obvious that those gamification elements which support the basic psychological needs and types of motivations characterizing a certain Hexad user type are particularly relevant for that user type.

As a consequence, we can conclude that considering a user's scores on each of the Hexad user types, which may reflect motivational orientations and the importance of certain needs, is important when personalizing gamified systems.

Fourth, we have seen that combining behavior change intentions and Hexad user types could be a promising research direction, since we found that the strength of the correlation between Hexad user types and gamification elements differs significantly between users in lower stages of change and those in higher stages of change.

Lastly, the fact that we used the same methodology to assess the perception of the same gamification elements across two different contexts allowed us to compare them. We found that the perception of certain gamification elements differs between contexts, and we were able to associate these differences with context-inherent need satisfactions and frustrations, which further supports the need to consider the application context when tailoring gamified systems (**RQ1**).

To sum up, we demonstrated that age, behavior change intentions, and Hexad user types, as well as the combination of the latter two, play a role in explaining interpersonal differences and thus are important factors to consider when tailoring gamified systems. We further showed that all our findings can be explained through the lens of SDT and that the concept of basic psychological needs (see Section 2.1.2) can be used to explain the differences in how gamification elements are perceived. All considered factors have been associated with changes in what motivates people and our results show that these changes are indeed reflected in the preferences people have for or against certain gamification elements. What is still missing at this point is empirical evidence for the actual effectiveness of these factors, i.e. whether the changes in the *perception* of gamification elements, which we have found in this chapter, actually can be used to increase motivation and improve behavioral outcomes in implemented gamified systems, when these systems are tailored based on our findings. We will contribute to answering this question in the next chapter.

Chapter 5
Effects of Personalized Gamified Systems on Motivation and Behavior

In the previous chapter, we showed that the *perception* of gamification elements differs between users and that personal factors – age, behavior change intentions and Hexad user types – explain these differences to a certain degree. Therefore, the factors could be used to tailor gamified systems to the user. However, whether taking these factors into account to tailor gamified systems to the user actually leads to measurable effects on motivation and behavior is unclear because so far, we have relied on storyboards to explain gamification elements. While this allows us to measure how participants perceive certain gamification elements, it does not allow users to experience the gamification elements in an implemented system. We will contribute to this open question in this chapter by implementing gamified systems and investigating whether gamified systems that are tailored to the user, using the findings from Chapter 4 and past research (see Section 2.4.3), lead to beneficial effects on motivation and behavior compared to non-tailored systems.

We will approach this question by implementing three different systems, each of which we used for one respective study. First, in Section 5.2 we will present findings from a gamified system called "Endless Universe", in which we offer a static set of gamification elements – which should be particularly suitable for people who have certain Hexad user types or who are in high stages of change – to motivate users to run on a treadmill. Thus, we contribute insights on the performance and motivation impacts of using behavior change intentions and Hexad user types to tailor a gamified system. In Section 5.3, we investigate the approach of dynamically (counter-) tailoring the set of gamification elements offered by a gamified system to users' Hexad types in an image-tagging context.

We measure whether activating gamification elements that should be particularly suitable leads to a higher motivation, performance, an increased prevalence of flow experiences and stronger positive affect, as compared to offering gamification elements which are not tailored. In this study, we complemented validated surveys with psychophysiological measures to obtain a more holistic picture about the effects on the user experience. Lastly, in Section 5.4, we integrated gamification elements that should be particularly suitable for certain Hexad user types into a gym's course booking system. Since the first two studies were short-term laboratory studies which did not allow us to investigate whether the positive effects of personalization on behavioral and psychological measures persist in the long run, we conducted an in-the-wild study over almost two years with participants visiting the gym.

Although we also contribute to **RQ1** by investigating the effects of gamification on motivation and behavior, we contribute mainly to **RQ3**, i.e. the question of what effects the personalization of gamified systems has on behavioral and psychological outcomes. Section 5.2 is based on [11], Section 5.3 on [21] and Section 5.4 on [17].

5.1 Motivation

As we have seen in the previous chapter, people differ in how they perceive gamification elements. Therefore, using a "one-size-fits-all" approach, i.e. providing users a static set of gamification elements, may not lead to optimal outcomes. This general conclusion can also be drawn based on previous research, which revealed inconclusive or even negative outcomes of using such a static approach [7, 146, 315]. In line with our findings, it was found that there are inter-personal differences in the *perception* of gamification elements [182], which could threaten static gamification approaches. As a consequence, past research focused on finding and investigating factors that moderate the perception of gamification elements. As we have seen in Section 2.4.3, demographic factors such as gender [258] or age [43], as well as personality traits [168] and user or player typologies [182] were investigated. These studies revealed correlations between the *perception* of gamification elements and the respective factors.

However, previous research investigating correlations between preferences for gamification elements and certain personal factors relied on non-interactive materials to explain gamification elements [182]. Consequently, participants in these studies had no chance to interact with actual applications, but instead rated their *perception* based on e.g. textual descriptions or storyboards. Thus, the actual *effects* on behavioral or psychological outcomes, when users interact with gameful applications that have been personalized based on these factors, remained unclear. This limitation also applies to our findings presented in Chapter 4, since we relied on storyboards to explain gamification elements. While this approach has several advantages (such as allowing a larger number of participants to be recruited more easily and abstracting from implementation-specific details), the actual effects on behavior and motivation cannot be studied.

We bridge this gap in this chapter by investigating whether personalizing gamified systems based on our findings from Chapter 4 improves behavioral and psychological measures. To do so, we focus on Hexad user types and behavior change intentions, and investigate whether the moderating role of these factors in explaining user preferences (as found in the previous chapter) translates to actual differences in implemented gamified systems.

5.2 Behavior Change Intentions and User Types in a Gamified Fitness System

The first section of this chapter continues our efforts to better understand the role of behavior change intentions and Hexad user types in explaining why the outcome of gamification differs across users. Similarly as in Section 4.3 and Section 4.5, we focus on the domain of physical activity. As already stated, the studies presented in the previous chapter, as well as studies from related work done in the past, investigated factors for personalization by using a survey-based methodology. This means that participants did not have the chance to interact with applications, but instead rated their perceptions by imagining how the gameful design elements would look in a real system. In contrast, we investigate the actual effect of tailoring gamification to such personal factors by allowing participants to interact with a gamified system.

More specifically, we contribute knowledge about the actual effectiveness of personalization based on Hexad user types and behavior change intentions by applying our findings from Chapter 4 in the context of "Endless Universe", a gameful application encouraging physical activity on a treadmill. We conducted a lab experiment in which participants were asked to run on a treadmill and thereby interacted with Endless Universe. The gamification elements used in the system were intended to be particularly suitable for certain Hexad user types and users in high stages of change, according to our findings regarding the impact of behavior change intentions and Hexad user types on the perception of gamification elements. In the lab experiment, we investigated whether users who are in high stages of change and score particularly high on certain Hexad user types increase their performance more than other users, whether they are more intrinsically motivated and whether they show stronger emotional responses. Besides investigating the effectiveness of behavior change intentions and Hexad user types as factors for personalization, we also investigate whether Endless Universe affects the aforementioned dependent variables without personalization, i.e. when not considering any of the aforementioned two factors, in order to replicate past research in the field of one-size-fits-all gamification.

5.2.1 Concept and System Design

The *actual effects* of personalizing gameful applications based on behavior change intentions and Hexad user types on task performance and user experience cannot be investigated without allowing users to interact and experience the gameful design elements in a real system. Therefore, we implemented Endless Universe, a gameful application that builds upon the results of the online study to investigate the effects of personalization on these aspects.

Endless Universe ties the distance covered on a treadmill to the progress within several gameful design elements. To investigate which effects personalization has on measures related to the users' performance and experience, we decided to use the findings from the storyboards-based online study presented in Section 4.5 to tailor Endless Universe to a specific user group.

Theme

We decided to use outer space as the main theme of the gameful application. This decision is based on previous research using gameful applications encouraging physical activity, which demonstrated that this theme is well-perceived within the physical activity context [68, 98, 108, 109, 124, 304]. The core mechanic in the gameful application is a spaceship exploring an endless universe. Hereby, the real-time distance covered by the user on the treadmill has a direct influence on the speed of the spaceship moving forward in the space exploration. The spaceship is shown prominently in the middle of the screen and a moving illusion is created by animating the background of the scene (i.e. stars and particles are moving faster or slower). The distance covered by the user is shown permanently in the application. When starting the application for the first time, an introduction is given to the users, explaining that they belong to an alien species which is competing to explore the universe with their spaceships. Figure 5.1 shows a screenshot of the application.

Goal Setting

Endless Universe establishes a target distance to cover, which is shown next to the distance covered in the main screen of the application. This target distance is personalized to the user, i.e. is based on a users' fitness level. This was done to make sure that the target distance is reachable to all users and thus comparable. This is in line with previous research within this context [204, 230]. More specifically, this target distance was 10% higher than the previously covered distance. The gameful design elements, which are described next, operate on this target distance.

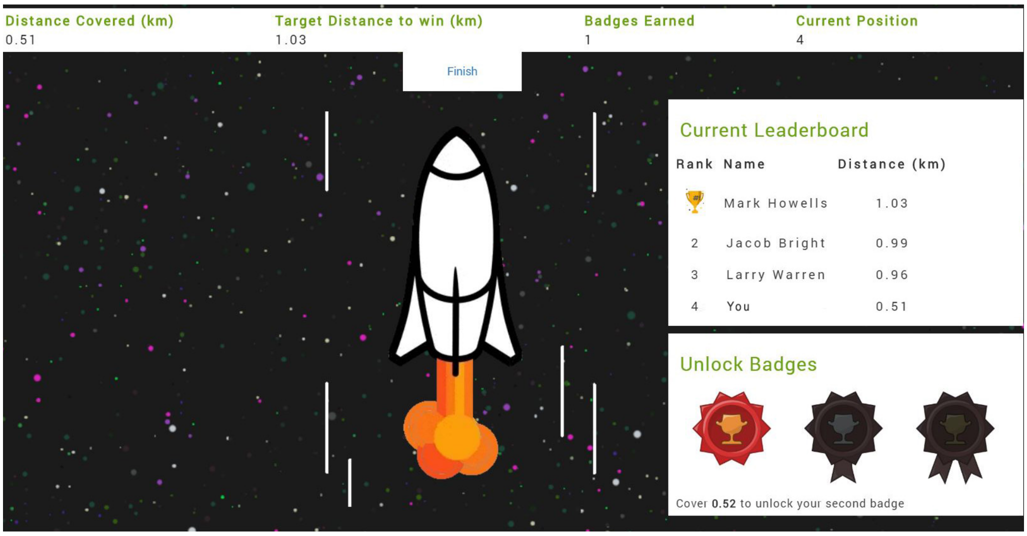


Figure 5.1: Screenshot of the Endless Universe application. The distance covered, target distance, number of badges unlocked and current position on the leaderboard are shown on the top. The leaderboard and badges are shown on the right side of the screen.

Gamification Elements

The findings of the storyboards-based online study presented in Section 4.5 show that behavior change intentions and Hexad user types are relevant factors for personalizing gameful applications encouraging physical activity. Based on these findings, we derived a set of gameful design elements to investigate the effects of personalization. As such, we decided to use the gameful design elements *Badges*, *Challenges* and *Social Competition*. These gameful design elements were shown to be perceived as significantly more persuasive among participants in higher stages of change ("High-SoC", see Section 4.5) than among participants in lower stages of change ("Low-SoC", see Section 4.5) as can be seen in Table 4.6. Also, these gameful design elements were shown to be positively correlated with the Achiever, Player and Socializer Hexad user types (see Table 4.7 in Section 4.5). Therefore, we expected that Endless Universe should be suitable for users belonging to the High-SoC group or scoring particularly high on Achiever, Socializer or Player. The gameful design elements are described in the following:

Badges: There are three different badges in the gameful application. To account for inter-personal performance differences, the thresholds to unlock badges was established relatively to the target distance. The first badge is unlocked when reaching 20% of the target distance and is visualized through a bronze trophy. The second badge, a silver trophy, is unlocked when reaching 50% of the target distance. Finally, the golden badge is unlocked when reaching 100% of the target distance. This progression concept follows the recommendations related to progression stairs in games by Werbach and Hunter [360]. The badges were shown on the right side of the screen and darkened until they were unlocked. The remaining distance until unlocking the next badge was shown permanently below the badges. Based on the findings from Section 4.5, this gameful design element should be perceived particularly well by users belonging to the *High-SoC* group and users scoring high on the *Achiever* or *Player* factors of the Hexad.

Challenges: The ultimate challenge of Endless Universe is to reach the target distance. This is explained to the user as part of the onboarding procedure before starting the gameful application. When reaching the target distance and thus mastering the main challenge of the application, a so-called "explorer of the day" trophy is unlocked and shown to the user. Based on Section 4.5, this gameful design element should be perceived particularly well by users belonging to the *High-SoC* group and users having a high *Achiever* score.

Social Competition: We used a leaderboard to introduce social competition to the gameful application, positioned on the right-hand side of the screen. In this leaderboard, fictitious users were shown, similar to previous gamification studies [224]. This was done to ensure the comparability across participants, i.e. that all participants had the same chance to rise in ranks, and to avoid introducing a confounding variable [354]. Similar to *Badges*, there were three other fictitious users who covered distances that were calculated in relation to the target distance described above. The fictitious user on the first rank covered the target distance, the fictitious user on the second rank covered 5% less than the target distance and the fictitious user on the third rank covered 8% less than the target distance. This follows the same progression scheme as was used for *Badges* and thus follows recommendations established by Werbach and Hunter [360]. Based on the findings from Section 4.5, this gameful design element should be perceived particularly well by users belonging to the *High-SoC* group and users scoring high on the *Achiever*, *Player* or **Socializer** factors of the Hexad.

Implementation

The user interface part of Endless Universe was implemented as a web application and capturing the distance covered on the treadmill was realized by using an Arduino Uno board and a QRE1113 infrared reflectance sensor comprised of an infrared emitting LED and an infrared sensitive phototransistor. The hardware and user interface are explained in the following.

Hardware to Capture the Covered Distance on the Treadmill Since the covered distance is a direct input to the gameful application, we implemented a system to track the distance covered on the treadmill. We placed reflecting light tape on the belt of the treadmill in equal, pre-defined distances and used an infrared reflectance sensor to detect the tape. We used an Arduino Uno, which was connected to a PC via USB to send an event to the main application running on the PC whenever a tape was detected.

User Interface The number of events that were triggered when the reflecting tape on the belt was detected by the Arduino was sent via USB to a NodeJS Express webserver running on a PC in real-time. The webserver calculated the distance covered based on the number of detections, i.e. the total distance could be derived with a maximum discrepancy of 3.1 meters (the tape was placed every 3.1 meters). Besides calculating the covered distance, the webserver is responsible for the game logic, i.e. deriving the current rank of the user on the leaderboard, checking whether a badge should be unlocked and whether the main challenge was completed. This information is populated to the frontend using bi-directional websockets. The frontend itself was realized using HTML, CSS and Javascript. Three.js was used for the visualization of the space, the rocket and to create the moving illusion with various speed. Moreover, Bootstrap was used to make sure that the application adapts to various screen sizes and jQuery was used to manipulate the DOM of the web application whenever updated data from the webserver has been sent.

5.2.2 Evaluation

To investigate whether the findings from the online study presented in Section 4.5, which were based on the perception of storyboards, lead to effects on a user's performance or experience when actually interacting with a gameful application, we conducted a lab study. In this lab study, participants were running on a treadmill and thus interacted with Endless Universe. In the following, the procedure, method and the results of this study are presented.

Procedure and Method

The study followed a within-subjects design with two conditions. When recruiting participants, we used the same validated questionnaire as in the online study to assess the SoC within the context of Physical Activity [212], to make sure that an equal number of Low- and High-SoC participants was recruited. In the baseline condition, participants were running on a treadmill without getting any kind of feedback (the display of the treadmill was covered using black foil). In the intervention condition, Endless Universe was deployed on a 10 inch tablet device, which was placed where the display of the treadmill is located, to ensure that participants can easily see the gameful application. The study started with the baseline phase to avoid detrimental effects when removing gameful design elements [146] and to establish the target distance in the intervention phase (to make sure that the target distance is reachable to all users [204,230]). After giving informed consent, participants were asked to fill out a survey. In this survey, demographic data was collected. Next, the Hexad user type was assessed using the validated questionnaire by Tondello et al. [337], followed by a validated questionnaire to assess the SoC within the context of Physical Activity [212].

After completing this survey, participants were asked to run on the treadmill for

10 minutes in a speed that they felt comfortable with. They were told to stop running when feeling uncomfortable. Drinks were provided.

After running for 10 minutes, participants were asked to complete a second survey. In this survey, the validated version of the Positive and Negative Affect Schedule ("PANAS") [358] was administered in order to assess affective experiences while running. Next, participants were asked to fill out the 22-item task evaluation questionnaire of the Intrinsic Motivation Inventory ("IMI") [216, 297] to assess intrinsic motivation and enjoyment of the running activity. Finally, Borg's Rating of Perceived Exertion ("RPE") [50] was administered to assess how exhausting participants perceived the activity. In this scale, users choose a number between 6 ("no exertion") and 20 ("maximum exertion possible") to describe their perceived exertion. Finally, a date for the intervention phase was scheduled. We made sure that there is a break of at least one full day between the baseline and intervention phase.

The intervention phase followed exactly the same procedure. The only difference was that Endless Universe was in place while running. The task was exactly the same, i.e. participants were asked to run on the treadmill for 10 minutes in a speed that they felt comfortable with. The target distance was established based on the covered distance in the baseline phase, as described in Section 5.2.1. After running for 10 minutes, the same questionnaires as in the baseline (PANAS, IMI, RPE) were administered.

Participants were compensated by a 10 Euro amazon gift card. The study has been reviewed and received ethics clearance through an institutional Research Ethics Committee[29].

Hypotheses

Based on the findings of the storyboards-based pre-study and previous research, we expected to find evidence for the following hypotheses:

H1: One-size-fits-all Gamification affects performance and experience
H1a: The covered distance is higher when using Endless Universe
H1b: Users perceive running as more enjoyable using Endless Universe
H1c: Users have stronger affective experiences with Endless Universe

H2: SoC affects performance and experience with Endless Universe
H2a: The improvement in distance is higher for High-SoC users
H2b: High-SoC users perceive Endless Universe as more enjoyable
H2c: High-SoC users have stronger affective experiences

H3: Hexad types affect performance and experience with Endless Universe

[29] Saarland University: *Ethical Review Board of the Faculty of Mathematics and Computer Science*, `https://bit.ly/2TK2Qii` (last accessed: 2021-12-01)

H3a: The improvement in distance is higher for AC, PL, SO
H3b: AC, PL, SO perceive Endless Universe as more enjoyable
H3c: AC, PL, SO have stronger affective experiences

H1 is motivated by previous work showing that gameful applications can increase physical activity and can have positive effects on the user experience when doing sports [7, 185]. Consequently, **H1** can be seen as a replication of previous work and is important to demonstrate the overall effectiveness and validity of Endless Universe. **H1** is analyzed by conducting paired samples t-tests or Wilcoxon signed rank tests (when the assumptions of the t-test were not met). **H2** stems from findings of the storyboards-based online study presented in Section 4.5. In this study, we found that the perceived persuasiveness of Social Competition, Badges and Challenges is significantly higher among High-SoC users. Since we are using these gameful design elements in Endless Universe, we expect that the increased perceived persuasiveness should be reflected in an increased actual performance and experience. **H2** is analyzed by splitting participants in Low- and High-SoC groups and conducting independent samples t-tests or Mann-Whitney U tests (when assumptions of the t-test were not met). Similarly, **H3** bases on our findings from the online study (Section 4.5), which revealed significant correlations between the Socializer, Achiever, Player and the aforementioned gameful design elements. Also, previous research has demonstrated similar correlations for these gameful design elements in different contexts [144, 189, 261, 342]. To analyze **H3**, we calculated bivariate correlation coefficients. Similar to the online study, we used Kendall's τ, since it is well-suited for non-parametric data [155]. Also, research has recommended using Kendall's τ when the sample size is rather low [44]. Since we established one-directional hypotheses beforehand (**H3a, H3b, H3c**) and to further increase the power of the correlation analysis, we used one-sided tests. Again, when interpreting the correlation coefficients, it should be considered that Kendall's τ is lower than Pearson's r for the same effect sizes.

Results

We recruited 20 participants. 11 self-reported their gender as male and 9 as female. Most participants (50%) were aged 25-31 years, followed by 18-24 (45%) and 32-38 (5%). The number of participants across the Low- and High-SoC groups was equal (10 participants in each group). Achievers (M=24.80, SD=2.35) and Philanthropists (M=24.10, SD=3.35) showed the highest and second-highest average scores, followed by Players (M=24.10, SD=2.92) and Free-Spirits (M=23.40, SD=3.12). Socializers (M=22.90, SD = 3.80) and Disruptors (M=17.40, SD=3.95) followed with lower average scores.

Effects of "One-Size-Fits-All" Gamification First, we investigated whether Endless Universe has an effect on the performance of users, i.e. whether it moti-

	Baseline N=20	**Intervention** N=20	**Diff. sig.**
Distance Covered [km]	M = 0.96, SD = 0.32, Mdn = 0.97	M = 1.13, SD = 0.36, Mdn = 1.12	p = 0.003 Z = 24.00
RPE [scale from 6-20]	M = 9.35, SD = 2.11, Mdn = 9.00	M = 11.10, SD = 2.95, Mdn = 11.50	p = 0.027 t = -2.40
IMI Enjoyment [scale from 1-7]	M = 4.88, SD = 1.97, Mdn = 5.50	M = 5.43, SD = 1.42, Mdn = 5.67	-
IMI Competence [scale from 1-7]	M = 4.42, SD = 1.89, Mdn = 4.84	M = 5.38, SD = 1.35, Mdn = 5.84	p = 0.008 t = -2.97
IMI Pressure [scale from 1-7]	M = 1.73, SD = 1.04, Mdn = 1.33	M = 6.07, SD = 1.09, Mdn = 6.50	p<0.001 t = -10.40
IMI Choice [scale from 1-7]	M = 6.03, SD = 1.27, Mdn = 6.67	M = 2.40, SD = 1.56, Mdn = 2.00	p<0.001 t = 7.42
PANAS pos. [scale from 1-5]	M = 3.03, SD = 0.56, Mdn = 2.90	M = 3.40, SD = 0.83, Mdn = 3.45	-
PANAS neg. [scale from 1-5]	M = 2.91, SD = 0.14, Mdn = 3.00	M = 2.90, SD = 0.0.17, Mdn = 2.90	-

Table 5.1: Dependent variables of the lab study for the baseline and intervention condition and results of paired samples t-tests / Wilcoxon signed rank tests ("Diff. sig.") comparing them.

vated participants to cover more distance than in the baseline. This is important to replicate previous research, which showed the effectiveness of one-size-fits-all gamification in this domain [14, 79, 185].

Table 5.1 shows the means, standard deviations, medians and significant differences for all dependent variables of the study for the baseline and intervention phase. We found a significant difference in the covered distance between the baseline and intervention condition (Z=24.00, p=0.003). Based on this, we establish result **R1: Participants covered a significantly higher distance when using Endless Universe**. Next, we analyzed whether RPE differs across the conditions. Again, we found a significant difference between the intervention and baseline phase in perceived exertion (t=-2.40, p=0.027). Thus, **R2: Perceived Exertion is higher when using Endless Universe** confirms that the increased distance (**R1**) is also reflected in the subjectively higher feeling of exertion. Regarding enjoyment and user experience, we compared the factors of the IMI and PANAS. Here, we found a significant difference for the competence (t=-2.97, p=0.008), pressure (t=-10.40, p<.001), and choice (t=7.42, p<.001) factors of the IMI. No significant effects were found for the enjoyment factor (p=0.20). Thus we establish **R3: Perceived competence and pressure is higher when using Endless Universe**

	Low-SoC N=10	**High-SoC** N=10	**Diff. sig.**
Distance Improvement [intervention/baseline]	M = 1.37, SD = 0.31, Mdn = 1.23	M = 1.11, SD = 0.31, Mdn = 1.15	-
RPE [scale from 6-20]	M = 11.90, SD = 2.89, Mdn = 12.00	M = 10.30, SD = 2.95, Mdn = 11.00	-
IMI Enjoyment [scale from 1-7]	M = 5.13, SD = 1.77, Mdn = 5.33	M = 5.73, SD = 0.97, Mdn = 5.83	-
IMI Competence [scale from 1-7]	M = 5.07, SD = 1.62, Mdn = 5.17	M = 5.70, SD = 1.01, Mdn = 6.00	-
IMI Pressure [scale from 1-7]	M = 5.80, SD = 1.25, Mdn = 6.00	M = 6.33, SD = 0.87, Mdn = 6.67	-
IMI Choice [scale from 1-7]	M = 2.80, SD = 1.63, Mdn = 3.33	M = 2.00, SD = 1.23, Mdn = 1.67	-
PANAS pos. [scale from 1-5]	M = 3.02, SD = 0.55, Mdn = 2.75	M = 3.77, SD = 0.92, Mdn = 4.10	p = 0.040 t = 2.21
PANAS neg. [scale from 1-5]	M = 2.80, SD = 0.13, Mdn = 2.80	M = 3.00, SD = 0.15, Mdn = 3.00	p = 0.005 t = 3.16

Table 5.2: Dependent variables of the lab study in the Low- and High-SoC groups and results of independent t-tests / Mann-Whitney-U tests ("Diff. sig.") comparing them.

and **R4: Perceived choice is lower when using Endless Universe**. Regarding affective experience, no significant effects were found for the positive (p=0.08) nor the negative affect factor (p=0.62).

Effects of SoC-Personalization Similar to the online study, we split participants in Low- and High-SoC groups and compared these two groups to check for significant effects. To ensure the comparability of the improvement of performance, we did not consider the absolute distance but calculated the relative improvement (i.e. we divided the distance covered in the intervention phase by the distance covered in the baseline phase). Table 5.2 provides an overview of descriptive data and significant differences. It can be seen that we could not find a significant effect in distance improvement between the Low- and High-SoC groups (p=0.07) and no significant effect was found for the perceived exertion between the groups (p=0.24). In addition, none of the factors of the IMI revealed a significant difference (enjoyment: p=0.36; competence: p=0.31; pressure: p=0.28; choice: p=0.23). However, we found a significant effect for affective experiences, i.e. a significant effect was found for both positive (t=2.21, p=0.040) and negative affect (t=3.16, p=0.005). Both positive and negative affect was significantly higher in the High-SoC group. Consequently, we establish **R5: Participants in**

	AC	DI	FS	PH	PL	SO
Distance Improvement [intervention/baseline]	-	**-.387****	-	-	-	-
RPE [scale from 6-20]	-	-	-	-	-	-
IMI Enjoyment [scale from 1-7]	-	-	-	-	-	-
IMI Competence [scale from 1-7]	-	-	-	-	-	**.304***
IMI Pressure [scale from 1-7]	-	-	**.467****	-	-	-
IMI Choice [scale from 1-7]	-	-	-	-	-	-
PANAS pos. [scale from 1-5]	-	-	-	-	-	-
PANAS neg. [scale from 1-5]	-	-	-	-	-	-

Table 5.3: Kendall's τ and significance between the Hexad user types and the dependent variables in the lab study. *p$<$.05, **p$<$.01

the High-SoC group had stronger affective experiences.

Effects of Hexad-Personalization The results of the correlation analysis can be seen in Table 5.3. When analyzing the significant correlations between the dependent variables of the lab study and the AC, PL, SO Hexad user types, we found that the score in the Socializer factor of the Hexad is positively correlated to the perceived competence of the IMI when interacting with Endless Universe, having a medium effect size. This suggest that Socializers perceived the feedback of the gameful design elements as particularly confirming and leads to **R6: Endless Universe positively affected the perceived competence of Socializers**. We also found correlations for Hexad user types besides AC, PL and SO. For these remaining Hexad user types, we expected to find either no conclusive correlations or expected that negative effects on the user experience or performance would be found. Since we did not have specific a-priori formulated assumptions for these user types, we used two-tailed tests for them. We found a negative, medium-sized correlation between the distance improvement and the Disruptor. This suggests that Disruptors were not encouraged to increase their performance by Endless Universe and leads to **R7: The performance of Disruptors was negatively affected by Endless Universe**. Also, we found a medium to strong positive correlation between the perceived pressure and Free Spirits. This means that **R8: Perceived pressure was particularly high for Free Spirits when**

using Endless Universe.

5.2.3 Discussion and Limitations

We implemented a gameful application aiming to motivate users to cover a higher distance on a treadmill to investigate whether the theoretical findings of the storyboards-based study presented in Section 4.5 lead to effects on performance, enjoyment or affective experiences when allowing users to interact with a real implementation of gameful design elements.

In the study, we used Endless Universe to investigate its effectiveness and the effects of behavioral intentions and Hexad user types. We relied on the findings from the online study presented in Section 4.5 to decide which gameful design elements to use for which Hexad user types and stage of change. Consequently, we ended up using Badges, Challenges and Social Competition. These elements were shown to be perceived as significantly more persuasive for users in the High-SoC group in the online study. Also, expected correlations were found between the perceived persuasiveness of these three elements and the Hexad user types Socializer, Achiever and Player. Thus, by using these gameful design elements, we expected to see positive effects on the aforementioned dependent variables for High-SoC users and users scoring particularly high on the Socializer-, Achiever-, or Player- factors of the Hexad.

As a first step of our analysis, we investigated whether the gameful elements used in Endless Universe are effective (**H1**). We found that Endless Universe led to a significant increase in covered distance on the treadmill (**R1**) and also to a subjectively higher exertion (**R2**), thus supporting **H1a: The covered distance is higher when using "Endless Universe"**. This finding is important as it replicates previous research [7, 185] and thus demonstrates the validity of the gameful application itself. We also analyzed whether there is a difference in factors of the IMI. We found that perceived competence and perceived pressure are significantly higher when using Endless Universe (**R3**) and that perceived choice is significantly lower (**R4**). The increased perceived competence is considered as a positive predictor of intrinsic motivation and thus contributes positively to the enjoyment of Endless Universe [300]. On the other hand, perceived pressure is considered as a negative predictor of intrinsic motivation [361]. However, the increase in perceived pressure might also be related to a higher immersion, an enhanced focus on the task and thus a higher sense of flow [94, 148]. Therefore, the significant increase of perceived pressure might be perceived both negatively or positively and should be studied in future work. The fact that perceived choice is significantly lower when using Endless Universe might be related to the introduction of gameful design elements, which establish certain goals and norms which might establish more guidance and thus lead to less choice. Taking **R3** and **R4** together, we consider **H1b: Users perceive running as more enjoyable using Endless Universe** as partially supported. Since no significant effects were found regarding positive or negative affect, **H1c: Users have stronger affective experi-**

ences with Endless Universe is not supported. These mixed results regarding user experience (**H1b, H1c**) might be related to inter-personal differences in the perception of gameful design elements, which have been shown by previous research [261, 342] and as part of the online study presented in Section 4.5.

Therefore, as a next step, we analyzed whether such inter-personal differences could be explained by considering the behavioral intention and Hexad user type of participants. Regarding behavioral intentions (**H2**), we did not find any significant effects between Low- and High-SoC users regarding distance improvement or perceived exertion. Thus **H2a: The improvement in distance is higher for High-SoC users** is not supported, given our data. A potential reason could be related to observer effects, i.e. the effect that participants act more ethically, more conscientiously or more efficiently when being observed [232]. During the experiment, one researcher was in the same room as the participant. This might have affected Low-SoC users more than High-SoC users to improve their performance, since Low-SoC users might have wanted to avoid drawing attention to the fact that they were performing worse than others. Consequently, they might have powered more in the baseline, but could not improve in the intervention. Regarding **H2b: High-SoC users perceive Endless Universe as more enjoyable**, we found no significant differences on the respective IMI factors (enjoyment, competence). Thus, this hypothesis cannot be supported. However, it should be noted that the sample size to compare the Low- and High-SoC users was rather small (10 participants per group), which means that the chance of not finding small to medium sized effects is relatively high. Therefore, we acknowledge that the absence of significant effects (**H2a, H2b**) should not be seen as supporting evidence for the respective null hypotheses. Descriptively, both factors were considerably higher in the High-SoC group, which might suggest that a significant difference could have been found with more participants in each group and that the size of the actual effect was too small to be detected with a total N of 20. Finally, we found a significant increase of both positive and negative affect among High-SoC users (**R5**). This supports **H2c: High-SoC users have stronger affective experiences**.

The fact that positive affect was significantly higher when using Endless Universe supports that tailoring a gameful application to the SoC of users positively affects the user experience. Given that also negative affect was significantly higher when using Endless Universe, these results need to be interpreted more carefully. There is a lot of criticism of considering positive and negative affect as polar opposites [296]. Research has found strong positive correlations between the latent factors of positive and negative affect. Also, the instrument that we used, PANAS, actually does not measure opposite affective experiences (as the names of the latent variables might suggest) [296]. In fact, the items of positive affect were chosen to represent a latent variable (named positive affect) which is defined as activation plus pleasantness. The negative items were chosen to represent a latent construct (named negative affect) defined as activation plus unpleasantness [296, 358]. This shows that these two latent constructs are not opposite on activation, which ultimately means that they are not opposite. We also found supporting evidence of this effect in our data. When analyzing a potential correlation between positive and negative affect, which should be strongly negative, assuming a bipolarity of both latent variables, we found that there exists an insignificant positive correlation between positive and negative affect (Kendall's $\tau = 0.25$, $p=0.17$). This supports the assumption that activation was the deciding cause for the increase in negative affect, instead of unpleasantness. This assumption is further supported by research showing that negative affect can lead to a positive user experience in games [48]. Thus, we conclude that the increase in negative affect seems to be related to higher arousal and activation. Considering a significant increase in positive affect, this allows to interpret the results related to affective experience in a way that supports the assumption of a better user experience when using Endless Universe.

Regarding Hexad user types, we found no evidence for **H3a: The improvement in distance is higher for AC, PL, SO**. Considering that correlations between gameful design elements and Hexad user types using self-reported measures were rather weak [261, 342], the absence of significant correlations between the improvement in distance and these Hexad user types might be related to the low sample size and the resulting low test power. Future work should consider a higher number of participants in order to be able to detect small to medium-size correlations. However, it should be noted that we found a negative correlation between the Disruptor and distance improvement (**R7**), suggesting that Hexad user types seem to have an actual effect on performance. Furthermore, we found that perceived competence was positively correlated to the Socializer user type (**R6**) and that perceived pressure was negatively correlated to the Free Spirit user type (**R8**). **R6** can be interpreted as partially supporting evidence for **H3b: AC, PL, SO perceive Endless Universe as more enjoyable**. For Players and Achievers no significant correlations were found, meaning that **H3b** is only supported for the Socializer. However, taking also **R8** into account, the importance of Hexad user types as a factor moderating the user experience in a gameful fitness application is strengthened and should be investigated further in upcoming interventions. Lastly, we did not find significant correlations regarding affective experiences, thus **H3c: AC, PL, SO have stronger affective experiences** is not supported, given our data. Potentially, this could indicate that tailoring for Hexad user types affects measures related to motivation and the perception of gameful design elements more than the measures related to emotional responses evoked by those gameful design elements. However, this needs to be investigated in future work. Also, it should be noted that we used concrete implementations of gameful design elements, implying that certain design decisions needed to be made, which in turn might have affected the perception of the gameful design elements.

Finally, regarding the question of whether gameful fitness systems should be personalized using behavior change intentions or Hexad user types, the short answer based on our findings is "most probably, yes". No evidence was found for personalization affecting immediate performance-related measures (**H2a,**

H3a). However, we found significant positive effects on the user experience of participants (**H2c, H3b**). This indicates that personalization using behavior change intentions or Hexad user types might affect the performance or behavior of users in the long-run, i.e. the improved user experience might lead to improved retention rates and participants might be more motivated to keep increasing their physical activity. Consequently, beneficial effects on the performance and behavior of users are expected when conducting studies over a longer time-span. This is an important direction that should be followed in future work.

5.2.4 Contribution to Research Questions

The findings from this lab study complement the findings from the online study presented in Section 4.5. While the latter provided supporting evidence for the assumption that Hexad user types and behavior change intentions affect the *perception* of gamification elements, we were able to contribute results which support that these changes in the perception of gamification elements seem to have an actual impact on motivation and affective experiences, when implementing a gamified system that is tailored to specific user groups.

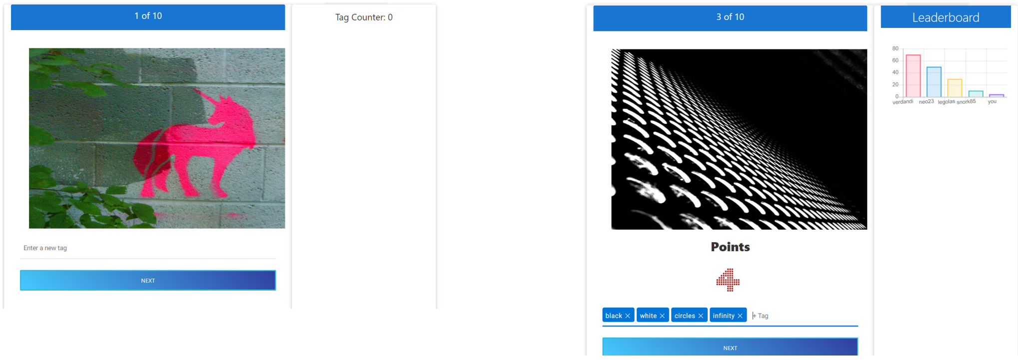
We found that using gamification elements which were perceived particularly well by users in high stages of change leads to stronger affective experiences among that category of users, when exposed to a gamified system implementing them. Moreover, regarding Hexad user types, we found that participants scoring high on the Socializer trait felt particularly competent when exposed to the gamification elements. Since the gamification elements that were implemented were shown to be perceived well by Achievers, Players and Socializers, these findings support the results of the online study presented in Section 4.5. They consequently show that using both Hexad user types and behavior change intentions to personalize gamified systems has an actual impact on user experience related measures of motivation and emotional responses. Additionally, we found a positive correlation between Free Spirits and the pressure factors of the IMI. Since pressure is seen as a negative predictor of intrinsic motivation, our findings show that gamification elements which are particularly suitable for some user types might detrimentally affect other user types. This finding supports the importance of personalizing gamified systems by considering Hexad user types, and could be seen as a possible explanation for the differences in whether gamification leads to positive, neutral or negative effects that were reported in previous research [7, 146, 315].

Overall, these findings mainly contribute to **RQ3**, i.e. the question of what effects personalization of gamified systems has on behavioral and psychological outcomes. In addition, we have found that Endless Universe in general led to an increased performance and perceived exertion. Thus, we contribute evidence for the effect of gamification on behavioral measures, independent of personalization. When analyzing reasons for this effect, we found that users' motivation was affected. We found that the system increased participants' perceived competence, which should contribute positively to intrinsic motivation. However, we also found that perceived choice decreased significantly and perceived pressure increased significantly. These findings suggest that the gamification elements served as rather controlling stimuli in this experiment, i.e. they seemed to undermine autonomy (since perceived choice significantly decreased). Although the increased perceived pressure could be an indication of the prevalence of flow experiences (as discussed in the previous section), it could also support the assumption that the gamification elements were perceived as controlling the behavior of participants. We see these findings as important contributions to **RQ1**, i.e. the question of how gamification affects motivation in behavior change contexts. They show that, while increasing the performance of participants, gamification elements might undermine perceived autonomy and introduce pressure. This suggests that participants increased their performance by employing rather controlled types of regulation (external or introjected regulation), instead of using more autonomous regulatory styles (identified regulation, integrated regulation).

5.3 Effects of (Non-) Personalized Gamification on Task Performance, User Experience and Psychophysiological Reactions

In this section, we focus specifically on the effect of Hexad user types on behavioral and psychological measures. Hexad user types have been used frequently in past research on explaining user preferences in gamified systems, e.g. to explain user preferences in the domain of education [236], physical activity [13], energy conservation [189] or health [261] (see Section 2.4.3). However, as already stated, past research did not investigate these preferences by using an implemented system, but relied on non-interactive materials.

We present findings from a lab study in which participants interacted with an image tagging platform. On this platform, various gamification elements could be dynamically activated or deactivated. Thus, instead of using a static set of gamification elements and investigating whether they are advantageous for user groups which have been shown to perceive these elements particularly well, in this study we dynamically adjust the set of gamification elements to the users' Hexad type. Furthermore, in order to get a more holistic picture of the effects of personalization on psychological measures, we complement survey-based measures with psychophysiological ones. Investigating whether personalizing gamified systems based on the Hexad types of users is important, since we know much about the role of Hexad types in explaining differences in the *perception* of gamification elements but very little about the *actual effects* of personalizing gamified systems based on them, i.e. whether the findings from survey-based studies translate to the real world. In the following, we will elaborate on the design and implementation of the image tagging platform and the user study we have conducted to investigate the aforementioned aspects.



(a) Image tagging task without any gameful design elements

(b) Gamified image tagging task, using a leaderboard and points

Figure 5.2: Image tagging platform

5.3.1 Concept and System Design

We implemented a web-based study platform to investigate the effectiveness of personalization of gameful systems based on Hexad user types. We thereby followed the approach of Mekler et al. [223, 224, 225] by using an image tagging context. The image tagging task on the platform and the set of gameful design elements is described in the following sections.

Image Tagging Task

To ensure comparability, the general task and platform were similar to those used by Mekler et al. [223, 224, 225]. To begin with, the platform allowed participants to get familiar with the tagging task in a tutorial, i.e. allowing participants to add tags for three consecutive images. After completing the tutorial, participants were asked to tag ten images, appearing one at a time and in a random order. We did not use the images employed by Mekler et al. [223, 224, 225]. The authors noted in their most recent study using this platform [224] that utilizing abstract paintings and asking participants to tag emotions makes it hard to objectively assess tag quality. Instead, we decided to consider images that are used for object detection. This allows us to assess the quality of tags in a more objective way. The participants were shown images from the MIRFLICKR-25000 image collection [157]. This collection consists of 25,000 images downloaded from the social photography site Flickr and has been widely used in machine learning research to train object detection algorithms. Participants were asked to type anything they thought of when seeing the image, and could provide tags in a free text field, separating them by pressing enter. Above every image there was a brief description on how to tag it. Figure 5.2a shows the image tagging task.

Gameful Design Elements

We implemented the image tagging platform in a modular way, such that gameful design elements can be activated or deactivated on an individual basis. This allowed for ad-hoc adaptations of the set of gameful design elements depending on the Hexad user type of the participant. We realized five gameful design elements, i.e. badges, points and leaderboard, virtual character, and unlockables, which are described in the following. We made sure that each Hexad user type has at least one suitable gameful design element, based on positively correlated gameful design elements described in the study by Tondello et al. [342]. One exception is the Disruptor, because it is negatively correlated (or not correlated at all) to most gameful design elements [261] making it difficult to find and include suitable gameful design elements. In fact, the Disruptor might also not be as practically relevant as the other user types, since a huge majority of users score lowest in this particular trait [13,342]. Except for leaderboards, all gameful design elements used three score thresholds, which led to a state change of the corresponding element (e.g. unlocking a badge, changing the mood of the virtual character). These thresholds were based on previous gamification research about image tagging [200,311]. The thresholds were the same across all gameful design elements to avoid a bias in the tag quantity depending on which elements are activated (the first state change happens after adding 20 tags, the second after adding 45 tags and the third after adding 70 tags across all images). To ensure comparability to Mekler et al. [223,224,225], who showed five users on their leaderboard, we slightly adapted these thresholds for the leaderboard, without changing the maximum amount to reach the first rank so as not to introduce ceiling effects (in line with all other gameful design elements, the first rank had 70 tags). The second rank had 50, the third 30 and the fourth 10 tags.

Badges This gameful design element is especially suitable for **Achievers** as it builds on the concept of mastery [213]. Previous research has shown that the perception of Badges is positively correlated to the Achiever user type [342]. On

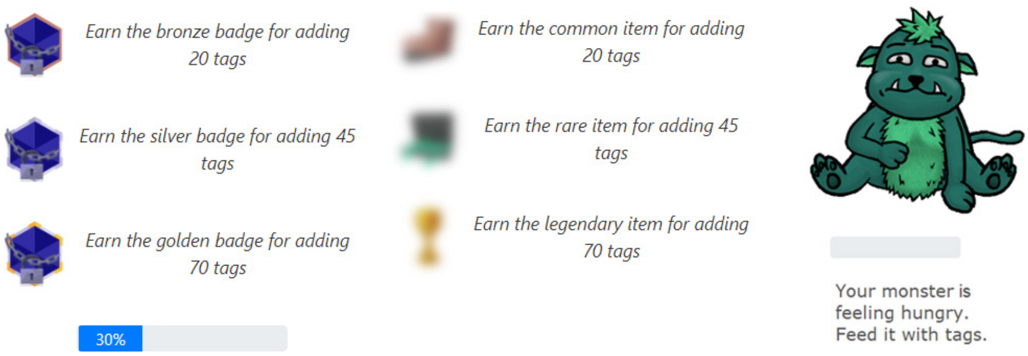


Figure 5.3: The gameful design elements Badges, Unlockables and Virtual Character

the platform, three different Badges (using the score thresholds mentioned before) can be unlocked: A *bronze badge* could be unlocked after adding 20 tags, a *silver badge* after adding 45 tags and a *golden badge* after adding 70 tags. The badges are shown on the right side of the screen. A progress bar indicates the progress towards unlocking the next badge (see Figure 5.3).

Points and Leaderboard Points have been shown to positively affect **Players** [213, 342] and **Socializers** [13, 261]. Similarly, both user types have been shown to be particularly driven by social competition on leaderboards [213,342]. In line with Mekler et al. [224], the leaderboard on our platform shows fictitious users with scores similar to the thresholds established before, to ensure that all participants have equal chances to rise in the ranks. For each tag, users received one point. The leaderboard is shown on the right side of the screen, while the user's current amount of points is shown right below the image (see Figure 5.2b).

Virtual Character **Philanthropists** are driven by purpose and like to care for others [213, 342]. Although past research has not revealed consistent correlations between the Philanthropist factor and the perception of gameful design elements, we expect that a virtual character should be particularly relevant for Philanthropists. A virtual character may induce feelings of care-taking and stimulate striving for purpose [20]. We used an animated virtual monster whose emotional state is coupled to the amount of tags. The three shifts in its emotional state are based on the score thresholds described before. A progress bar indicates the progress towards reaching the next emotional state of the monster (see Figure 5.3). The virtual character field was placed on the left side of each image.

Unlockables Unlockables, i.e. unlocking unknown virtual items, are expected to motivate **Free Spirits** because they are mainly driven by autonomy and curiosity [213,342]. To realize Unlockables, we provided virtual items on the image tagging platform that can be unlocked by adding tags. Reflecting the score thresholds, there were three items differing in rarity (common, rare, epic). The virtual items were blurred and gradually became more visible when adding tags, with the intention to make users curious and more motivated to explore (which is particularly interesting for Free Spirits [342]). The more tags the user added, the clearer the virtual item would become and the closer the user would get to unlocking it. A progress bar indicates progress towards unlocking a certain item. Unlockables were placed on the right side of each image, as can be seen in Figure 5.3.

5.3.2 Evaluation

We used the previously described study platform to investigate the effects of personalization based on Hexad user types on task performance and user experience.

User experience entailed enjoyment, affective experiences and flow and was measured by both questionnaires and psychophysiological reactions. Moreover, we aimed to replicate previous research [200,223,224,225] by investigating the effect of gamification on task performance.

This user study has a within-subjects repeated measures design, in which participants received three conditions: Control, Tailored Gamification, Contra-Tailored Gamification. The selection of (un-) suitable gamification elements for the gamified conditions was based on Hexad user types. The study has been reviewed and received ethics clearance through an institutional Research Ethics Committee[30].

Conditions

The user study had three different conditions, which all participants took part in. There were three conditions differing solely in the type of feedback that was given to users while tagging images. The conditions are explained in the following:

Control ("CO"): In this condition, participants were asked to complete the image tagging task while no gameful design elements were activated.

Tailored Gamification ("TG"): In this condition, we activated gameful design elements that correspond to the Hexad user type of the user (as described in Section 5.3.1). If the user scored (equally) high in two or more Hexad user types, all related gameful design elements were activated.

Contra-Tailored Gamification ("CG"): In this condition, gameful design elements were activated that correspond to the Hexad user types that the user scored lowest on (as described in Section 5.3.1). These elements should be least relevant to the user. If two or more user type scores were equally low, all related gameful design elements were activated.

The Disruptor user type was not considered for assigning suitable or unsuitable gameful design elements (see Section 5.3.1). This is in line with previous research, excluding this user type due to a lack of practical relevance [236]. We decided to activate multiple gameful design elements when a participant had an equal score on their highest or their lowest user type to reflect the traits-based nature of the Hexad model. In case of a conflict, i.e. when a participant scored highest on Player and lowest on Socializer (or vice-versa), we activated the gameful elements corresponding to their second-lowest score, since Players and Socializers are both motivated by points and competition [342] and have been shown to be positively correlated [337].

[30] Saarland University: *Ethical Review Board of the Faculty of Mathematics and Computer Science*, `https://bit.ly/2TK2Qii` (last accessed: 2021-12-01)

Subjective Measures and Analysis

We used the validated Hexad user type questionnaire [337] to assess participants' *user type*. *Enjoyment* and intrinsic motivation were measured using the task evaluation questionnaire of the Intrinsic Motivation Inventory ("IMI") [216, 297]. *Affective experiences* were assessed by using the Positive and Negative Affect Schedule ("PANAS") [358]. The Activity Flow State Scale ("AFSS") [267, 268] was used to measure *flow*. All questionnaires were analyzed as instructed by the authors of the corresponding instruments [216, 297, 337, 358].

Physiological Measures and Analysis

We used features of the electrodermal and cardiovascular system to analyze psychophysiological responses, i.e. flow, enjoyment and affective experiences. The electrodermal system, or sweat gland activity, is solely innervated by the sympathetic "fight-or-flight" nervous system [69]. It was previously used to research games [110,240], and persuasive messages [325] and is among the most commonly used measures to assess *flow* [183]. An important feature of the cardiovascular system is heart rate variability. The variability in time between heartbeats is caused by an interplay between the sympathetic and parasympathetic nervous system [69]. It is the most commonly used measure to detect flow states [183] and was used previously to investigate the persuasiveness of messages [323,324] or assess arousal and affective responses in game-related contexts [351]. It was shown that increases in workload or arousal are associated with decreases in HRV [173]. Changes in blood flow beneath the skin, induced primarily by the sympathetic nervous system, lead to changes in ST [287]. It was found that increased ST at the hands is associated with positive arousal [287] and that skin temperature slopes were more positive for conducive events within games [351]. For a full review on the psychophysiology of emotions and cognition see Jänig [165], Kreibig [190], or Posner et al. [272]. Based on previous research, we included the following physiological measures:

Skin conductance level (SCL): includes the tonic component measured in micro Siemens, μS. In our analysis, SCL was operationalized by dividing the average skin conductance in each condition by the average skin conductance while watching the relaxation video [110].

Skin conductance responses (SCR): concerns the phasic component in electrodermal activity, i.e. the number of abrupt increases in the skin conductance (peaks) [52, 69]. SCR was operationalized by counting the number of skin conductance peaks during a condition and dividing it by the minutes it took to complete that condition. Thus, we analyzed the average number of peaks per minute. To count peaks, we used the scientific python package SciPy[31].

[31] SciPy: *SciPy Reference Guide*, `https://bit.ly/3pFis2f` (last accessed: 2021-12-01)

Heart rate variability (HRV): was measured based on the inter-beat interval in milliseconds, which was used to calculate the square root of the mean squared successive heart period differences ("RMSSD"), a commonly used HRV statistic [69], for each condition. The RMSSD was normalized by dividing it by the RMSSD measured when watching the relaxation video.

Skin temperature (ST): was measured in °C and normalized by dividing the average measure in each condition by the average measure when watching the relaxation video [351].

Similar to Drachen et al. [110] we used a simple form of normalization, in which we divided the average value of a measure by the average value of the respective measure while watching the relaxation video, except for SCR.

Procedure

The user study was conducted in a laboratory. 30 participants were recruited via social media and flyers on the university campus. This number of participants was chosen based on an a-priori calculated power analysis, assuming a medium effect size of $\eta_p^2 = .06$, a power of 80% and a correlation among repeated measures of .5, revealing a minimum number of 27 participants. The expected effect size was informed by analyzing the effect sizes of previous research in the same context [224], which were between $\eta_p^2 = .02$ (medium-small) and $\eta_p^2 = .10$ (medium-large) on relevant measures. The study took approximately 60 minutes to complete. Participants were compensated with an 8 Euro Amazon gift card. Upon their arrival at the study site, the procedure was explained to the participants. After giving consent to participate, participants were asked to take a seat in front of a desktop computer. Next, the validated Empatica E4 wristband [217], a medical-grade wearable device to measure physiological data[32], was put on participants' non-dominant wrist. In the task explanation partial deception was used, since we did not want to reveal that the gamification elements were (contra-) tailored to the participants' Hexad user types. We told them that the purpose of the study was to advance the field of image classification and investigate the perception of different feedback mechanisms in this context. This was done to avoid introducing a potential bias due to participants trying to figure out which condition was being presented to them (which might affect their behavior or flow experiences). After the introduction, participants were asked to complete an initial survey, consisting of demographical data and the validated Hexad user types questionnaire [337]. After they filled out this survey, we assessed a baseline of psychophysiological measures. For this, participants were asked to relax while watching a 5-minute video of sea life [263], in the absence of any discrete environmental event/external stimulus. This video has been successfully used in previous research for the purpose of getting baseline measurements of

[32] E4 wristband: *Real-time physiological data streaming and visualization*, `https://bit.ly/3g86FX6` (last accessed: 2021-12-01)

physiology [263,323]. While participants were watching the video, we prepared the *Tailored Gamification* and the *Contra-Tailored Gamification* conditions based on the results of the Hexad user types questionnaire by activating suitable and unsuitable gameful design elements on the study platform.

Next, participants completed the tutorial consisting of three image tagging tasks (as described in Section 5.3.1). After they completed the tutorial and became familiar with the task itself, the main part of the study followed. Here, participants were asked to tag ten images (one by one) in each of the three conditions. Since the number of conditions was odd, we decided against using a full balanced Latin square to counterbalance conditions (as it would have implied that each participant would have needed to run through each condition twice [171], resulting in a study duration of more than 3 hours, which we deemed ethically unacceptable). Instead, we followed the recommendations provided by Kantowitz et al. [171] and incorporated practice trials (as described before). Thus, the subjects could familiarize themselves with the task beforehand, reducing learning effects [171, 274]. To further minimize order effects, we used random counterbalancing [274] instead of a full Latin square: The order of the conditions as well as the order of the images shown to the user were fully randomized. After tagging ten images, participants were administered a set of questionnaires in each condition, to assess enjoyment, affective experiences and flow.

In order to distinguish psychophysiological measures between conditions in the analysis, the study platform stored the current action of the user (e.g. starting/-completing an image or the whole condition, watching the relaxation video etc.) in the physiological recording. This allowed us to consider solely the physiological data stored while the user was performing the task and exclude all other measures. After completing all three conditions, the participants were debriefed and the full purpose of the study was revealed.

Hypotheses

We investigated the following hypotheses:

H1: Task performance differs across conditions

- **H1a:** Tag quantity is higher in gamified conditions than in *Control*
- **H1b:** Tag quantity is higher in *Tailored Gamification* than in *Contra-Tailored Gamification*
- **H1c:** Tag quality is higher in gamified conditions than in *Control*
- **H1d:** Tag quality is higher in *Tailored Gamification* than in *Contra-Tailored Gamification*

H2: User enjoyment differs across conditions

- **H2a:** User enjoyment is higher in gamified conditions than in *Control*

H2b: User enjoyment is higher in *Tailored Gamification* than in *Contra-Tailored Gamification*

H3: The strength of affective experiences differs across conditions

H3a: Positive affective experiences are stronger in gamified conditions than in *Control*
H3b: Positive affective experiences are stronger in *Tailored Gamification* than in *Contra-Tailored Gamification*

H4: The prevalence of flow experiences differs across conditions

H4a: Experiences of flow are more prevalent in gamified conditions than in *Control*
H4b: Experiences of flow are more prevalent in *Tailored Gamification* than in *Contra-Tailored Gamification*

In general, **H1** is motivated by previous research showing that gamification has an impact on the performance of users when tagging images [200,223,224]. Specifically, we considered both tag quality as well as tag quantity as indicators of task performance (based on [73]) and expected that gamification (independent of whether it is tailored or not) should increase both (**H1a, H1c**). For tag quantity (**H1a**), this assumption is based on previous research showing that gamification increases the number of tags in an image-tagging context [200,224]. We further hypothesized that gamification should lead to an enhanced tag quality (**H1c**) since a meta-analysis on performance predictors came to the conclusion that motivation (especially intrinsic motivation, but also extrinsic incentives), which should be positively affected by gamification and goal-setting [206,223], predicts quality [73]. **H1b** and **H1d** build on the assumption that gameful design elements which are tailored to a users' Hexad type lead to an additional increase on both performance measures, due to previous research showing correlations between user preferences for gameful design elements and their Hexad user type [342].

H2 refers to the enjoyment of tagging images. Based on literature reviews by Seaborn and Fels [315] and Hamari et al. [146], we expected that enjoyment should be improved by gamification (whether it is tailored or not) (**H2a**). We also hypothesized that a set of gameful design elements which is tailored to a user's Hexad type should lead to an increased enjoyment when compared to a contra-tailored set of gameful design elements (**H2b)**, since user preferences [342] should have an impact on the user experience of a gameful system. **H3** follows the same argumentation. We expect that an increased enjoyment is related to positive affective experiences and thus assume that gamification should lead to an increase in positive affective experiences (**H3a**), especially when tailored to the user's Hexad type (**H3b**).

Lastly, flow experiences, which can be defined as "the holistic sensation that people feel when they act with total involvement" [93], are related to optimal

P	PL	SO	AC	PH	FS	DI		P	PL	SO	AC	PH	FS	DI
1	26	24	27*	22*	27*	21		**16**	24	24	23*	26*	23*	13
2	17*	28	27	22	28*	28		**17**	25	27*	27*	27*	25*	15
3	19*	27	26	28*	23	12		**18**	28*	21	27	20*	22	16
4	21*	23	23	25*	25*	15		**19**	26*	24	26*	24*	25	20
5	25*	17	23	22	21*	12		**20**	21*	21*	23	24*	24*	19
6	28*	27	27*	28*	27*	16		**21**	19	23	25*	23	19*	12
7	21*	27	24	28*	21*	18		**22**	16*	17	18	23*	22	17
8	23	26*	24	23	20*	15		**23**	19	14*	19*	19*	19*	15
9	27*	19	23	21*	26	13		**24**	16*	24	26	28*	27	23
10	20*	23	24	27*	24	14		**25**	25*	24	24	22	21*	13
11	20*	24	26*	26*	25	21		**26**	26*	17	25	24*	25	13
12	25	27*	25	23	21*	8		**27**	19*	28	28	28*	28*	10
13	17*	18	19	21*	19	12		**28**	24*	18	23	23	22*	18
14	23	24*	22	23	21*	16		**29**	22	26	27*	26	22*	16
15	26*	22	23	15*	21	19								

Table 5.4: Hexad scores of all participants. Green cells represent particularly high scores in the respective user type, red cells represent particularly low scores. An asterisk marks which user type was considered for the personalization of the gameful design elements.

task performance [268]. This includes being completely focused on the task and an increased engagement [268]. Thus, following from **H1–H3**, we expect that flow experiences are more frequent in the gamified conditions (**H4a**) and that personalizing gameful design elements to a user's Hexad type further increases the prevalence of flow experiences (**H4b**).

Results

In the following, we will present the results of our user study. After presenting descriptive results regrading our sample, findings related to task performance, user experience and psychophysiological reactions will be described.

To investigate the aforementioned hypotheses, we used repeated measures ANOVAs to compare the dependent variables between the three conditions. When assumptions for the ANOVA were not met, Friedman tests were used as non-parametric counterparts. When using Friedman tests, the Durbin-Conover method was used for post-hoc analysis. The Bonferroni-Holm method was used in both cases to control the family-wise error rate.

Participants Out of 30 participants, one had to be excluded due to technical problems during the study, leading to a total sample size of 29 which was considered for the analysis.

Out of these participants, 10 self-reported their gender as female and 19 as male. Regarding age, 10 participants were aged 18-24 years, 18 participants were aged 25-31 years, and 1 participant was aged 32-38 years. We assessed gaming familiarity with 3 items ("I consider myself as gaming-affine", "I frequently play video games", "I have a passion for video games") with 5-point scales (1=strongly disagree). The means were rather neutral: 3.12, 2.82 and 2.92, respectively. The Hexad user types average scores are similar to the averages reported in the validation study of the Hexad questionnaire by Tondello et al. [337]. Achievers showed the highest average scores (M=24.28, SD=2.52), followed by Philanthropists (M=23.83, SD=3.10), Free Spirits (M=23.20, SD=2.71) and Socializers (M=22.90, SD=3.80). Players (M=22.34, SD=3.58) and Disruptors (M=15.86, SD=4.13) followed with lower average scores. Table 5.4 shows the Hexad scores of all participants as well as which user types were highest (marked green) or lowest (marked red). An asterisk marks which user type was considered for the personalization of the gameful design elements (sometimes, the highest/lowest score could not be used for the selection of suitable/unsuitable gameful design elements due to a conflict in the reported preferences in the literature; see Section 5.3.2).

Task Performance Overall, participants provided 3,967 individual tags (1,114 in Control ("CO"), 1,402 in Tailored Gamification ("TG") and 1,451 in Contra-Tailored Gamification ("CG")). Table 5.5 provides an overview of the mean and median tag count per condition. We compared the average number of tags per condition and found they differed significantly ($F(2, 56) = 13.56, p <.001$, $\eta_p^2 = .33$). Pairwise comparisons revealed result **R1: The number of tags in both gamified conditions is significantly higher than in CO** ($p_{\text{holm}} <.001$ each). When comparing the TG and CG conditions, no significant result was found ($p_{\text{holm}} =.48$).

To analyze tag quality, we followed a qualitative coding process, similar to Mekler et al. [224]. The coding process was conducted by two independent raters who manually inspected each of the 3,967 individual tags provided for the images and rated whether the tag was: neither related to any given object in the image nor captures a specific mood, or was just nonsense (value 1); describes a mood or color scheme that was present in the pictures but not a specific object (value 2) or describes a concrete object in the picture (value 3). After both raters rated all tags, the inter-rater agreement was calculated using Cohen's Kappa κ. The result was κ=.66, which is considered as substantial agreement [221]. The average quality of tags for each condition is shown in Table 5.5. We compared the average rating per participant across conditions and found that it differed significantly ($F(2, 56) = 3.97, p =.024$, $\eta_p^2 = .02$). As part of the post-hoc procedure, we found that **R2: The average tag quality is significantly higher in both gamified conditions than in CO** ($p_{\text{holm}} <.05$ each). However, no significant difference was found between TG and CG ($p_{\text{holm}} =.82$).

Also, the amount of tags per minute differed significantly ($F(2, 56) = 16.64, p <.001$,

		Control			Tailored			Contra-Tailored		
		Mean	SD	Median	Mean	SD	Median	Mean	SD	Median
Perfor.	**Tag Count***	38.41	21.44	31.00	48.34	24.71	46.00	50.03	24.99	45.00
	Tag Qual.*	2.66	0.21	2.71	2.73	0.19	2.79	2.72	0.23	2.79
	Tags/min*.	3.17	1.53	2.92	3.97	1.73	3.98	4.40	2.06	4.25
IMI [sum]	**Enjoyment***	26.72	10.36	26.00	28.03	10.80	27.00	24.97	9.98	26.00
	Competence	19.69	4.74	19.00	20.76	6.90	20.00	19.90	5.17	20.00
	Choice	23.03	5.33	23.00	23.10	4.88	23.00	22.45	5.21	22.00
	Pressure	13.03	5.05	14.00	14.55	6.49	14.00	14.38	5.45	14.00
PANAS [sum]	**Pos. Affect***	27.07	8.15	27.00	27.62	8.94	27.00	22.79	10.39	18.00
	Neg. Affect	16.59	4.79	16.00	15.90	5.03	15.00	15.00	3.32	14.00
AFSS [sum]	**MAA***	9.52	2.72	9.00	10.48	2.68	11.00	9.17	3.11	9.00
	CG	11.03	2.34	12.00	11.59	2.51	12.00	10.52	2.49	10.00
	CO	15.41	2.28	16.00	15.03	3.17	15.00	14.28	3.32	14.28
	UF	6.59	1.82	6.00	7.48	1.72	8.00	6.83	2.28	8.00
	CS	9.28	2.37	9.00	9.79	2.72	10.00	9.34	2.35	9.00
	TT	9.07	2.81	9.00	9.66	3.25	10.00	8.86	3.03	9.00
	CN	7.72	1.36	8.00	7.66	1.59	8.00	7.48	1.62	8.00
	SC	10.52	2.57	11.00	10.59	2.85	10.00	9.83	2.52	9.00
	AE	8.97	2.74	9.00	10.17	3.12	11.00	9.28	3.28	9.00
Psy.phys. [normalized]	RMSSD	1.01	0.04	1.00	1.01	0.05	1.00	1.02	0.07	1.00
	SCL	1.22	0.39	1.13	1.33	0.56	1.18	1.34	0.76	1.10
	SCR* [p/min]	44.27	14.26	50.34	49.42	9.92	52.40	50.86	7.28	51.85
	ST*	1.02	0.03	1.01	1.04	0.04	1.03	1.04	0.05	1.03

Table 5.5: Mean, standard deviation ("SD") and median for each dependent variable and condition. Bold entries with * represent dependent variables for which a significant difference across conditions was found.

$\eta_p^2 = .37$). We found that **R3: The amount of tags per minute is significantly higher in both gamified conditions than in CO** (p_{holm} <.001 each). Also, when comparing the tailored (TG) and contra-tailored (CG) conditions, we found that **R4: The amount of tags per minute in the Contra-Tailored Gamification condition is significantly higher than in the Tailored Gamification condition** (p_{holm} =.049). Together with **R2**, **R3** suggests that gamification might have helped participants to come up with good tags, since the average time per tag decreased.

Subjective User Experience To analyze the user experience, we considered enjoyment or intrinsic motivation, affective and flow experiences. In this section, we report the results of the survey-based measures. All descriptive data can be found in Table 5.5.

Regarding the IMI factors, we did not find a significant effect for the competence ($F(2, 56) = .65, p = .52$), choice ($F(2, 56) = .32, p = .73$) nor pressure ($F(2, 56) = 1.51, p = .23$) factors. However, the enjoyment factor differed significantly across the conditions ($F(2, 56) = 3.45, p = .039$, $\eta_p^2 = .11$). While there were no significant differences between the gamified conditions and CO (p_{holm} =.28

each), we found that **R5: Enjoyment is significantly higher in the Tailored Gamification condition than in Contra-Tailored Gamification** (p_{holm} =.003).

When analyzing the positive and negative affect factors of the PANAS, we found that positive affect differed significantly between the three conditions ($F(2, 56) = 6.39, p$ =.003, $\eta_p^2 = .19$). Post-hoc comparisons showed that positive affect in TG was not significantly higher than in CO (p_{holm} =.71). However, positive affect was significantly higher in TG than in CG (p_{holm} =.006), leading to **R6: Positive affect is significantly higher in Tailored Gamification than in Contra-Tailored Gamification**. In addition, we found that **R7: Positive affect is significantly lower in Contra-Tailored Gamification than in Control** (p_{holm} =.001). These results not only show that selecting gameful design elements matching the users' Hexad types leads to increased positive affective experiences, but also that choosing unsuitable gameful design elements is worse than having no gameful design elements at all, regarding affective experiences. Concerning negative affect, no significant differences were found ($F(2, 56) = 2.06, p$ =.14).

To measure self-assessed flow states, we relied on the AFSS, having nine factors (see Table 5.5). We found a significant effect on the Merging Actions and Awareness factor ($F(2, 56) = 3.31p$ =.044, $\eta_p^2 = .11$), indicating that flow experiences differed across the three conditions [94]. Based on the pairwise comparisons, we found that there were no significant differences between CO and TG (p_{holm} =.15) nor between CO and CG (p_{holm} =.52). However, similar to **R5** and **R6**, we found that **R8: Flow experiences are significantly more prevalent in Tailored Gamification than in Contra-Tailored Gamification** (p_{holm} =.048). This indicates that the selection of gameful design elements has an effect on task immersion and flow experiences [94] and that personalizing the gameful design elements to a user's Hexad type positively affects these experiences. For the remaining factors of the AFSS, no significant differences were found. However, a similar pattern as in the MAA factor can be seen in the remaining factors as well, almost reaching significance on the clear goals ("CG", $F(2, 56) = 2.36, p$ =.10), concentration on task at hand ("CO", $F(2, 56) = 1.98, p$ =.15), unambiguous feedback ("UF", $F(2, 56) = 2.45, p$ =.10) and autotelic experience ("AE", $F(2, 56) = 2.80, p$ =.07) factors. We also did not find any significant effect on the challenge skill balance ("CS", $F(2, 56) = .67, p$ =.52), transformation of time ("TT", $F(2, 56) = 1.09, p$ =.34), sense of control ("CN", $F(2, 56) = .35, p$ =.71) nor the loss of self-consciousness ("SC", $F(2, 56) = .94, p$ =.40) factors.

Physiological User Experience To complement the survey-based measures, we used physiological measures to assess the user experience in a multi-faceted way. These measures were analyzed using a Friedman test instead of an ANOVA, because the assumption of normality and/or the assumption of sphericity were violated. Table 5.5 provides an overview of the descriptive data across conditions.

For RMSSD, we found no effects between conditions ($\chi^2(2) = 1.10, p$ =.58). Also, no effects were found regarding potential changes in SCL ($\chi^2(2) = .48, p$ =.79).

However, we found that the number of SCRs differed between the conditions ($\chi^2(2) = 18.83, p < .001$). As revealed by post-hoc comparisons, both gamified conditions showed an increase in SCRs compared to CO, leading to **R9: The number of peaks in the EDA (SCRs) is significantly higher in both gamified conditions than in Control** (p_{holm} <.001 each). This shows an increased sympathetic arousal in both gamified conditions and hints at either increased flow states (assuming that the arousal is positively valenced) or increased pressure or tension (assuming that the increased arousal is negatively valenced) [69]. When comparing TG and CG, no significant difference was found regarding SCR (p_{holm} =.64). We also found a significant difference in skin temperature ($\chi^2(2) = 37.72, p$ <.001). In line with **R9**, both gamified conditions differed from the CO condition, i.e. **R10: Skin temperature is significantly higher in both gamified conditions than in Control** (p_{holm} <.001). No effects were found between TG and CG (p_{holm} =.83).

5.3.3 Discussion and Limitations

We investigated whether personalization based on Hexad user types affects task performance, user experience or psychophysiological reactions by letting participants tag images in Control (without any gameful design elements), Tailored Gamification (using gameful design elements tailored to the users' highest-scored Hexad type) and Contra-Tailored Gamification (using gameful design elements matching the users' lowest-scored Hexad type). Our results show that both gamified conditions lead to an increase in the amount of tags, compared to the Control condition (**R1**). This supports **H1a: The number of tags is higher in gamified conditions than in *Control***. It shows that gamification, independent of whether it is tailored or not, increases the number of tags in an image tagging context. This is in line with previous research by Mekler et al. [223,224,225] as well as Lessel et al. [200] and therefore contributes a replication of previous results using a static set of gameful design elements. When comparing the number of tags between the Tailored Gamification and the Contra-Tailored Gamification conditions, we did not find a significant difference. Thus, **H1b: The number of tags is higher in *Tailored Gamification* than in *Contra-Tailored Gamification*** is not supported, given our data. This might be explainable by the fact that all gameful design elements, regardless of their suitability, introduce goals. According to goal-setting theory [206], goals motivate people by introducing a state of tension that activates actions. Also, the experimental setting and the fact that participants were compensated for participating might have led to participants feeling obligated to meet these established goals, independent of the gameful design elements that were activated and their user experience. Therefore, this aspect needs further research and should be investigated in in-the-wild studies over a longer time-span.

Related to tag quality, we found that the average quality of tags was significantly higher in both gamified conditions (**R2**), and that the time users took to create a tag was significantly lower in the gamified conditions (**R3**), both adding support for **H1c: Tag quality is higher in gamified conditions than in *Control***. This finding is explainable by the fact that increases in a user's motivation to perform a task (which likely occur due to the gamification that was used [206, 223]) have been shown to lead to increases in the quality of the task outcome [73]. Moreover, the fact that the mean time to add a tag to an image was significantly lower in both gamified conditions than in CO (**R3**) suggests that participants had to think less about which tags to provide, which might have been caused by the potentially stimulating gamification environment. However, it should be considered that previous work in the same context did not find significant effects regarding tag quality [200, 224, 311]. In contrast to these previous studies, we used images showing actual real-world objects which participants had to tag, instead of using abstract paintings and asking participants to tag the mood that the images might evoke. This allowed us to assess tag quality in a more objective way and might be the reason why we were able to find an effect of gamification on tag quality. Additionally, we found that the average time per tag was significantly lower in the NG than in the TG condition (**R4**). Since no difference between the CG and the TG condition could been found regarding tag quality, the reason for this result needs further investigation in the future. Also, based on these results, we cannot support **H1d: Tag quality is higher in *Tailored Gamification* than in *Contra-Tailored Gamification***.

Regarding the enjoyment of tagging images, we found no significant difference between the gamified conditions and CO. Thus, **H2a: User enjoyment is higher in gamified conditions than in *Control*** is not supported. This is similar to previous research in the same context, which did not find any effect on the IMI enjoyment factor [200, 311]. A potential reason might be that the task itself, i.e. tagging images, was perceived as unexciting or boring. However, we found a significant effect between the Tailored Gamification and the Contra-Tailored Gamification condition regarding enjoyment. Our results show that participants in the Tailored Gamification condition enjoyed tagging images significantly more than in the Contra-Tailored Gamification condition (**R5**). Thus, **H2b: User enjoyment is higher in *Tailored Gamification* than in *Contra-Tailored Gamification*** is supported. This shows that personalizing a gameful application based on Hexad user types leads to an increased task enjoyment.

Related to this, we investigated whether positive or negative affect differs across conditions. Again, no significant difference between CO and both gamified conditions was found. Therefore, **H3a: Positive affective experiences are stronger in gamified conditions than in *Control*** is not supported. Similar to the absence of an effect regarding enjoyment between gamified conditions and CO, the repetitive nature of the task itself could be the reason here again. However, in line with **R5**, a significant difference was found between Tailored Gamification and Contra-Tailored Gamification. Positive affect was significantly higher when participants were exposed to gameful design elements that were suitable for their highest-scored Hexad type (**R6**), adding support for **H3b: Positive affec-**

tive experiences are stronger in *Tailored Gamification* than in *Contra-Tailored Gamification*. **R7**, i.e. the fact that positive affect was even significantly lower in CG than it was in CO, further supports **H3b** and underlines the importance of personalization for the users' experience in gameful systems.

The same trends and effects as for enjoyment and affective experiences can be seen when analyzing flow experiences. Based on the AFSS questionnaire, we found a significant difference between the TG and CG conditions on the Merging Actions and Awareness (MAA) factor (**R8**), whereas no significant effects were found on the remaining eight factors of the AFSS. However, since the results on the remaining factors mostly followed the same pattern as on the MAA factor, we see the significant effect on the MAA factor together with the same trend on most of the remaining factors as supporting evidence for **H4b: Experiences of flow are more prevalent in *Tailored Gamification* than in *Contra-Tailored Gamification*.** Since no significant effects were found between CO and both gamified conditions, we cannot support **H4a: Experiences of flow are more prevalent in gamified conditions than in *Control*** using the survey-based flow assessment.

Taking together the aforementioned results related to the user experience covering self-reported enjoyment, affective experiences and flow (**R5–R8**), we see that all these factors differ significantly between Tailored Gamification and Contra-Tailored Gamification. Thus, our results indicate that selecting a suitable set of gameful design elements, based on the users' Hexad type, leads to improvements in enjoyment, positive affect and flow. These are important results that could be found because participants were actually exposed to an interactive system instead of presented storyboards [261] or textual descriptions [342], which have been used in previous work.

Lastly, we analyzed physiological reactions (RMSSD, SCL, SCR, ST) to complement the survey-based measures. While we did not find significant effects for heart rate variability nor skin conductance level, a significant effect was found for skin conductance responses as well as skin temperature. Both the amount of peaks in the skin conductance signal as well as the skin temperature were significantly higher in the gamified conditions (**R9, R10**), showing that participants were more aroused when interacting with a gameful system than when there were no gameful design elements at all. This is a noteworthy contribution on its own, since there is no previous research showing that gamification affects psychophysiological measures, as far as we know. Thus, our findings provide evidence for the fact that design elements providing gameful feedback affect users' state of arousal. However, these psychophysiologial measures do not allow us to assess whether participants are positively or negatively aroused. Therefore, to interpret these findings, we consider the results from the survey-based instruments measuring flow, affective experiences and enjoyment, since both skin temperature and skin conductance responses were shown to be linked to such measures [183]. Combining them suggests that the significant increase in SCR and ST seems to be related to positive experiences in the Tailored Gamification condition (supported by the increase in positive affect, enjoyment and flow) whereas it seems to be related to negatively valenced arousal in the Contra-Tailored Gamification condition (supported by the fact that positive affect, enjoyment and flow are rated lower in Contra-Tailored Gamification than in the CO condition, and significant effects were found between TG and CG).

Limitations

Our study has several limitations, which should be considered when interpreting our findings. Although the selection of suitable gameful design elements is based on previous research, certain design decisions when realizing these gameful design elements are inherently a matter of interpretation, which might affect the external validity of our results. The fact that we investigated a specific context (image tagging) adds to this, although it should be noted that we contribute to the replication of previous findings in this context, since image tagging is frequently used for basic gamification research [199, 200, 223, 224, 225].

Another limitation concerns the approach we followed to select which gameful design elements to activate in the TG and CG conditions. Here, we decided to activate the gameful design elements, which were shown to be particularly relevant for the participant's Hexad type having the highest (TG) or lowest (CG) score. While this approach was straightforward to implement for most participants, we had two special cases that should be considered: First, since the Hexad model is a traits model, it could happen that participants scored highest/lowest on multiple user types. In this case, we considered these multiple user types equally to select which gameful design elements to use.

Second, due to the fact that the mapping between relevant gameful design elements and Hexad user types is not one to one, it could happen that the set of suitable gameful design elements overlaps with the set of unsuitable gameful design elements (i.e. Socializers and Players both have a strong preference for leaderboards [342], which means that participants scoring highest on Socializer and lowest on Player would get the leaderboard in both TG and CG conditions). To avoid this and ensure that participants actually are presented with irrelevant gameful design elements, we selected the user type where participants had the second lowest score to decide which gameful design elements to activate in the CG condition. These two decisions need to be considered when replicating and interpreting our results.

Related to this, it must be noted that we excluded the Disruptor type (similar to previous research by Mora et al. [236]), since no clear relationships to gameful design elements have been shown previously. However, since the Disruptor is by far the least common user type [342], we do not see a major limitation in terms of the practical relevance of our findings. Lastly, we would like to acknowledge that the validity of the psychophysiological measures is tightly coupled to the technical specification of the Empatica E4 wristband which was used. Although the validity of the band has been demonstrated [217], and participants were not moving a lot (due to the task itself), a certain level of noise is unavoidable.

5.3.4 Contribution to Research Questions

Our findings demonstrate that personalization, i.e. activating gamification elements that are suitable for users' highest-scored Hexad types, has an effect on psychological measures. For instance, we found that participants enjoyed the system significantly more in the tailored than in the contra-tailored condition. This suggests that they were more intrinsically motivated when presented with suitable gamification elements. This is supported by the finding that positive affect was significantly higher in the tailored condition than in the contra-tailored condition. Notably, positive affect was even significantly lower in the contra-tailored condition than in the control. In particular, this suggests that using an unsuitable set of gamification elements may lead to detrimental effects on the users' experience. In line with the previous findings, we were able to show that the prevalence of flow experiences is higher in the tailored than in the contra-tailored condition. Overall, these results show that personalization based on Hexad user types provides benefits over using a one-size-fits-all approach.

Although we could not show that these benefits in terms of motivation and the user experience translate to benefits in the performance of users in the system, we assume that these benefits may become visible when investigating a personalized system over a longer time span in an in-the-wild study. These results and insights provide answers to **RQ3** and underline the importance of personalizing gamified systems to the user. In addition, we also contribute important results for **RQ1**. First, we provide evidence for the fact that gamification not only affects self-reported measures of motivation and user experience, but also affects psychophysiological measures of arousal. This contributes a more objective perspective to gamification research and complements previous findings. We were also able to show that gamification increased performance on the task: Participants provided more tags, and tags of significantly higher quality. This shows that gamification affected the participants' behavior. When analyzing motivation-related measures, we see that this increase in performance seems to be caused by different types of motivation.

Since enjoyment, positive affect and the prevalence of flow experiences were significantly higher when participants received suitable gamification elements, it seems that their motivation may have been more autonomously regulated in this condition. Nevertheless, since the performance was also significantly higher in the contra-tailored condition, it seems that although the type of motivation may have differed, gamification was still capable of increasing performance. This underlines the importance of investigating *how* gamification motivates, and whether the type of motivation is controlling or autonomously regulated, to better understand the users' experience in gamified systems, instead of solely focusing on whether it leads to an increased performance.

5.4 A Long-Term Investigation on the Effects of (Personalized) Gamification on Course Participation in a Gym

In the previous two sections, we investigated the actual effects of personalized gamification on psychological and behavioral outcomes in two controlled lab experiments. In these experiments, we found that psychological measures of affect, flow and motivation were positively affected when using gamification elements personalized to the users' Hexad type or stage of change. However, we could not find supporting evidence for the fact that these positive effects on psychological measures also lead to an increased performance, or affect behavioral outcomes more than non-personalized gamification does. We assumed that this is because the duration of the interventions was too short to find such differences and expected that the increased motivation and improved user experience could affect performance and behavior over a longer time span.

In this section, we investigate this assumption by conducting an almost-two-year-long in-the-wild study on the effects of using personalized versus non-personalized gamification elements on the course booking behavior of users in a gym. In line with the previously described studies, we used Hexad user types and the findings from previous research and from Chapter 4 to decide whether certain gamification elements are particularly suitable for certain users or not. Besides contributing to the question of whether personalized gamification has advantages regarding long-term effectiveness over non-personalized gamification, we also contribute to gamification research in general, which lacks empirical evidence for its long-term success [7,185]. Aldenaini et al. [7] found in their meta-review that a huge majority of previous studies had a rather short duration. Thus, whether gamification has a lasting impact on the physical activity levels of participants remains unclear. This in turn poses the question, to which this study contributes, of whether gamification is a suitable approach to motivate users to change their behavior sustainably.

5.4.1 Concept and System Design

To investigate the effectiveness of gamification elements and personalization on the amount of booked fitness courses, we were allowed to integrate gamification elements into an existing web-based course booking system, which will be described in the following.

Gym and Course Booking System

The local gym in Germany, which we collaborated with, does solely offer courses, i.e. users have to register beforehand for a course manually and are not allowed to participate in any fitness activity when not registered. Thus, course bookings are binding. To book fitness courses, users have to login to the website of the gym with personal credentials. Next, they can see which courses are offered in the following two weeks and can freely decide which courses to book. It is important to note that courses need either to be payed manually or users may have a subscription. In case of the latter, users pay a monthly fee and are allowed to participate in a fixed number of courses per week. If they book more courses in a week than is covered by their subscription, the respective courses will be charged manually. This means that each booked course needs to be payed.

There are different types of courses. While some focus on strength, others focus on endurance or combine fitness activities and exercises with teaching the proper technique. The number of people in a course typically is bound to ten or twelve. Also, courses are offered throughout the day and usually take an hour. The earliest courses start at around 7am and the latest courses end at 9pm. There are courses on each day of the week, but the number of courses per day is slightly reduced on Sundays.

Gamification Elements

We included four different gamification elements, which are described in the following. Additionally, we implemented an option to upload profile pictures and set a username, which would be shown on the leaderboard and when booking a course (such that users could see which other users take part in the course). This was done to allow for social comparison outside of the leaderboard. The selection of gamification elements orientated on frequently used gamification elements in fitness contexts [185]. We also implemented an option to stay anonymous. This means that the username is not shown when booking courses and is also hidden in the leaderboard. Per default, users are anonymous and need to enable that they would like to be shown on the leaderboard/when booking courses in their profile settings.

Activity points Points have been shown to positively affect **Players** [213, 342] and **Socializers** [13, 261]. We implemented a point system based on so-called Activity Points ("AP"). When users book a course, they receive 10 AP. The current amount of earned AP is permanently shown on the left side of the navigation section as well as when users navigate to their profile page (see Figure 5.5). When collecting APs, users make progress towards reaching their next activity level, which is explained in the following section.

Levels Levels and progression have been shown to be particularly motivating for **Achievers** and **Players** [213,342]. We introduced the concept of activity levels, meaning that collecting Activity Points leads to increases in the activity level of a user. The activity level is visualized together with the current amount of Activity Points in a users' profile page as well as permanently on the navigation side bar. Below the current activity level, a progress bar indicates the current

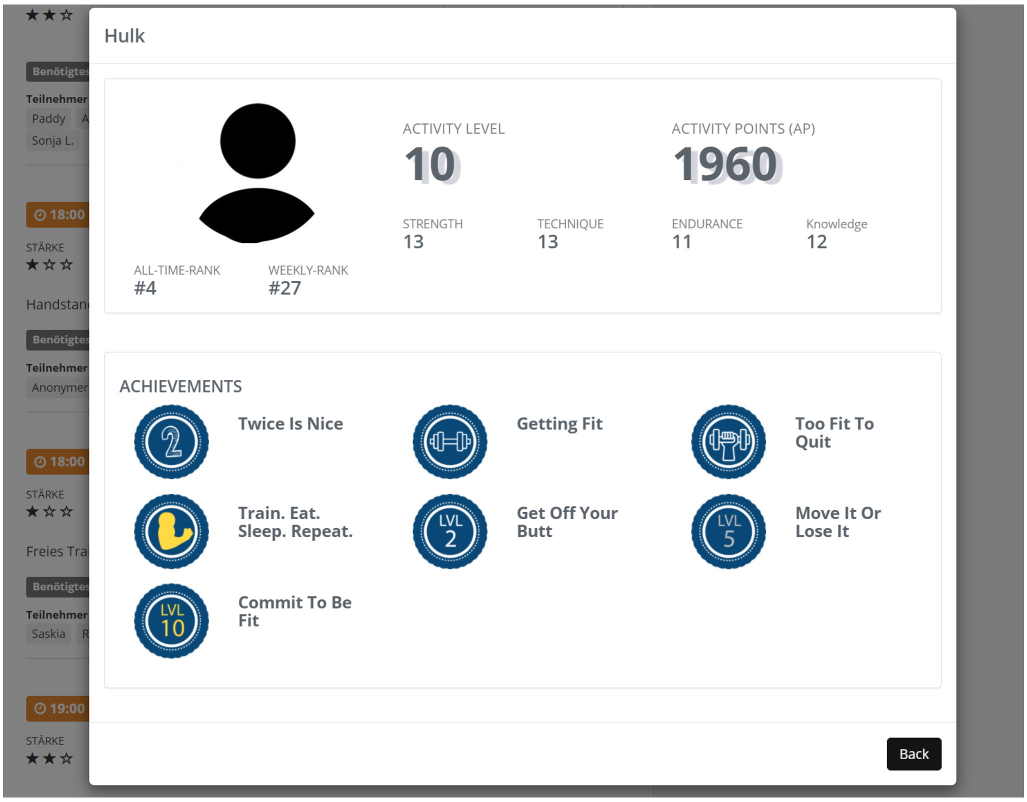


Figure 5.4: Profile information, after clicking on a user when booking a course or on the leaderboard

progress towards the next activity level and shows how many Activity Points are missing. The number of Activity Points to reach the next activity level grows logarithmically with increasing levels.

Besides the activity level, we also introduced four different attribute levels: The Strength, Endurance, Technique and the Knowledge levels. In consultation with the professional fitness trainers, all courses were rated in terms of these attributes, such that users who book a certain course, receive attribute points and make progress in corresponding attribute levels (depending on the type of course). When booking a fitness course, we added this information such that users know how to improve their attributes and to give them an indication whether their training reflects their personal fitness goals. Besides showing the fitness attributes, we visualized the users who have booked the course (as stated before). When clicking on a user, their profile picture, current rank on the leaderboard, unlocked badges, activity level and the level in each attribute is shown (if the user is not anonymous), see Figure 5.4.

Badges This gameful design element is especially suitable for **Achievers** as it builds on the concept of mastery [213]. Previous research has shown that the perception of Badges is positively correlated to the **Achiever** and **Player** user types [342]. We integrated nine different badges, of which three are triggered by making progress in the activity level, four by completing a certain number of courses per week and two by participating in particularly early or late courses. Figure 5.5 shows a subset of badges offered.

Social Competition / Leaderboard **Players** and **Socializers** were shown to be particularly driven by Social Competition and Leaderboards [213, 342]. We implemented a weekly and an all-time leaderboard to both allow new users to keep up with others and also reward long-term participation. The leaderboard was shown on the profile page (see Figure 5.5) and the current rank of users was also shown when clicking on a certain user when booking a course (see Figure 5.4). When clicking on an entry in the leaderboard, the same dialog opened as when clicking on a user when booking a course.

5.4.2 Evaluation

To investigate the long-term effects of gamification on the number of booked courses as well as the role of personalization based on Hexad user types, we analyzed a fully anonymous dataset after the gamification elements were active for one year (*Gamification* phase) on the booking system and compared it to the year before gamification was introduced (*Baseline* phase).

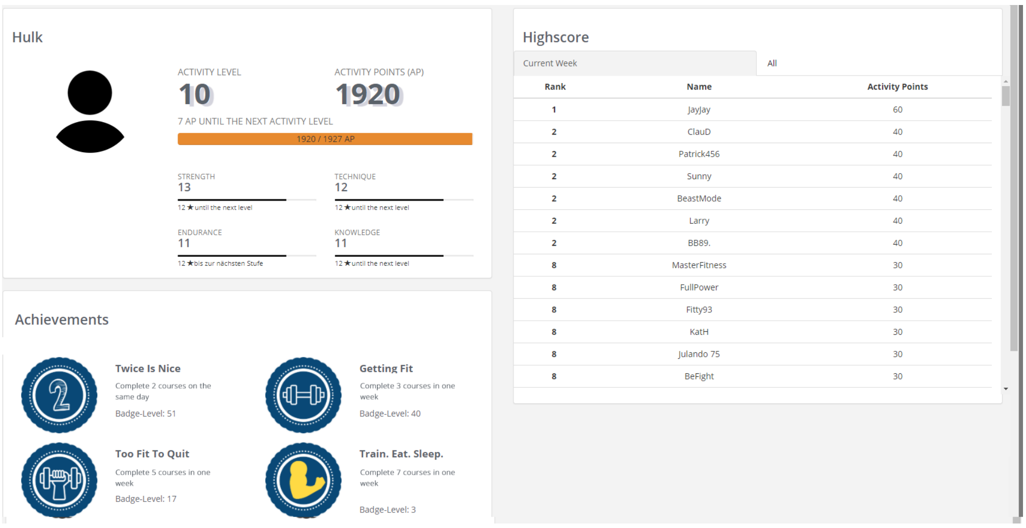


Figure 5.5: Gamification elements shown in the personal profile of the user. The user interface shows the current activity level and the level for each attribute as well as the current progress towards the next level. Also, Activity Points, unlocked badges and the leaderboard are shown. If users have the same score, they are ranked on the same position.

Method

After the gamification elements were activated, users of the booking system were asked to fill out the validated Hexad user types questionnaire [337] voluntarily. They also had the option to skip filling out the questionnaire. When selecting users to be analyzed, we had the following criteria: First, users should be registered for at least two years, i.e. before 2018-10-08 (we considered data between 2018-10-08 and 2019-10-07 as Baseline). Second, users should have booked at least one course in the month before the end of the Baseline and before the end of the Gamification phase (the Gamification phase started on 2019-10-08 and ended on 2020-10-07). This was done to decrease the chance of including users who quit going to the gym during the study duration. Lastly, we only included users who voluntarily filled out the Hexad user types questionnaire completely. To investigate the effects of personalization, we split participants into two groups. Since the gamification elements described in Section 5.4.1 were shown to be particularly suitable for Achievers, Players and Socializers [213, 342], users scoring highest on at least one of these traits were matched to a group of users who received a suitable set of gamification elements. This group was compared to the remaining users, for who the implemented gamification elements should not be particularly suitable, according to the Hexad model. Data was analyzed using paired/unpaired t-tests or the non-parametric counterpart, when assumptions were not met (determined by conducting Levene's/Shapiro-Wilk tests).

Hypotheses

We had the following hypotheses:

H1: The number of booked courses per participant is significantly higher in Gamification than in Baseline.

H2: Users scoring highest on the Hexad user types Achiever, Socializer or Player – and thus receive a suitable set of gamification elements – increase their number of booked courses significantly more than other users.

H1 is based on previous research demonstrating that gamification leads to an increased physical activity [10, 14, 74]. Thus, we expect to find similar effects. In contrast to previous work, the study duration is much longer in our study. Therefore, we expect to find a smaller effect than was reported in previous work, because novelty effects decrease over time. Since previous research conducting rather short studies has found medium to large effect sizes [14], we calculated an a-priori power analysis to detect a small to medium sized effect of dz=.40 with a power of 80%, thus revealing a minimum number of 41 participants. **H2** is based on findings of past research, showing that there are correlations between Hexad user types and the perception of gamification elements [13, 342]. Thus, we expect that these self-reported preferences should be reflected in the behavior of users, i.e. that receiving gamification elements that are suitable to the highest Hexad user types should have an effect on the behavior. Given the minimum number of 41 participants as calculated by the aforementioned power analysis, we are able to find large effects of d=.80 with a probability of 80%. **H1** is evaluated using paired tests while **H2** is evaluated using unpaired tests.

Dataset

We received the aggregated number of booked courses for each month and per study phase for each eligible user together with the information whether users scored highest on the Hexad factors Achiever, Player or Socializer. Besides that, we received the following aggregated information: the average scores of all Hexad factors, the average levels, the aggregated number of users who decided to provide a username and the average number of unlocked badges. We did not receive any non-aggregated data nor personal information such as age or gender such that the dataset can be considered fully anonymous. This was important to prevent any GDPR related issues. Due to the COVID-19 pandemic and the resulting lockdown in Germany, we had to exclude data between 2020-03-02 and 2020-06-01. To ensure the comparability to the Baseline phase and prevent seasonal effects, we excluded data from the same timespan in the Baseline, i.e. between 2019-03-02 and 2019-06-01.

Results

Overall, 52 eligible users were considered. The Hexad user types average scores are similar to the ones reported in the validation study of the Hexad questionnaire by Tondello et al. [337]. Philanthropists showed the highest average scores (M=23.79, SD=3.15), followed by Achievers (M=23.60, SD=3.33), Socializers (M=22.75, SD=3.62) and Free-Spirits (M=22.13, SD=3.55). Players (M=19.60, SD=5.34) and Disruptors (M=14.37, SD=4.87) followed with lower average scores. Based on the Hexad user type scores, our sample consisted of 33 users who received a suitable set of gamification elements (i.e. scoring highest on Achiever, Player or Socializer) and 19 users who did not receive a suitable set of gamification elements (i.e. who did not score highest on Achiever, Player or Socializer). The Achiever, Player and Socializer scores of users who received suitable gamification elements were on average significantly higher than the respective scores of users who did not receive a suitable set of gamification elements, with large effect sizes of $d>0.5$ (Achiever: $t(25.11)=2.25$, $p<0.05$, $d=0.74$; Player: $t(40.44)=1.88$, $p<0.05$, $d=0.53$, Socializer: $t(31.86)=2.40$, $p<0.05$, $d=0.73$). Users unlocked 4.60 (SD=1.80) badges and a substantial majority of 92% actively decided to be shown on the leaderboard by selecting a nickname.

Effect of Gamification on Course Bookings To analyze whether gamification had an effect on the number of booked courses (**H1**), we compared the number

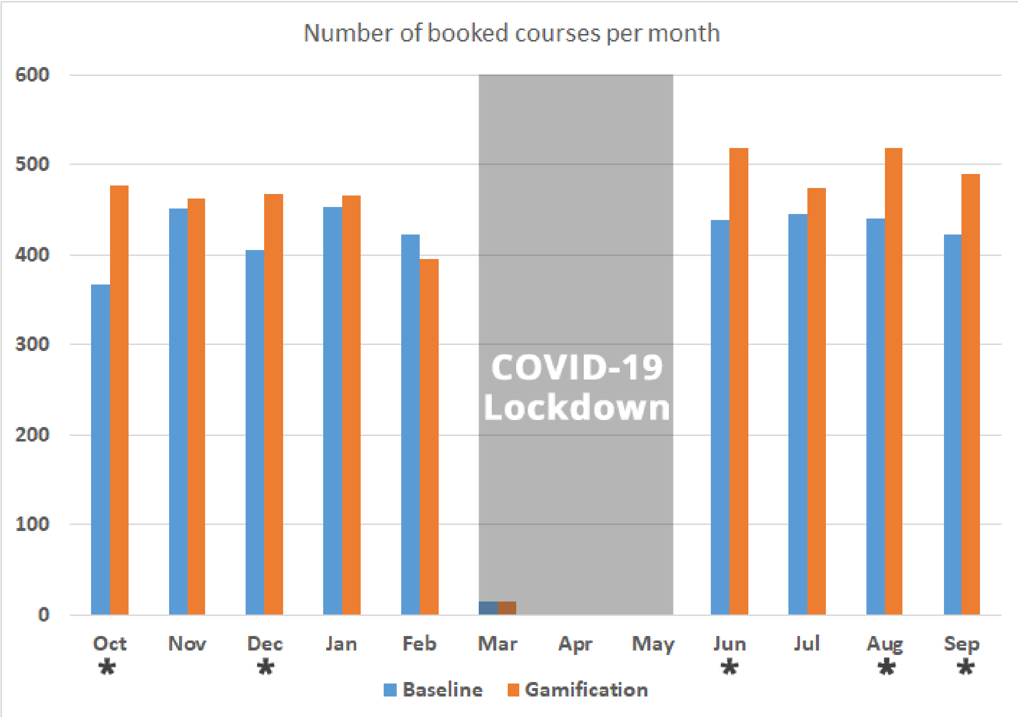


Figure 5.6: Number of booked courses per month. Blue bars (left) represent the number of bookings in Baseline, orange bars in Gamification. An asterisk represents a significant difference between Gamification and Baseline

of booked courses per day between Baseline and Gamification. In the Baseline phase, users booked 0.28 courses on average per day (Mdn=0.28, SD=0.14). This number significantly increased in the Gamification phase ($Z=491.00$, $p=0.036$, $d=0.31$) to an average of 0.30 courses per day (Mdn=0.29, SD=0.15). Thus, we derive result **R1: The number of booked courses per day is significantly higher in Gamification than in Baseline**.

In addition to comparing the full Baseline versus the full Gamification phase, we also compared the number of booked courses on a monthly basis between Baseline and Gamification phase. This was done to abstract from the fact that the time of the year might have a general influence on the behavior and motivation of users to participate in fitness courses in a gym. When looking at Figure 5.6, it can be seen that the number of booked courses is descriptively higher in all months (but February, which might be due to COVID-19) in the Gamification phase. Significant effects were found for five out of nine months (October: $t(51)=-3.38$, $p<0.01$, $d=0.47$; December: $t(51)=2.13$, $p<0.05$, $d=0.29$; June: $t(51)=1.95$, $p<0.05$, $d=0.27$; August: $t(51)=1.76$, $p<0.05$, $d=0.24$; September: $t(51)=1.68$, $p<0.05$, $d=0.23$). This is summarized as **R2: In the month-by-month comparison, the number of booked courses per day is significantly higher in Gamification for the majority of months**.

Effect of Personalization using Hexad User Types on Course Bookings Next, we analyzed if users scoring highest on the Achiever, Socializer or Player factor of the Hexad – and thus received a suitable set of gamification elements – were more driven by the gamification elements than other users. Therefore, we calculated the ratio between the average number of courses booked per day in Baseline and the average number of courses booked per day in Gamification for each user, to have a relative number indicating the difference in booked courses per day between Baseline and Gamification. This relative metric was chosen to abstract from the fact that users might have different subscriptions, which might introduce a bias to comparing the users who received a suitable set of gamification elements against users who did not. We then split users into a group that received suitable gamification elements (according to previous literature on correlations between Hexad user types and gamification elements, as described in Section 5.4.1; N=33) and a group that did not receive suitable gamification elements (N=19).

On average, users who did not receive a suitable set of gamification elements have a ratio of 1.01 (SD=0.29). This indicates that the number of booked courses remained almost the same in the Gamification phase. When considering users who received a suitable set of gamification elements, according to their Hexad user type, the average ratio is 1.31 (SD=0.90). This shows that these users increased their number of bookings in Gamification by more than 30% on average. When comparing this ratio between the users who received a suitable set of gamification elements and those who did not, we found a significant difference ($t(42.31)=1.78$, $p<0.05$, $d=0.41$). Thus, we derive **R3: Users who received a suitable set of gamification elements improved their participation during Gamification significantly more than others**. This suggests that personalization had an effect on the number of courses that users booked.

Since users who received a suitable set of gamification elements increased their number of course bookings significantly more than others, we also investigated whether the results that were found regarding the impact of gamification (**R1, R2**) still persist among users who did not receive suitable gamification elements as compared to those who did. Indeed, for users who did not receive a suitable set of gamification elements, none of the significant differences in course bookings reported in Figure 5.6 were found (i.e. there is no significant difference between Gamification and Baseline for users who did not receive suitable gamification elements), whereas the same significant differences where found for users who received a suitable set of gamification elements. Thus, we establish **R4: The significant differences in the month-by-month comparisons between Gamification and Baseline seem to be attributable to users who received a suitable set of gamification elements**. Furthermore, we analyzed whether there are differences in the amount of interactions with gamification elements. We found that more users who received a suitable set of gamification elements updated their profile (2.09 vs. 1.37 times), and wanted to show their name on leaderboards (94% vs. 89%), as compared to user who did not receive suitable gamification elements, without reaching significance. Regarding the amount of unlocked badges, we found a significant increase (4.94 vs. 4.00) among users receiving suitable gamification elements (t(33.38)=1.79, p=<0.05, d=0.54). These interaction-related results lead to **R5: User who received a suitable set of gamification elements unlocked more badges than others**.

5.4.3 Discussion and Limitations

The findings show that users increased their participation significantly during the year in which gamification was activated (**R1**). Also, when analyzing the number of booked courses per day on a monthly basis, it could be seen that the number of booked courses per day was higher in all months but February, with five out of nine months reaching significance (**R2**). This adds further to the fact that gamification affected users positively, apparently even in the long-run. We see **R1** and **R2** as supporting evidence for **H1: The number of booked courses per participant is significantly higher in Gamification than in Baseline**. On a more abstract level, these results contribute novel insights into the long-term effectiveness of gamification, which has been controversially discussed in the field [7, 185, 239].

We found that users who received a suitable set of gamification elements based on their Hexad type increased their participation in fitness courses significantly more than users who did not receive particularly suitable gamification elements (**R3**). In fact, users receiving suitable gamification elements increased their participation by more than 30%, while other users did not increase their participation considerably. Furthermore, when only considering users who did not receive a suitable set of gamification elements, the significant differences in the month-by-month comparisons between Gamification and Baseline disappear (**R4**). This together with **R3** suggests that the increased participation in Gamification (**H1**) might actually be caused by the group of users receiving suitable gamification elements, which undermines **H1** to a certain extent. Consequently, this could mean that the suitability of gamification elements plays a substantial role in the success of gamification. This poses the question, if the variety of positive, neutral or negative outcomes in previous literature [7, 146, 185] is due to the selection of suitable or unsuitable gamification elements. We also found that users for whom the gamification elements were suitable unlocked significantly more badges and interacted (descriptively) more with gamification-related features of the system (**R5**). **R3–R5** are important findings, since previous research has not considered behavioral data but solely focused on self-reported preferences, as far as we know. We see these results as supporting evidence for **H2: Users scoring highest on the Hexad user types AC, SO or PL—and thus receive a suitable set of gamification elements—increase their number of booked courses significantly more than other users**.

Limitations

Our study has several limitations which should be considered. First, we would like to acknowledge that the users we considered had to pay for every single course they booked, which limits the autonomy of their decision. Thus, it could be that the effect sizes we reported are different in contexts were users have a free choice of how much physical activity they would like to perform. Also, we selected participants who were participating in at least one course both in the last month of the Baseline as well as in the last month of the intervention phase. While this ensures that users who quit the gym due to external factors (such as changing the place of residence) are not considered in the sample, users who quit due to other reasons are also excluded. Therefore, future work should follow a study design which separates intervention and control groups, instead of doing a within-subjects study or ask users who quit about their reasons. In addition, it should be noted that our target group were users who already decided to visit the gym, which might have an impact on the success of gamification elements as reported in past research [13, 285]. Regarding **H2**, it should be considered that we used a dichotomous approach in deciding whether a certain user received suitable gamification elements or not, which was based on whether the users scored highest on the Achiever, Socializer or Player factor of the Hexad (because the gamification elements that we implemented were shown to be perceived particularly well among these user types). This has the advantage of an increased statistical power (due to less factors to differentiate), but comes at the cost of potential simplification (since the Hexad consists of six factors) and should be considered when interpreting our findings. Lastly, it should be noted that the Gamification phase was affected by the COVID-19 pandemic. Due to a nation-wide lockdown and the closure of the gym, we had to remove roughly three months from the Gamification phase. To account for this limitation and ensure the month-wise comparability of the data, we removed the corresponding days from the Baseline phase. However, we do not know in how far the pandemic has influenced the behavior of users. The fact that the number of booked courses was (descriptively) lower solely for the month February in the Gamification phase suggests that the COVID-19 pandemic already had an effect on the behavior of users in February 2020. Furthermore, we do not know whether the closure might have led to users booking courses more frequently when the gym was re-opened.

5.4.4 Contribution to Research Questions

The findings presented in this section show that gamification increases course participation, even over a longer time span. This contributes to gamification research, since past studies in the field had a rather short study duration, which did not make it possible to investigate whether the positive outcomes could be retained. Therefore, these findings provide answers to **RQ1**, since they show that gamification affects how people behave in the context of a course booking system, when using gamification elements like badges, points, and a leaderboard.

However, when investigating why gamification worked in this context by considering the individual Hexad type of users, we found that the increase in course participation can largely be attributed to users receiving suitable gamification elements according to their Hexad user type. This is an important finding: Without analyzing Hexad user types, we might have concluded that one-size-fits-all gamification was successful in this context, even though users receiving a suitable set of gamification elements was the main cause for the positive effect on booking behavior. This underlines the importance of better understanding how and why gamification works, and contributes answers to **RQ3** which strengthen the argument for considering personal factors – especially Hexad user types in this case – and tailoring gamified systems to the user.

5.5 Summary

This chapter mainly investigated what effects personalization, based on the factors investigated in Chapter 4, has on behavioral and psychological measures in gamified systems (**RQ3**). Overall, we conducted three studies – two in the lab and one in the wild – to contribute answers to this question.

In the first study, we implemented a gamified system encouraging physical activity on a treadmill to evaluate whether the findings regarding how Hexad user types and behavior change intentions affect the perception of gamification elements could be replicated when actually implementing a gamified system. Besides showing the general effectiveness of gamification and providing insights on how gamification affects motivation in this context, we found that both factors had an influence on measures related to affective experiences and intrinsic motivation. This provides initial support for the assumption that differences in the perception of gamification elements seem to translate into differences in motivation in implemented systems. However, we could not demonstrate that behavioral measures were affected when providing users gamification elements that were suitable for their respective stages of change or Hexad user types.

In the second study, we dynamically adjusted the set of gamification elements on an image tagging platform such that users received no gamification elements, or gamification elements which were suitable or unsuitable for their Hexad type, in three randomly ordered conditions. We first demonstrated that gamification, independent of personalization, increased task performance and affected psychological measures of arousal, contributing novel insights to the field. We were also able to find a broad range of effects, supporting evidence for the benefits of personalizing gamified systems. We found that enjoyment and positive affect were significantly higher in the condition in which suitable gamification elements were offered. Similarly, we found that flow experiences were more prevalent in this condition. However, in line with the previous study, we did not find supporting evidence that the differences in the motivation of users and their experience led to an increased performance.

In the third study, we aimed to better understand whether receiving gamification elements that were suitable for the users' Hexad type would have an impact on their behavior. We assumed that this would be the case, based on the findings of the two studies mentioned earlier. In those studies, we found that receiving suitable gamification elements affected psychological measures positively, which should affect behavioral outcomes in the long run. Therefore, we conducted an almost-two-year-long in-the-wild study in a gym. The results of this study supported our assumption: Participants who scored particularly high on Hexad user types which were shown to be positively linked to the perception of the gamification elements we used increased their number of course bookings significantly more than the others.

To sum up, our findings provide evidence for the benefits of personalizing gamified systems. They show that personalization is essential to increase both psychological and behavioral outcomes of gamification. In addition, our findings suggest that the results regarding the perceptual differences of gamification elements and their relationships to Hexad user types and behavior change intentions that we reported in Chapter 4, in which participants had to imagine how the gamification elements might look and feel, translate into actual differences in motivation and behavior when being implemented.

Chapter 6

Unobtrusive Ways to Facilitate Personalization in Gamified Systems

In the past two chapters, we demonstrated that accounting for inter-personal differences is important to increase motivation in gamified systems. In Chapter 4 we showed that the *perception* of gamification elements differs between users and identified factors such as age, behavior change intentions and Hexad user types to explain such differences. In Chapter 5, we demonstrated that accounting for these factors in implemented gamified systems has *actual effects* on behavioral and psychological measures. However, to allow utilizing factors like Hexad user types to personalize gamified systems in practice, it is crucial to provide ways of assessing these factors. While demographic factors like age can be assessed in an unobtrusive way by e.g. using machine learning methods and very short instruments are available to assess the stage of change of users [212], Hexad user types require more effort to assess. This is problematic, since our findings, and findings from related research in this domain (see Klock et al. [182] for an overview, and Section 2.4.3), have demonstrated the usefulness of the Hexad model for personalization purposes and showed that the model is superior over other factors for personalizing gamified systems [144]. Therefore, in this chapter, we focus on ways to facilitate the personalization of gamified systems based on Hexad user types in practice. In particular, we are interested in exploring and evaluating ways of predicting Hexad user types, without detrimentally affecting the gameful experience of a gamified system or breaking the immersion of users.

Moreover, using an implicit way of assessing Hexad user types could help to overcome issues being inherent to questionnaires. When users are asked to self-report answers in questionnaires, problems such as social desirability bias (participants answering in a socially desirable manner) and acquiescent responding (participants tending to agree with statements) were found [290]. Also, constraints on self-knowledge can lead to less accurate conclusions [290].

To overcome these issues and facilitate personalization in gamification practice, we first explore the feasibility of predicting Hexad user types based on smartphone data (such as which apps a user has installed and the communication patterns of users) in Section 6.2. Afterwards, we investigate whether Hexad user types can be assessed in a gameful way and whether the interaction behavior of users can be used to predict them in Section 6.3. The findings of both studies mainly contribute to **RQ4**, i.e. the question of how we can unobtrusively assess personal factors to personalize gamified systems without disturbing the gameful experience. Section 6.2 is based on [18] and Section 6.3 on [20].

6.1 Motivation

As we have seen in Section 2.4.3, the Hexad user types model was used to investigate user preferences in gamified systems across different contexts, including physical activity [13], education [236], energy conservation [189], health [261], and others. These investigations and the results of our studies show the usefulness of the Hexad user types model for personalizing gameful systems. The correlations between the perception of gamification elements and Hexad user types [13,342] enable dynamic adjustments to the elements of a gameful system, and our results from Chapter 5 demonstrate that gamified systems which use Hexad user types to personalize the gameful experience to the user are beneficial over non-tailored gamification in terms of psychological and behavioral measures.

However, determining the Hexad user type requires people to fill out a 24-item questionnaire. While this is appropriate in academic contexts so as to preserve the psychometric properties of the Hexad model and ensure scientific rigour, it may be disadvantageous when using the Hexad model to tailor gamified systems dynamically to their users in practice. Because gamified systems usually aim at providing an enjoyable and gameful user experience, requiring users to fill out a survey up front may break immersion, lead to frustration and thus detrimentally affect the overall user experience of a gamified system, as suggested by previous work in the context of gamified surveys [148,149,175,344]. Therefore, researching ways to tailor gamified systems without negatively affecting the user experience of a gameful system is important.

In this section, we will contribute to this by evaluating the feasibility of alternative approaches to determine Hexad user types without requiring users to fill out questionnaires. We aim at establishing statistical models to predict Hexad user types based on more subtle factors, which should not detrimentally affect the gameful experience of a system. While we first try to predict Hexad user types by using smartphone data, we investigate whether we can predict Hexad user types based on the interaction patterns of users and whether we can assess Hexad user types in a gameful way in the second study of this chapter.

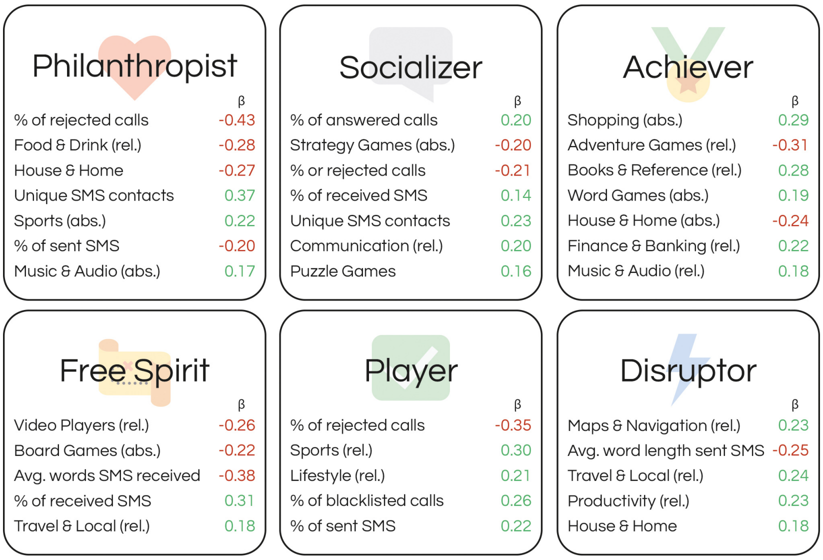


Figure 6.1: Regression models predicting user types from smartphone data. β **= standardized regression coefficient**

6.2 Towards Predicting Hexad User Types from Smartphone Data

In this section, we investigate whether smartphone data such as the types of installed apps and the communication behavior of users can be used to predict Hexad user types. This approach would allow us to assess Hexad user types in a subtle way, i.e. with the user's consent, the data needed to predict Hexad user types could be acquired in the background. This would not burden users with filling out a questionnaire and could be seamlessly integrated into existing gamified systems. The idea to predict Hexad user types based on smartphone data is inspired by the studies presented in Section 2.5, showing that smartphone data is linked to personality traits and that personality traits correlate with Hexad user types [342].

We developed a smartphone app gathering smartphone data and obtaining Hexad user types, which was used in a user study (N=122). We found regression models for each user type, indicating that deriving user types on the basis of smartphone data is promising (see Figure 6.1). In the following, the application, methodology and study results will be presented.

6.2.1 Concept and System Design

We developed a smartphone application to gather data, which is summarized and for which reasons are provided below. The application starts with a view in which the aim of the study is explained, which data will be collected, and on which participants are able to give consent to participate. Afterwards, participants are asked to fill out a survey.

Collecting Data

While filling out the survey, the application gathers the following smartphone data in the background:

- **Installed applications**: Package name, application name, install date, category of the app in the Google Play Store
- **Phone calls**: Average call duration, percentage distribution of types (initiated, answered, missed, rejected, blacklisted)
- **Short Message Service ("SMS") messages**: Average number of words for sent and received messages, average word length for sent and received messages, percentage distribution of sent and received messages
- **Contacts**: The number of unique message and call contacts

Attributes related to **communication data** (phone calls, SMS messages and contacts) were inspired by Chittaranjan et al. [80], who found these attributes to be related to the Big 5 personality traits. Similarly, installed applications were used because Lane et al. [192] and Seneviratne et al. [316] found these to be related to personality traits, too.

Displaying Study Results to Participants

Since we collect sensitive data, we aimed to make the data collection process as transparent as possible. Therefore, after filling out the survey and gathering smartphone data, the app visualizes all results in a graphical way using appropriate charts (see Figure 6.2). This is in line with [30], showing that users are unaware of what data is collected by smartphone apps and appreciate receiving feedback on that. First, our app shows participants the number of installed apps and their category distribution (cf. Figure 6.2a), followed by the duration and amount of their phone calls (cf. Figure 6.2b), the number of SMS messages they wrote and received (cf. Figure 6.2c) and the average length of written and received SMS messages. Afterwards, the distribution of the user type (cf. Figure 6.2d) and Big 5 scores is shown. We ensured that participants could only participate once and showed them their results once they had participated, whenever the app was opened.

Implementation

The smartphone application was realized as a native Android app (minimum API Level: 19). After giving consent to participate, the survey was shown while

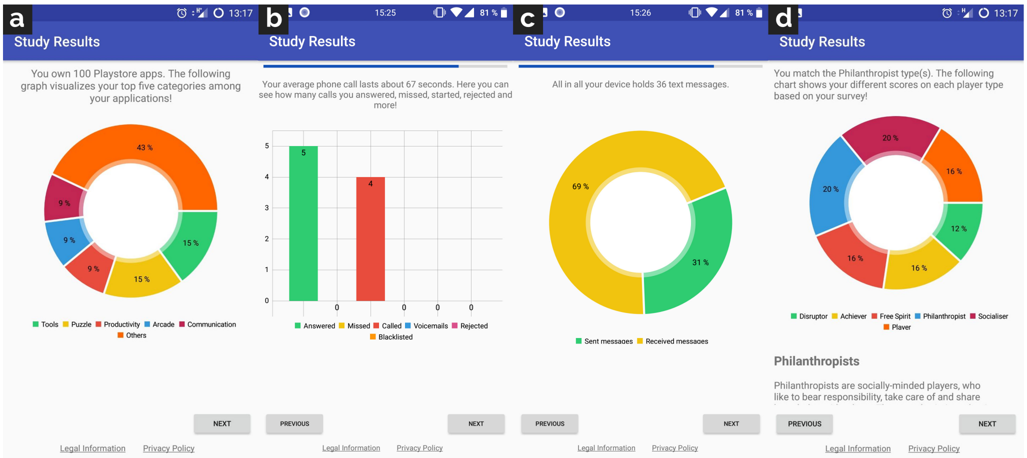


Figure 6.2: Prototype used to gather data and visualize study results to participants. a) Number of installed apps and their category distribution, b) duration and amount of phone calls, c) number of SMS messages written and received, d) user type score distribution

the required smartphone data was gathered in the background. To obtain a category for each installed smartphone app, the app crawled the Play Store (similar to [316]), offering 40 different categories overall. Once the data gathering process finished, all data was sent to a webserver, which stored them in a database. All data was completely anonymized as soon as it was sent to our server, i.e. it is impossible for us to relate the gathered data back to an individual person. We stored the Android ID, a 64-bit number which is unique to each combination of app-signing key, user and device to check whether participants already took part in the study (even re-installing the application would not change this ID). To ensure privacy, the Android ID was stored in a separate table and not linked to a specific participant.

6.2.2 Evaluation

Our study was approved by an institutional Ethical Review Board[33]. It was designed to investigate the relationships between smartphone data and Hexad user types. More specifically, we tried to find evidence for the following hypothesis:

Smartphone data can be used to predict the score of each of the six Hexad user types.

The hypothesis follows transitively from both studies showing that personality traits and smartphone data are correlated [80, 192, 269, 316] and studies showing that personality traits and Hexad user types are correlated [338, 342].

[33] Saarland University: *Ethical Review Board of the Faculty of Mathematics and Computer Science*, `https://bit.ly/2TK2Qii` (last accessed: 2021-12-01)

Method

Participants were asked to install our smartphone app from the Play Store. While the app gathered the above-mentioned smartphone data, they filled out a survey covering the following:

- Demographical data
- Hexad user type (four items per user type to be answered on 7-point Likert scales) [342]
- Big Five Inventory (BFI-10, ten items to be answered on 5-point Likert scales) [280]

Participants

Roughly half of the participants (63) were recruited through mailing lists and social media. The remaining participants (59) were recruited through Amazon Mechanical Turk ("AMT"), resulting in 122 participants in total (gender: 41.8% female; age <18: 2.5%, 18-24: 32%, 25-31: 29.5%, 32-38: 13.1%, 39-45: 11.5%, 46-52: 7.4%, 53-59: 4.1%). On AMT, we restricted the selection to US Turkers having an Android smartphone as their primary device. AMT Workers received $2 as compensation (the study took approximately 10 minutes to complete). The user types score distribution was found to be almost the same as in the validation study of the Hexad user type questionnaire [342]. Also, we found most of the correlations between user types and personality traits from [342]. The average scores for each user type (4 is the lowest, 28 the highest possible score) and personality trait (2 is the lowest, 10 the highest possible score) and correlations between user types and personality traits can be found in Table 6.1.

Smartphone Data

The average age of participants' smartphones was 506.11 (SD=393.09) days (difference between the time of participation and the first app installed). On average, participants had 73.11 (SD=35.45) apps installed, excluding system apps, i.e. apps that are pre-installed on the smartphone and respectively 276.30 (SD=100.00) apps including system apps. Furthermore, most of the installed apps belong to the "Tools" (20.63%) , "Productivity" (12.99%) and "Communications & Messaging" (11.29%) categories. The whole app distribution among categories along with descriptions, the number of users having at least one app belonging to each category and correlations to the Hexad user types can be found in Table 6.2. The most frequently installed apps were YouTube (99.18%), Google Maps (96.72%), Google Mail (94.26%), Google Hangouts (71.31%), Facebook (51.64%), WhatsApp (50.82%), Facebook Messenger (50.00%), Instagram (43.44%), Samsung Push Service (39.34%) and Spotify (38.52%).

	Mean	SD	OP	CO	EX	AG	NE
PH	23.16	3.89	.21**	.16*	.14*	.35**	
SO	19.76	5.40			.31**		
FS	22.63	3.11	.29**	.22**			-.27**
AC	21.94	3.86		.42**		.17*	-.20**
PL	21.36	4.52		.19**			-.13*
DI	14.81	5.07	.18**	-.17*		-.22**	
OP	7.75	1.94	1		.19**		
CO	7.36	2.02		1		.25**	-.29**
EX	5.91	2.41	.19**		1		-.17*
AG	6.80	2.06		.25**		1	-.16*
NE	5.34	2.28		-.29**	-.17*	-.16*	1

Table 6.1: Mean & standard deviation (SD) of the user types and the Big 5 and bivariate correlation coefficients (Kendall's τ). OP= Openness, CO= Conscientiousness, EX= Extraversion, AG= Agreeableness, NE=Neuroticism. * Significant at $p<.05$. ** Significant at $p<.01$.

Regarding communication data, participants had 65.43 (SD=58.93, Mdn=55.00) call and 30.07 (SD = 33.55, Md n= 18.50) message contacts. An average call lasted 3.33 minutes (SD = 2.77, Mdn = 2.52). Of all calls, 24.97% (SD = 13.06%, Mdn = 23.74%) were answered, 21.10% (SD = 15.37%, Mdn = 16.87%) were missed, 2.94% (SD = 5.98%, Mdn = 0.71%) were rejected, 50.10% (SD = 18.33%, Mdn = 50.00%) were outgoing, and 0.41% (SD = 2.04%, Mdn = 0.00%) were blacklisted. Moreover, 25.66% (SD = 19.21%, Mdn =2 7.93%) of all SMS messages were sent while 74.34% (SD = 19.21%, Mdn = 72.08%) were received. On average, received SMS messages contained 15.74 words (SD = 8.86, Mdn = 14.82) with a word length of 5.85 (SD = 4.33, Mdn = 5.33) while sent SMS messages contained 8.80 words (SD = 6.87, Mdn = 8.28) having a word length of 3.91 (SD = 2.16, Mdn = 4.36). Correlations between the aforementioned communication attributes and user types can be found in Table 6.3.

Predicting Hexad User Types

We investigate the hypothesis using stepwise multiple regressions to find a suitable model predicting the score for each user type (the score is between 4 and 28 for each user type). In stepwise multiple regression, a combination of forward selection and backward elimination is used. In the forward method, the highest simple correlated predictor is added to the model iteratively. Each time a predictor is added to the model, a removal test is made (backward elimination) to check whether redundant predictors can be removed [123]. It should be noted

Category	Description from Play Store	%Apps	#Users	PH	SO	FS	AC	PL	DI
Books & Reference	Wikis, book readers, dictionaries	2.14%	88						
Business	Document editor, remote desktop, email	1.23%	68						
Communications	Messaging, chat/IM, address books	11.29%	122		.14*/.13*			-.20**/-	
Education	Exam prep., educational games, vocabulary	1.76%	61						
Entertainment	Movies, TV, interactive entertainment	5.26%	117				- /.16*	- /.15*	
Finance	Banking, financial news, tip calculators	2.79%	80	- /-.14*			- /.15*	.18**/.24**	
Food & Drink	Food guides, wine tasting & discovery	0.67%	40	-.22**/-.22**				.14*/.18*	
Health & Fitness	Workout tracking, nutritional tips, health	2.59%	88			.14*/.15*		- / .17*	
House & Home	Home improvement, interior decoration	0.18%	15	-.15* / -				.16*/.15*	
Lifestyle	Style guides, party planning, how-to guides	2.41%	75					.16*/.19**	
Maps & Navigation	GPS, transit tools, public transportation	1.52%	72						.17**/.19**
Music & Audio	Music services, radios, music players	4.51%	122			- /.15*			
News	Newspapers, magazines, blogging	2.93%	101						
Personalization	Wallpapers, home screens, ringtones	1.90%	88	-.16*/ -					
Photography	photo editing, photo sharing	3.46%	117						
Productivity	Notepad, to do list, calculator	12.99%	122					- /.15*	
Shopping	Price comparison, product reviews, shopping	2.43%	78					- /.16*	
Social	Social networking, check-in	3.32%	111					- /.18**	
Sports	Sports news, fantasy team management	0.71%	27				.16*/.17*	.24**/.25**	
Tools	Tools for Android devices	20.63%	122					- /.14*	
Travel & Local	Ride sharing, city guides, local information	3.34%	122			- /.16*	- /.16*	- /.16*	
Video Players	Video players, video editors	3.85%	122			-.14*/-.15*			
Weather	Weather reports	0.57%	44						
Action Games	Action games	0.40%	22						
Adventure Games	Adventure games	0.39%	28				-.15*/-		
Arcade Games	Arcade games	0.74%	31			-.15*/-			
Board Games	Board games	0.29%	15						
Card Games	Card games	0.36%	21						
Casual Games	Casual games	0.77%	39						
Puzzle Games	Puzzle games	1.24%	44					-.19**/-.17*	
Role Playing Games	Role playing games	0.43%	21						
Simulation Games	Simulation games	0.44%	24		- /-.16*				
Strategy Games	Strategy games	0.33%	17						
Trivia Games	Trivia games (quizzes)	0.41%	27						
Word Games	Word games	0.35%	19						-.17*/-.17*

Table 6.2: Categories, their description based on Play Store guidelines, percentage share of apps for each category, number of users having at least one app from the category and bivariate correlation coefficients (Kendall's τ) of categories with Hexad user types. The number before "/" denotes a correlation between the *relative* amount of apps in a category and a user type, the number after "/" denotes a correlation between the *absolute* amount. * Significant at $p<.05$. ** Significant at $p<.01$.

Attribute	PH	SO	FS	AC	PL	DI
Avg. call duration				.16*		
Unique call cont.						
% of answered calls		.21**				
% of outgoing calls						
% of missed calls						
% of rejected calls	-.14*				.21**	
% of blacklisted calls						
% of sent SMS					.17**	
% of received SMS		.13*				
Unique SMS cont.		.16*		.19**	.19**	
Avg. W SMS rcv.		.13*				
Avg. WL SMS rcv.		.13*				.13*
Avg. W SMS sent						
Avg. WL SMS sent			-.13*			

Table 6.3: Bivariate correlation coefficients (Kendall's τ) of communication attributes with each user type of the Hexad model. * Significant at $p<.05$. ** Significant at $p<.01$. Cont=contacts, W=words, WL=word length, avg=Average, rcv=received

that this method was chosen since we did not have specific assumptions about which predictors are most relevant for each user type and because our goal was in the first place to explore the potential and feasibility of using smartphone data to infer user types implicitly. Given this kind of exploratory model building research, stepwise regression is a suitable method [123]. To reduce Type 1 errors, which might occur due to the multiple iterations performed by the stepwise method, the Benjamini-Hochberg false discovery rate [40] was used to adjust significance values for multiple comparisons. We chose this procedure as it maintains more statistical power and since it is more suitable for stepwise regressions and the increased amount of statistical tests than other correction methods like e.g. the Bonferroni method [40].

As potential predictors, the absolute number of installed apps for each Play Store category and participant, the relative number of installed apps per participant for each Play Store category (the absolute number divided by the total number of installed apps) and communication data (phone calls, SMS messages and contacts) were entered into each model. We decided to include the absolute and the relative amount of apps per category to reflect both personal preferences and overall app distribution. Moreover, categories for which less than 15 unique users had at least one app, were excluded (cf. [123]). Table 6.2 and Table 6.3 summarize all attributes that were initially included in the stepwise regression.

For all multiple regression analysis, the assumptions of normality, linearity, homoscedasticity and independent errors were met (checked using histograms, ZPRED vs. ZRESID plots, P-P plots and Durbin-Watson tests which were all between 1.61 and 2.16). Additionally, indicators of multicollinearity (tolerance and variance inflation factors) were within acceptable limits (for all tolerance>0.6; VIF<2) [123].

In the following paragraphs, we provide the results of the multiple regression analysis and the correlations we have found between Play Store categories and Hexad user types. Based on these results, we discuss potential explanations for each predictor and correlation we have found.

Predictor	**b**	**SE B**	β	$p_{adj.}$
% of rejected calls	-0.16	0.03	-0.43	0.000
Food & Drink (rel.)	-1.27	0.33	-0.28	0.000
House & Home (rel.)	-2.47	0.66	-0.27	0.000
Unique SMS contacts	0.04	0.01	0.37	0.000
Sports (abs.)	0.62	0.21	0.22	0.004
% of sent SMS	-0.04	0.16	-0.20	0.021
Music & Audio (abs.)	0.34	0.15	0.17	0.021

Table 6.4: Predictors for the Philanthropist with beta values, standard error (SE), the standardized beta values (β) and the adjusted p-value

Philanthropists We found a significant regression equation to predict the score of the "Philanthropist" user type scale (R=.67, R^2 adjusted=.41, F(7,114)=13.03, p<.000). As illustrated in Table 6.4, the percentage share of rejected calls, the relative amount of apps in the "Food & Drink" and "House & Home" category, the number of unique SMS contacts, the number of applications in the "Sports" category, the percentage share of sent SMS messages and the number of apps in the "Music & Audio" category all significantly predict the score on the "Philanthropist" scale of the Hexad model.

The negative influence of the percentage share of rejected calls and the positive influence of the number of SMS contacts might be well explained by the basic social attitude of the Philanthropist. Taking into account that Philanthropists are socially-minded but not primarily interested in initiating social interaction, the negative influence of the percentage of sent SMS messages also seems reasonable. Furthermore, the positive influence of "Sports" apps relates well to the preference of Philanthropists for administrative roles, as many fantasy team management apps belong to this category. However, the influence of "Music & Audio", "Food & Drink" and "House & Home" apps is not directly explainable by the definition and motivational factors of the Philanthrophist user type.

Moreover, we also found several correlations to smartphone data, which can be found in Table 6.2 Table 6.3. In addition to the percentage share of rejected calls, the "Food & Drink" and "House & Home" category, negative correlations to the "Finance" and "Personalization" categories were found. For both a potential reason is the lack of meaning: Philanthropists strive for meaningful goals and purpose, which they might not find in monetary things or by changing their wallpapers, home screens or ringtones.

Predictor	**b**	**SE B**	β	$p_{adj.}$
% of answered calls	0.08	0.03	0.20	0.032
Strategy Games (abs.)	-1.52	0.63	-0.20	0.032
% of rejected calls	-0.11	0.04	-0.21	0.032
% of received SMS	0.02	0.01	0.14	0.076
Unique SMS contacts	0.03	0.01	0.23	0.032
Communication (rel.)	0.27	0.12	0.20	0.032
Puzzle Games (rel.)	0.40	0.20	0.16	0.057

Table 6.5: Predictors for the Socializer with beta values, standard error (SE), the standardized beta values (β) and the adjusted p-value

Socializers For the score on the Socializer scale, we found a significant regression equation (R=.55, R^2 adjusted=.26, F(7,114)=7.07, p<.000) including seven predictors as can be seen in Table 6.5. While the percentage share of answered calls, received SMS messages, unique SMS contacts and the relative amount of puzzle games and communication apps positively influence the score, the percentage share of rejected calls and the absolute number of strategy games have a negative influence. Given that relatedness is the most important motivational factor, the positive impact of answered calls, received SMS messages, unique SMS contacts and the relative number of communication apps together with the negative influence of rejected calls is not surprising and fits the motivational aspects of the Socializer user type very well. However, the potential reasons for why puzzle games positively and strategy games negatively influence the score are not so obvious and do not directly fit the characteristics of this user type.

While we also found positive correlations to communication data (see Table 6.3) and the "Communications" category, a negative correlation was found to "Simulation Games" (see Table 6.2), which might be explainable by the extraordinarily low social interaction in this kind of game (this category mainly includes games like "Sim City" or "Car Mechanic Simulator").

Predictor	b	SE B	β	$p_{adj.}$
Video Players (rel.)	-0.42	0.14	-0.26	0.005
Board Games (abs.)	-0.93	0.36	-0.22	0.014
Avg. words SMS received	-0.12	0.03	-0.38	0.000
% of received SMS	0.03	0.01	0.31	0.005
Travel & Local (rel.)	0.28	0.13	0.18	0.039

Table 6.6: Predictors for the Free Spirit with beta values, standard error (SE), the standardized beta values (β) and the adjusted p-value

Free Spirits For the "Free Spirit" user type scale, we found a regression model consisting of five predictors (R=.48, R^2 adjusted=.19, F(5,116)=6.80, p<.000), which are shown in Table 6.6. While the preference for "Travel & Local" apps might be well explainable by the need to explore and discover, which is likely satisfied when traveling, the negative impact of board games might be related to the fact that board games usually have a fixed rule-set and thus potentially compromise the need for autonomy. However, why Free Spirits seem to receive many SMS messages with a low number of average words and why the relative amount of apps in the "Video Players" category negatively influence the score on the Free Spirits scale is not directly explained by the characteristics of this user type.

All positive correlations between app categories and the Free Spirit user type ("Health & Fitness", "Music & Audio", "Travel & Local", see Table 6.2) have in common that these categories usually allow users to explore new things like new exercises or healthy recipes, new songs and artists or new destinations for traveling. Furthermore, the negative correlation to "Arcade Games" might be explainable by the usually very restricted number of ways to play arcade games and thus low autonomy. As stated above, the reason for the negative correlation to the "Video Players" category is not as obvious.

Achievers Seven predictors (see Table 6.7) predict the "Achiever" user type scale (R=.51, R^2 adjusted=.22, F(7,114)=5.83, p<.000). The "Shopping", "Books & Reference" and "Finance" category have in common that apps in this category often convey competence (price comparison and product reviews, "Wikis" and dictionaries, financial news and tip calculators; cf. Table 6.2). Also, "Word Games" (e.g. "Scrabble") build on competence and often require players to overcome challenges demanding mental abilities, which relates well to the characteristics of Achievers. However, the negative influence of "Adventure Games" and "House & Home" is not directly explainable by the specific needs of Achievers.

In addition to these significant predictors, several correlations were found between the "Achiever" user type and the Play Store categories (see Table 6.2). Besides "Finance" and "Adventure Games", correlations were found for "En-

Predictor	b	SE B	β	$p_{adj.}$
Shopping (abs.)	0.47	0.14	0.29	0.002
Adventure Games (rel.)	-1.55	0.42	-0.31	0.000
Books & Reference (rel.)	0.53	0.16	0.28	0.002
Word Games (abs.)	1.08	0.48	0.19	0.032
House & Home (abs.)	-2.56	0.91	-0.24	0.009
Finance (rel.)	0.32	0.13	0.22	0.025
Music & Audio (rel.)	0.36	0.17	0.18	0.034

Table 6.7: Predictors for the Achiever with beta values, standard error (SE), the standardized beta values (β) and the adjusted p-value

tertainment", "Sports" and "Travel & Local" (all positive). While the latter two categories might satisfy the need for competence in a way (sports news, fantasy team management games and hotel/trip comparisons), the "Entertainment" category does not as strongly support the need for competence. We also found positive correlations to the average call duration and to the number of SMS contacts (see Table 6.3), which might be explainable by the positive correlation between the Achiever and the Socializer user type [342].

Predictor	b	SE B	β	$p_{adj.}$
% of rejected calls	-0.15	0.04	-0.35	0.000
Sports (rel.)	0.93	0.25	0.30	0.000
Lifestyle (rel.)	0.42	0.18	0.21	0.018
% of blacklisted calls	0.61	0.19	0.26	0.003
% of sent SMS	0.05	0.02	0.22	0.018

Table 6.8: Predictors for the Player with beta values, standard error (SE), the standardized beta values (β) and the adjusted p-value

Players Also for the "Player" user type scale, a significant regression equation was found (R=.51, R^2 adjusted=.26, F(5,116)=8.11, p<.000). As illustrated in Table 6.8, the percentage share of rejected calls has a negative influence, while the relative amount of "Sports" and "Lifestyle" apps together with the percentage share of sent SMS messages and blacklisted calls positively influence the score. While the positive influence of rejected calls might be explainable by the tendency of players to take care of their own needs, the negative influence of rejected calls and the positive impact of sent SMS messages seem contrary. However, the strong positive correlation between the Player and Socializer, which was shown in [342], might explain these findings. Also, the positive impact of "Sports" apps might be explainable by the strong correlation between the Player and the Achiever [342], as this predictor was also found for the Achiever user type. However, no clear potential explanation can be given for the "Lifestyle" category.

The highest number of correlations was found between the Player and Play Store categories (see Table 6.2 and Table 6.3). Besides the ones that were also found as predictors in the regression model above, we found that "Entertainment", "Finance", "Food & Drink", "Health & Fitness", "House & Home", "Productivity", "Shopping", "Social", "Tools" and "Travel & Local" are all positively correlated. Moreover, we found negative correlations to the "Communications" and "Puzzle Games" categories. Finding the highest number of correlations for the Player type is explainable by previous work [342], showing that the Player type is the one showing the most correlations to other user types (i.e. it is correlated with the Socializer, Free Spirit and Achiever). This is also reflected in the high overlap of correlations to Play Store categories to these three user types. Interestingly, the "Social" category is positively correlated while the "Communication" category shows a negative correlation. This might indicate that Players seek attention (which might be rewarding) in social networks, but are not primarily interested in communicating with others, i.e. the "communication" might be rather one-directional. This would also possibly explain why blacklisted calls positively influence the user type score, while the percentage share of sent SMS messages also has a positive influence.

Predictor	**b**	**SE B**	β	$\mathbf{p_{adj.}}$
Maps & Navigation (rel.)	0.76	0.29	0.23	0.011
Avg. word length sent SMS	-0.52	0.17	-0.25	0.011
Travel & Local (rel.)	0.62	0.22	0.24	0.011
Productivity (rel.)	0.26	0.10	0.23	0.011
House & Home (abs.)	2.54	1.20	0.18	0.036

Table 6.9: Predictors for the Disruptor with beta values, standard error (SE), the standardized beta values (β) and the adjusted p-value

Disruptors The Disruptor score was found to be predictable using five variables (R=.46, R^2 adjusted=.18, F(5,116)=6.29, p<.000). As illustrated in Table 6.9, the relative number of apps in the "Maps & Navigation", "Travel & Local" and "Productivity" categories and the absolute number of apps in the "House & Home" category positively influence the score, while the average word length in outgoing SMS messages has a negative impact. Considering that autonomy and creativity are also important motivators for Disruptors [342], the positive influence of "Maps & Navigation" and "Travel & Local" categories is not surprising as they both relate well to the need to explore and discover. The positive influence of the "House & Home" category also relates well to the importance of creativity as this category deals with apps about interior decoration and home improvement (cf. Table 6.2). However, the positive influence of the “Productivity” category and the negative impact of the average word length in sent SMS messages are not directly explainable via the main motivations of Disruptors.

In addition to these predictors, we found two correlations to the Play Store categories (see Table 6.2), of which one (“Word Games”) is not included as a predictor. The negative correlation to the “Word Games” category seems to make sense, as word games are typically challenging and build on the need for competence rather than autonomy or creativity (following a clear rule-set and demanding mental effort to win).

6.2.3 Discussion and Limitations

The overall goal was to explore whether smartphone data can be used to infer the user type scores of users to tailor gamified systems implicitly, i.e. without the burden of filling out questionnaires. Summing up, we found regression equations that can be used to predict the score of each of the six user types of the Hexad model. This is supported by the amount of variance these models explain. All models we have found explain between 18% and 41% of the variance, showing that we found medium [86] (explaining at least 13% percent of the variance in the model: Free Spirits, Achievers, Disruptors) to large [86] (explaining at least 26% of the variance in the model: Players, Socializers, Philanthropists) effect sizes. Moreover, we found several significant correlations between smartphone data and user types. Therefore, our study results suggest that inferring user types from smartphone data is feasible to a certain extent and could be used to tailor gamified systems automatically (supporting evidence for our hypothesis). This finding is relevant for gamified smartphone applications that could adapt their game elements without the need for explicit user input and thus could motivate their users more effectively. However, our results are also relevant for gamified systems in general, as users could also provide their smartphone data to these systems in an automated way, which allows tailoring even such systems without manual user effort.

Our descriptive findings are in line with previous research: The user type distribution is nearly exactly the same as the distribution in the paper by Tondello et al. [342]. Also, the majority of correlations between the Hexad user types and the Big 5 personality traits are in line with findings from [342]. Furthermore, participants had a similar app category distribution as in [316]. This suggests that our sample was comparable in terms of user type distribution and app categories.

Even though we were not able to explain all predictors and correlations found, overall the most important motivational factors of each user type were reflected in the corresponding model (e.g. “relatedness” was reflected by positive impacts of communication data for Socializers, “autonomy” by the positive impact of exploration and creativity applications for Free Spirits etc.). On a meta-level, this suggests that preferences for smartphone application categories and smartphone communication behavior could explain the personal importance of motivational needs (as defined by the SDT [298]), which might be a relevant result also outside of the gamification domain.

Limitations

Although using stepwise regression is a suitable method for exploratory model building [123], the method itself has two main issues that should be considered. First, it is prone to model selection bias, resulting from including explanatory variables because of significant F statistics, which might in reality have no (or very weak) relationships to the response variable ("Freedman's paradox") [207]. This can lead to overestimations of the importance of certain predictors, which should be considered. As a consequence, the models we have found are not necessarily the only possible models, nor the best ones. Second, stepwise regression involves a large number of tests, inflating the probability for Type 1 errors. However, we considered this by adjusting our alpha thresholds accordingly.

In addition, roughly half of our participants were recruited from AMT. Even though we required an Android phone to be the primary device, we cannot guarantee that actual personal smartphones were used, rather than "dummy" smartphones, to earn the financial reward. However, since the Play Store category distribution of our sample was in line with previous research [316] and because we did not find significant differences between participants from AMT and the rest of our sample regarding the number of apps, we think that using AMT did not affect our results. On the contrary, we argue that using AMT was advantageous in terms of sample diversity.

Lastly, it should be considered that Play Store categories are assigned by the publishers of smartphone applications. Even though they should be interested in assigning a suitable category, a certain amount of fuzziness is unavoidable.

6.2.4 Contribution to Research Questions

Although it should be noted that the models require validation and that the statistical procedure to analyze our data was rather exploratory, we showed the general feasibility of predicting Hexad user types based on smartphone data in this study. The fact that there seem to be relationships between smartphone data and Hexad user types, which can be explained well by the characteristics of each user type, provides support for our approach of predicting Hexad user types based on smartphone data, and may serve as a way of facilitating personalization of gamified systems in practice, contributing answers to **RQ4**.

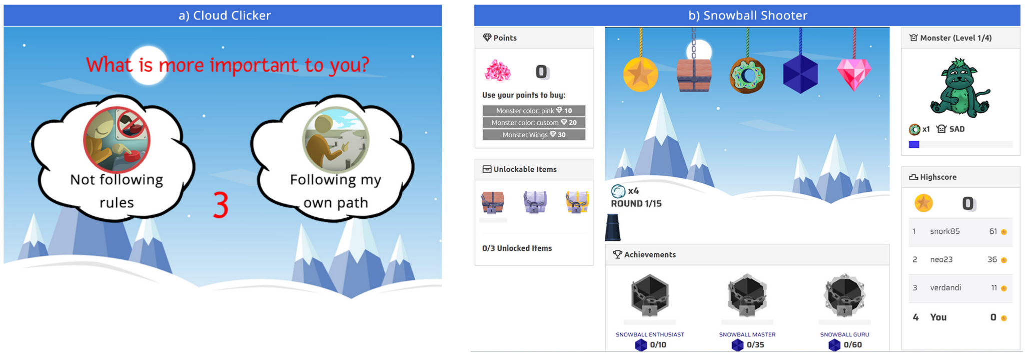


Figure 6.3: Gameful applications used in the user study. Cloud Clicker (a) asks users to decide which statement is more important to them. Snowball Shooter (b) provides several gamification elements that users can interact with.

6.3 HexArcade: Predicting Hexad User Types By Using Gameful Applications

Here, instead of predicting Hexad user types based on smartphone data in the background, we investigate whether Hexad user types can be assessed in a gameful way and whether they can be predicted based on how users interact with a gameful application. Since the regression models we established in the previous section leave room for improvement, we want to investigate whether using gamified applications to predict Hexad user types and explicitly tracking how users interact with such applications increases the variance explained. To do so, we conceptualized and implemented two gameful applications – Cloud Clicker and Snowball Shooter – which will be explained in the following. While Cloud Clicker focuses more on assessing Hexad user types in a gameful way by turning the 24-item Hexad questionnaire into a gameful application inspired by popular cookie clicker games, Snowball Shooter allows users to interact freely with several gamification elements and builds on the assumption that users scoring high on a particular Hexad type will interact more frequently with corresponding gamification elements.

We present findings from an online study in which these two gameful applications were investigated. Besides analyzing whether we could use these applications to predict or assess Hexad user types, we were also interested in understanding whether they were enjoyable and immersive. We compared these measures with assessing Hexad user types using the validated questionnaire to better understand whether assessing Hexad user types in a gameful way has advantages over the questionnaire in terms of the user experience. Lastly, we discuss the findings regarding the prediction of Hexad user types and the user experience of the gameful applications in comparison to using the validated questionnaire.

6.3.1 Gameful Applications

We implemented two gameful web applications using AngularJS[34] and Phaser[35]. In the following, the concepts behind both gameful applications are explained and discussed. For both applications, interactive step-by-step tutorials were developed, explaining how to interact with them. We used the systematic literature review by Keusch et al. [175] to ensure that most relevant papers on gamified surveys have been considered to inform the design of the gameful applications.

Gameful Application 1: Cloud Clicker

Similar to Triantoro et al. [344], we decided to transfer the 7-point Likert scales, which the Hexad questionnaire uses into binary choice questions in the gameful application. The whole design of the first gameful application ("Cloud Clicker") followed the process proposed by Harms et al. [148, 150] and considered recommendations and lessons learned from relevant previous work [148, 149, 175, 344], as described in the following. In Cloud Clicker, users see two statements in each of 15 rounds and then have to decide which of these two statements is more relevant to them (see Figure 6.3a). We selected one statement for each Hexad user type, which grasps its underlying motivation. We based the statements on the definitions given by Tondello et al. [342] and on items with a substantial factor load in the confirmatory factor analysis in the validation study of the Hexad user types questionnaire [337]. This follows the same procedure as Triantoro et al. [344] proposed to translate the Big-5 survey into a gamified counterpart.

Next, as part of the "aesthetics and relationship" layer of the design process for gamified surveys by Harms et al. [148, 150], we decided to present the two statements shown to the user using cloud visualizations to create visual sensation [148] and a convincing and motivational environment [175]. To enhance questions and support the comprehensibility of the statements [148, 175], we created visual illustrations for each statement, explaining the statement by using a gender-neutral avatar (Figure 6.4). For the Philanthropist, we decided to focus on the aspect of helping others because the statements *It makes me happy if I am able to help others* and *I like helping others to orient themselves in new situations* had the highest factor load for this type [337]. For Socializers, we focused on being part of a team because the corresponding statement *I like being part of a team* had the highest factor load among this trait [337].

Similarly, we used the statements having the highest factor load in a trait for the Free Spirit (*It is important to follow my own path*), the Achiever (*I like overcoming obstacles*) and the Player (*Rewards are a great way to motivate me*) [337]. Because the item having the highest factor load for the Disruptor type (*I see myself as a*

[34] AngularJS: *Superheroic JavaScript MVW Framework*, `https://angularjs.org/` (last accessed: 2021-12-01)
[35] Phaser: *A fast, fun and free open source HTML5 game framework*, `https://phaser.io/` (last accessed: 2021-12-01)

Figure 6.4: Illustrations and statements used in Cloud Clicker.

rebel) was hard to illustrate visually, we decided to use the item *I dislike following rules* instead, which had the second highest factor load [337]. Similar to Hallifax et al. [144], statements were presented to users using a full paired-comparison design (i.e., each user was asked to evaluate all possible pairs of statements, resulting in 15 rounds of comparisons in Cloud Clicker). The order of the statements as well as whether a cloud with a statement was shown on the left or on the right side of the screen in a round was randomized to avoid biasing results [148]. Cloud Clicker provides a ranking of the statements, in which scores of 0–5 are distributed across each statement representing its corresponding user type, because each statement is compared to every other statement.

Clouds were shaking and dropping coins when being clicked (i.e., when a user decided on a particular statement) to increase the gameful experience of Cloud Clicker. Also, we added sound effects to indicate interactions with the gameful applications. Both follow recommendations by Harms et al. [148], stating that gameful feedback in surveys should be provided by using indicators such as coins as rewards and supported by using auditory feedback. The coins were colour-coded to represent the corresponding Hexad user type and were showing a miniature version of the illustration of the related user type. Similar to Triantoro et al. [344], we introduced time pressure when participants were asked to decide between two statements in each round. This mechanic had three reasons: First, it emphasizes and stimulates the gameful experience of the application [344]. Second, it supports spontaneous responses, which was shown to increase the reliability of responses [247]. Third, it limits the time it takes to complete the application and thus to assess the Hexad user type, which might be important as we aim to provide a practical way of assessing Hexad user types in gamified systems. To allow researchers and practitioners to use Cloud Clicker, we published the source code as well as all graphical assets on GitHub[36].

Gameful Application 2: Snowball Shooter

Cloud Clicker aims at providing a gameful way of assessing Hexad user types and, thus, builds on the original items from the Hexad questionnaire [337]. In contrast, Snowball Shooter focuses on user behaviour when interacting with gameful elements and whether it is possible to use this input to predict Hexad user types. As such, Snowball Shooter provides some gamification elements that users can interact with (see Figure 6.3b): The user controls a snowball cannon and shoots snowballs at items representing each gamification element. These items are randomly positioned in each round. Shooting at an item increases the internal score of the corresponding gamification element. Similar to Cloud Clicker, the application consists of 15 rounds in which users may shoot five snowballs. It uses feedback sounds when shooting.

Following design suggestions from Harms et al. [148], we integrated progression loops in each gamification element. Consequently, there are three score thresholds, which lead to a state change of the corresponding element (e.g., unlocking a virtual item). To ensure comparability, these thresholds were kept constant across all gamification elements (given that a maximum score of 75 can be reached, the first state change happens at a score of 10, the second at 35 and the third at 60). As a result, it is impossible to complete all gamification elements. This was explained to users in the tutorial. It is important to note that – in contrast to previous work – users could experience the gamification elements instead of being given storyboards or textual descriptions only. Based on the user type descriptions by Marczewski [213] and the proposed gamification elements by Tondello et al. [342], we integrated the following gamification elements (Figure 6.3b):

Unlockables: Unlockables are expected to motivate **Free Spirits** because they are mainly driven by autonomy [213, 342]. In Snowball Shooter, we decided to provide treasure chests which could be unlocked to obtain virtual items. Reflecting the score thresholds, there are three different types of treasure chests (wooden, silver, golden) unlocking items of different rarity (common, rare, epic; see Figure 6.5).

Achievements: This element was shown to be especially suitable for **Achievers** as it supports mastery [213, 342]. In Snowball Shooter, three Achievements (using the score thresholds mentioned before) can be unlocked: "Snowball Enthusiast" (bronze frame), "Snowball Master" (silver frame) and "Snowball Guru" (golden frame), see Figure 6.5.

[36] GitHub: *m-altmeyer/cloud-clicker*, `https://bit.ly/3zjaDUo` (last accessed: 2021-12-01)

Figure 6.5: Visualizations of the score levels for Unlockables, Achievements and Virtual Character in Snowball Shooter.

Points: Points have been shown to positively affect **Players** [213, 342]. To underline the value of points as virtual currency (which is important for Players [342]), points can be used to buy modifications for the Virtual Character (see below). The amount of points that needs to be spent to buy all modifications (change or customize the colour of and add wings for the virtual character, see Figure 6.3b) equals the maximum score threshold described above.

Leaderboard: Social gamification elements such as Leaderboards are relevant for **Socializers** [213, 342]. However, findings by Tondello et al. [342] and other researchers [13, 189, 261] consistently demonstrate that Leaderboards are also positively correlated to **Players**, **Achievers** and **Disruptors**, which is why we expect to also find such correlations in Snowball Shooter. Similar to Mekler et al. [224], we decided to show fictitious users, having scores based on the thresholds established before, to ensure that all participants have same chances to rise in ranks.

Virtual Character: **Philanthropists** are driven by purpose and like to care for others [213, 342]. Although no significant correlations have been shown, we expect that this gamification element should be particularly relevant for Philanthropists because it may induce feelings of care-taking. We used a virtual monster whose emotional state is coupled to the number of snowballs shot at the respective item, a green doughnut. The three changes in its emotional state (Figure 6.5) are coupled to the thresholds described before.

6.3.2 Evaluation

We used the aforementioned gameful applications to investigate the following hypotheses stemming from our review of the related literature:

H1: Gameful applications can be used to predict Hexad types.

- **H1a**: The score of the statements in Cloud Clicker is correlated to the corresponding Hexad user types and thus may be used to predict them.
- **H1b**: The amount of interactions with gameful elements in Snowball Shooter is correlated to the corresponding Hexad user types and thus may be used to predict them.

H2: The users' perception of the gameful applications differs compared to their perception of the Hexad questionnaire.

- **H2a**: Both applications are perceived as more enjoyable (measured by the IMI enjoyment subscale) than the Hexad questionnaire.
- **H2b**: Participants feel more competent (measured by the IMI competence subscale) using both applications than using the Hexad questionnaire.
- **H2c**: Participants feel more pressure (measured by the IMI pressure subscale) in both applications than in the Hexad questionnaire.
- **H2d**: Both applications are perceived as more immersive (measured by the PXI immersion subscale) than the Hexad questionnaire.

H1 is motivated by previous work showing that questionnaires can be transformed into gameful applications without heavily affecting their validity [344]. Triantoro et al. [344] demonstrated that the Big-5 personality traits can be predicted based on gameful, binary choices in their survey, which is similar to our approach and thus motivates **H1a**. The subjective assessments of preferences for gamification elements using textual descriptions [342] or storyboards [13, 261] are correlated to the Hexad user types, which motivated us to find similar correlations when investigating actual interaction with implemented gamification elements (**H1b**).

H2a is mainly based on previous work in the domain of gamified surveys [148, 149, 344], where positive effects on enjoyment-related measures have been demonstrated. **H2b** relates back to feedback provided by gamification elements having been shown to increase perceived competence [303]. We expect to see an increase in perceived pressure in the gameful applications mainly because of the time pressure that is induced by Cloud Clicker and because of the round-based nature of both applications (**H2c**). An increase in pressure does not necessarily affect user experience negatively but might help to shape optimally challenging systems [148]. This supports users in reaching a flow state, which is described as a state of increased concentration and enjoyment [94]. To better understand whether the perceived pressure related to feelings of flow and immersion, we also evaluated immersion as part of the PXI questionnaire and expected it to be higher in the gameful applications (**H2d**).

Procedure

We conducted an online study on Prolific[37], an online platform specifically targeted at recruiting participants for scientific research studies. The only requirement was an understanding of the English language. The study has been reviewed and received ethics clearance through a University of Waterloo Research Ethics Committee (ORE#41608). It took approximately 15–20 minutes to complete and participants were paid £2 GBP. After giving informed consent, they were asked to provide demographic data including age and gender. Next, the 24-item Hexad user types questionnaire [337] was administered. The questionnaire consists of four items for each of the six user types, being measured on 7-point scales. To obtain a baseline for how participants perceived filling out the Hexad questionnaire, they were asked to fill out the 22-item task evaluation questionnaire of the Intrinsic Motivation Inventory ("IMI") [216, 297] as well as the "Immersion" subscale of the Player Experience Inventory ("PXI") [1]. Both the IMI items and the PXI items are measured on 7-point scales.

Next, participants were asked to interact with the gameful applications. The order of the gameful applications was randomized. Before starting the actual application, participants had to complete a tutorial explaining how to interact with them. In Cloud Clicker, we measured how often participants chose each statement. Similarly, the number of interactions with each gamification element in Snowball Shooter was measured. After interacting with each gameful application, participants were asked to fill out the IMI questionnaire and the PXI "Immersion" subscale.

Participants

After removing participants who preferred not to answer questions of the Hexad questionnaire, 147 participants were considered for the analysis. Of those, 49% self-reported their gender as female, 49% as male, 0.7% as non-binary and 1.3% preferred not to answer this question. The mean age was 33 years (SD=11.5,

[37] Prolific: *Quickly find research participants you can trust*, `https://www.prolific.co/` (last accessed: 2021-12-01)

Mdn=30, Min=18, Max=66). The Hexad user types average scores are similar to the averages reported in the validation study of the Hexad questionnaire by Tondello et al. [337]. Achievers showed the highest average scores (M=23.6, SD=2.98), followed by Philanthropists (M=22.8, SD=3.18), Players (M=22.8, SD=3.53) and Free Spirits (M=22.3, SD=3.52). Socializers (M=18.7, SD=4.89) and Disruptors (M=15.0, SD=4.49) followed with lower average scores.

Results

In this section, we present results related to predicting Hexad user types based on each gameful application as well as findings related to the enjoyment and perception of them.

Cloud Clicker and Hexad User Types To analyze whether Cloud Clicker may be used to predict Hexad user types, a canonical correlation analysis ("CCA") was conducted using the score of the six statements in the gameful application as predictors of the six Hexad user types measured by the Hexad user types questionnaire. A CCA is preferable when analyzing the association strength between two sets of variables and allows to evaluate the multivariate shared variance between them (i.e., between the six statement scores of the gameful application and the six scores of the Hexad subscales) [318]. Next, this method is explained based on Sherry and Hanson's guide on using CCA [318].

The core idea of CCA is that the set of predictor variables and the set of criterion variables are combined into a synthetic variable each (i.e., there is a synthetic predictor and a synthetic criterion variable). The canonical correlation is the correlation between these synthetic variables. Each pair of synthetic variables is called a canonical function ("CF"). Canonical functions are comparable to principal components in Principal Component Analyses (PCA) with the main difference that the CFs are composed of two different variable sets and thus can be seen as an extension of PCA [357]. In line with this, CCA was loosely defined as "a double-barreled principal components analysis" [332].

As long as there is residual variance left in the two variable sets which cannot be explained by the already derived canonical functions, the above process is repeated. This continues until either no residual variance is left to be explained or there are as many canonical functions as there are variables in the smaller variable set. Although CCA can accommodate variables without relying strictly on multivariate normality [357], multivariate normality was assessed by inspecting univariate Q-Q plots, skewness, and kurtosis of each variable included in the CCA. The Q-Q plots mainly supported the assumptions of normality, whereas some variables were shown to be slightly skewed. However, all skewness and kurtosis values were within the acceptable thresholds of skewness < 3 and kurtosis < 8 [181], given that the maximum absolute values of skewness and kurtosis were found to be 2.5 and 6.8 respectively such that the CCA could be conducted.

	CF 1		**CF2**		**CF3**		**CF4**		**CF5**	
Pred.	*co*	*rs*	*co*	*rs*	*co*	*rs*	*co*	*rs*	*co*	*rs*
G_DI	.10	**.46**	.04	-.21	.19	-.03	-.06	-.10	1.34	**.76**
G_FS	.38	**.73**	.11	-.12	.32	**.36**	-.12	-.22	.49	-.17
G_AC	-.53	**-.37**	.21	.01	.89	**.85**	.34	**.36**	.76	-.07
G_PL	-.11	.18	.99	**.91**	-.13	-.19	.45	.31	.90	.01
G_PH	-.22	-.30	-.21	**-.64**	-.29	**-.46**	.72	**.45**	.69	.10
G_SO	-.64	**-.67**	.40	-.04	.10	-.30	-.57	**-.65**	.98	.04
Crit.										
Hex_DI	.30	**.41**	-.31	**-.37**	-.18	-.08	-.08	-.17	1.00	**.78**
Hex_FS	.75	**.39**	-.01	-.34	-.09	.19	-.50	**-.41**	-.70	-.08
Hex_AC	-.56	-.33	-.07	-.05	1.24	**.59**	-.11	-.19	.37	.30
Hex_PL	.24	-.21	.92	**.60**	-.40	-.12	.26	.05	.16	**.36**
Hex_PH	-.16	-.31	-.57	**-.51**	-.21	-.12	.99	**.36**	-.11	.07
Hex_SO	-.64	**-.63**	-.19	-.17	-.58	**-.35**	-.87	**-.46**	.03	.13

Table 6.10: Structure coefficients (rs) and standardized canonical function coefficients (co) for predictor variables (statement scores in Cloud Clicker: G_DI etc.) and criterion variables (user type scores: Hex_DI etc.) for the canonical functions. Bold entries represent loads higher than $|.35|$, underlined entries represent loads higher than $|.50|$.

Given that 10 participants per observed variable are recommended to reach a reliability of 80% [330], our sample size can be considered as adequate.

Overall, the full model across all CF was statistically significant using the Wilks's λ=.256 criterion, $F(36, 595.59) = 6.01, p < .001$. This shows that the variance unexplained by the model is 25.63%. Consequently, the full model is able to explain 74.37% of the variance (the r^2 type effect size is $.74$) shared between the two variable sets. Given that the recommended threshold for strong effects was derived to be $r^2 = .64$ [121], the model can be considered to explain a substantial amount of variance between the two variable sets. Based on this, we derive **R1: The score of statements in the gameful application is substantially associated to the scores of the Hexad user type questionnaire.** This result shows that the two variable sets are strongly related. As a next step of the CCA, we will consider the results of the dimension reduction analysis to analyze whether the predictor variables (the score of the statements in Cloud Clicker) load on the same canonical functions as the corresponding Hexad user types. This is important to investigate whether the statements we have chosen for a certain user type actually represent this user type, given our data.

The dimension reduction analysis yielded six canonical functions (CF1–CF6) with squared canonical correlations of .36, .33, .22, .17, .09 and .00 each. The first five canonical functions were statistically significant whereas CF6 did not explain a statistically significant amount of shared variance between the variable sets (CF1–CF4: $p < .001$, CF5: $p = .011$, CF6: $p = .99$). Therefore, CF6 will not be interpreted as part of the analysis. Figure 6.6 presents the structure coefficients

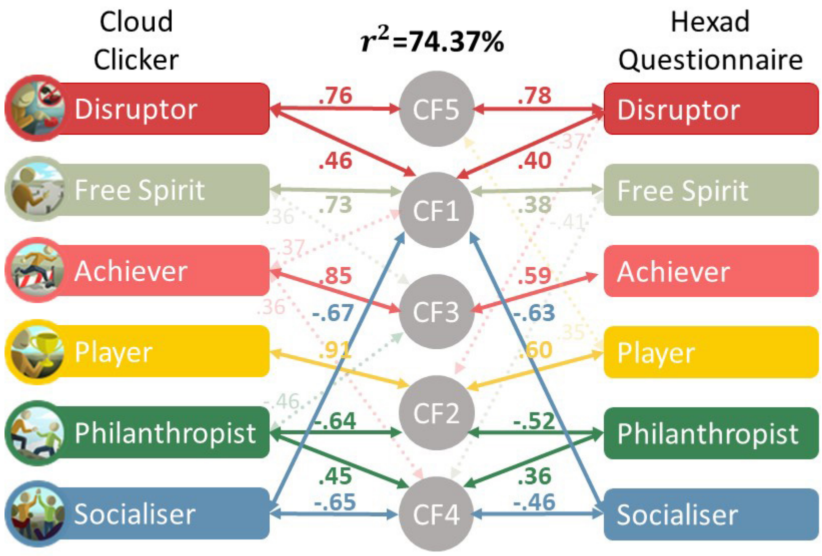


Figure 6.6: Structure coefficients for CF1–CF5 being stronger than $|.35|$ for Cloud Clicker. Dotted and transparent lines indicate relationships which differed between the predictor and criterion variables.

for CF1–CF5 being stronger than $|.35|$ (an upper threshold for weak factor loads established in [367]). Dotted and transparent lines indicate relationships which differed between the predictor and criterion variables. All standardized canonical function coefficients and structure coefficients can be found in Table 6.10.

While standardized canonical function coefficients represent the weights applied to the observed variables to combine the unobserved synthetic variables, structure coefficients are simple bivariate correlations between observed variables and synthetic variables [255]. It can be seen that most predictor and criterion variables have large structure coefficients loading substantially (i.e., $> |.5|$ according to [86]) on the same canonical functions. This is supported by the symmetry of the relationships, which can be seen in Figure 6.6. In terms of strength of the correlation, the Free Spirit subscale is an exemption to this as it is the only variable having no structure coefficient higher than $|.5|$ but loads moderately [86] on CF1.

Thus, we formulate **R2: All predictor and criterion variables have medium to large structure coefficients loading on the same canonical functions**. This result shows that there is not only a substantial relationship between the variable sets (**R1**) but also that the predictor variables load on the same canonical functions as the criterion variables. This means that the statements we have chosen represent the corresponding Hexad user types.

To further analyze the correlations between the Hexad user types questionnaire scores and the scores obtained through the gameful application, we calculated bivariate Spearman's rank correlation coefficients. Since the gameful application

	DI_R	FS_R	AC_R	PL_R	SO_R	PH_R
G_DI	**.32**				**-.23**	
G_FS	.21	**.36**			-.36	
G_AC			**.35**			
G_PL				**.50**		
G_SO	**-.32**				**.44**	
G_PH				**-.22**		**.56**
DI	**.44**			-.22	**-.37**	
FS		**.52**		-.38		
AC			**.45**			
PL		-.37		**.60**		-.32
SO	**-.52**	-.31			**.64**	
PH		-.31	-.23	**-.26**		**.54**

Table 6.11: Spearman's rank correlation coefficients between the ranked Hexad user type scores (DI_R etc.) and the score of each statement in Cloud Clicker (G_DI etc.) as well as between the ranked Hexad user type scores and the absolute Hexad user type scores (DI etc.). All p<.01.

requires users to make a binary choice whereas the original Hexad user types questionnaire allows to have the same score in multiple user types, we ranked the Hexad user type scores by assigning values from 0–5 to ensure comparability. These ranked Hexad scores are denoted by "_R" in the results. As a reference for interpretation, we also added the correlations between the absolute score of each user type of the Hexad and the ranked Hexad user type. The results are shown in Table 6.11. Providing further support for the results of the CCA, it can be seen that there are medium to large size correlations [86] between the scores of each statement of the gameful application and the ranked Hexad user types. This leads to **R3: The ranked Hexad scores are positively correlated to the corresponding scores of each statement of Cloud Clicker having medium to large effect sizes**. This provides further support for the suitability of the statements and visualizations used in Cloud Clicker. It can also be seen that the correlations between the ranked Hexad user type scores and the score of each statement in the gameful application are similar to the correlations between the absolute Hexad user type scores calculated using the Hexad questionnaire and the ranked Hexad user type scores concerning both strength and direction of the correlations. Thus, we formulate **R4: The correlations between the scores of the statements of Cloud Clicker and the ranked Hexad scores are similar to the correlations between the absolute Hexad user type scores and the ranked Hexad user type scores**. **R4** is reflected visually by the two highlighted diagonals in Table 6.11. In line with the results of the CCA, this indicates that assessing the

	CF 1	
Predictor	*co*	*rs*
Collectibles	.42	-.05
Achievements	.20	.01
Points	.37	.06
Leaderboard	1.08	**.89**
Virtual Character	-.06	**-.72**

	CF1	
Criterion	*co*	*rs*
Hex_DI	.37	**.37**
Hex_FS	-.35	.10
Hex_AC	.62	**.69**
Hex_PL	.51	**.79**
Hex_PH	-.40	.09
Hex_SO	.22	**.46**

Table 6.12: Structure coefficients (rs) and standardized canonical function coefficients (co) for predictor variables (number of interactions with gamification elements in Snowball Shooter) and criterion variables (score of each user type: Hex_DI etc.) for CF1. Bold entries represent loads higher than $|.35|$, underlined ones higher than $|.50|$.

ranking of Hexad user types with Cloud Clicker is comparable to assessing the ranking of Hexad user types based on the questionnaire. Consequently, taking **R1–R4** together, our results demonstrate that Cloud Clicker explains a substantial amount of shared variance between the predictor and criterion variable sets, that the statements and visualizations used in Cloud Clicker successfully represent their corresponding Hexad user types and that Cloud Clicker can be used to assess the ranking of a user's Hexad type scores.

Snowball Shooter and Hexad User Types Again, a CCA was conducted to investigate the shared variance between the amount of interactions with each gamification element in Snowball Shooter as predictor variables and the Hexad scores of each user type. The analysis yielded five canonical functions with squared canonical correlations of .20, .12, .07, .01, and .00. The full model was statistically significant (Wilks's $\lambda = .659$ criterion, $F(30, 546.00) = 2.00, p = .001$). This leads to result **R5: The amount of interactions with gamification elements in Snowball Shooter is moderately associated to the scores of the validated Hexad user type questionnaire.** Similar to Cloud Clicker, this means that the amount of interactions with gamification elements and the Hexad user types are related. However, the variance shared between the two sets of variables was considerably lower than in Cloud Clicker as the model of Snowball Shooter accounts for 34.1% of the shared variance. This indicates a moderate effect size [121]. As part of the dimension reduction analysis, it was found that solely CF1 was explaining a significant amount of variance. Therefore, only CF1 was considered for the interpretation of the canonical correlation analysis. An overview of the structure coefficients and standardized canonical function coefficients for CF1 can be found in Table 6.12. The structure coefficients for CF1 being stronger than $|.35|$ are show in Figure 6.7.

Looking at the CF1 coefficients, it can be seen that the score in the gamification elements Leaderboard and Virtual Character strongly contribute to the synthetic predictor variable. While the score in Leaderboard contributes positively to CF1,

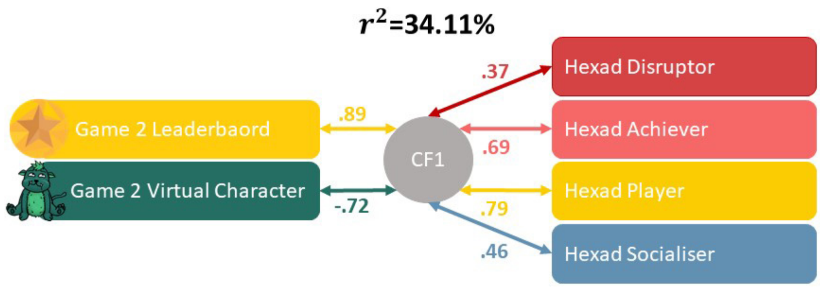


Figure 6.7: Structure coefficients for CF1 in Snowball Shooter being stronger than $|.35|$.

Virtual Character contributes negatively. Regarding the criterion variable set in CF1, Achiever and Player were the primary contributors to the criterion synthetic variable, with secondary contributions by Socializer and Disruptor. All of the aforementioned variables add positively to CF1. Thus, the amount of interactions with the Leaderboard is positively related to the score in the Achiever, Player, Disruptor and Socializer factors of the Hexad. This is in line with previous results based on self-reported preferences for gamification elements [342]. That the Virtual Character is negatively contributing to CF1 indicates that participants interacting with it were likely not interested in interacting with Leaderboards and tended to score lower on the Player, Achiever, Socializer and Disruptor types.

Analyzing the relationships between interaction with gamification elements and Hexad user types further, we again calculated bivariate Spearman's rank correlations. Similar to Cloud Clicker, we considered the absolute score in each Hexad user type as well as the ranked Hexad user type. Table 6.13 shows significant correlation coefficients.

Overall, the correlations support the results from the canonical correlation anal-

	DI	AC	PL	SO	PH_R
Collectibles					
Achievements		.19*			
Points					
Leaderboard	.19*	.30**	.33**	.23**	-.23**
Virtual Character		-.25**	-.22**		.19*

Table 6.13: Spearman's rank correlation coefficients between the amount of interactions with each gamification element in the Snowball Shooter application and the absolute Hexad user type scores (DI etc.) as well as the ranked Hexad scores (PH_R). Hexad user types having at least one significant correlation are shown. *$p < .05$, **$p < .01$

	Hexad	G1	G2
IMI Competence	5.00 /1.05	5.08 / 1.13	4.84 / 1.32
IMI Choice	5.29 / 1.47	5.29 / 1.37	5.48 / 1.36
IMI Enjoyment*	3.87 / 1.34	4.56 / 1.45	4.81 / 1.62
IMI Pressure*	1.96 / 0.98	2.44 / 1.30	2.34 / 1.27
IMI Immersion*	4.91 / 1.10	5.24 / 1.21	5.40 / 1.23

Table 6.14: Mean / Standard Deviation for each condition of the study. All variables are measured on 7-point scales. Variables for which the Friedman ANOVA was significant are marked (*). G1=Cloud Clicker, G2=Snowball Shooter.

ysis. It can be seen that most correlations were found for the Leaderboard and Virtual Character gamification elements. The positive correlation between the ranked Philanthropist score and the amount of interactions with the Virtual Character gamification element indicates that, as expected based on the definition of the user type [342], the Virtual Character seems to be particularly relevant for the Philanthropist. Also, the positive correlations between the Achiever, Player, Disruptor and Socializer user types and the Leaderboard are in line with the results from the canonical correlation analysis and were expected based on previous work [342]. In addition, the positive correlation between Achievements and the Achiever was expected and is in line with previous findings [342].

Taking these results all into account together, we establish **R6: The amount of interactions with gamification elements correlates to their corresponding Hexad user types**. On a more abstract level, **R5** and **R6** mean that users interact with gamification elements that correspond to their Hexad user types. This is an important result for the validity of the Hexad model, as previous research did not consider actual user behaviour within gameful applications, as far as we know. However, it should be noted that we could not find correlations between Collectibles and Free Spirits as well as between Points and Players.

Perception of Cloud Clicker and Snowball Shooter To analyze the perception of the gameful applications compared to completing the Hexad user types questionnaire, a repeated measures Friedman ANOVA was calculated for the IMI and PXI factors (the responses in the IMI and PXI responses were not normally distributed). The Durbin-Conover method was used for post-hoc analysis and the Benjamini-Hochberg false discovery rate [40] was used to adjust significance values for multiple comparisons. For this analysis, all participants who decided to not answer either the PXI or IMI questions after the Hexad questionnaire, Cloud Clicker, or Snowball Shooter, were excluded. Thus, the responses of 113 participants were considered.

Table 6.14 provides an overview of the mean and standard deviations of the

IMI subscales and the PXI Immersion subscale in each condition (Hexad questionnaire, Cloud Clicker, Snowball Shooter). When analyzing the Competence subscale of the IMI, we did not find significant differences between the conditions ($\chi^2(2) = 5.54, p = .063$). Similarly, no significant differences were found for the Choice subscale of the IMI ($\chi^2(2) = 4.94, p = .085$). However, the Enjoyment score differed significantly ($\chi^2(2) = 35.8, p < .001$). The post-hoc analysis revealed that both gameful applications were significantly more enjoyable than completing the Hexad questionnaire ($p < .001$ each). Also, Snowball Shooter was perceived as more enjoyable than Cloud Clicker ($p = 0.012$).

We summarize these effects by **R7: Both gameful applications are perceived as more enjoyable than completing the Hexad questionnaire**. This result indicates that gameful approaches might be more suitable to be used within gameful systems than the Hexad questionnaire, when a gameful experience is important. In addition, we found that the perceived pressure differed significantly between the conditions ($\chi^2(2) = 11.7, p = .003$). Both gameful applications scored higher in the Pressure factor of the IMI (both $p < .001$), whereas no difference was found between the gameful applications themselves ($p = 1.00$), leading to **R8: The perceived pressure is significantly higher in both gameful applications**. This finding is likely related to the timed and round-based nature of the gameful applications (i.e., the fact that we used a timer in Cloud Clicker and 15 rounds of interaction in both gameful applications).

Finally, we analyzed whether the immersion, as measured by the Immersion subscale of the PXI, differs across the conditions. The Friedman ANOVA revealed a significant effect ($\chi^2(2) = 26.5, p < .001$). Both gameful applications scored significantly higher on the Immersion subscale of the PXI (each p<.001) while there were no differences between the two gameful applications ($p = .127$). Thus, we derive **R9: Both gameful applications were perceived as more immersive than completing the Hexad questionnaire**. Taking **R7–R9** together, it seems like the higher pressure is not perceived negatively but may cause a feeling of higher immersion leading to a more enjoyable experience [94].

6.3.3 Discussion and Limitations

Our results show the scores of the statements ("predictor variables") in the Cloud Clicker application and the Hexad user type scores ("criterion variables") are substantially related to each other. They share 74.37% of their variance (**R1**). We also found the structure coefficients of the predictor variables and the structure coefficients of the criterion variables load on the same canonical functions, with large effect sizes (**R2**). In addition, we found there are medium-to-large correlations between the ranked Hexad scores and the scores of the statements in Cloud Clicker (**R3**), which were shown to be comparable to the correlations between the absolute Hexad scores and the ranked Hexad scores (**R4**). Taking **R1–R4** together, we conclude that Cloud Clicker can be used to predict Hexad user types in a gameful way, when an order of user types is sufficient (which is likely the case when personalizing gamification elements set in a gameful system).

We suggest using the order of scores, since Cloud Clicker uses a binary choice instead of allowing users to rate their agreement with statements on an ordinal scale (as was done in the validated Hexad questionnaire). Ultimately, these results support **H1a: The score of the statements in Cloud Clicker is correlated to the corresponding Hexad user types and thus may be used to predict them**. This is explainable because we used statements which were similar to the Hexad questionnaire items with the highest factor load.

Regarding the Snowball Shooter application, in which we analyzed whether the amount of interactions with gamification elements ("predictor variables") could be used to predict Hexad user types ("criterion variables"), we found that in general, there is a relationship between the predictor and criterion variables (**R5**). This is an important finding on its own because it shows that the correlations between the Hexad user types and preferences for gamification elements – which have been identified based on self-reports using textual descriptions or storyboards in previous work [13, 342] – can be replicated based on actual user behaviour. This finding supports the suitability of the Hexad model to explain user behaviour in gameful systems. When analyzing correlations between the amount of interactions with gamification elements and Hexad user types further, we found correlations that were expected based on previous work (**R6**). This is in line with the findings by Hallifax et al. [144], and supports the validity of the Hexad model.

However, it should be noted that two correlations that were expected could not be found (between Free Spirits and Unlockables as well as between Players and Points). A reason might be that Snowball Shooter did not motivate a specific behaviour (as gameful systems usually do [146]) but rather encouraged users to try out different gamification elements. This would be likely for Free Spirits who like to explore [342] and thus might explain the absence of correlations for this user type. The unlocked items could only be collected and not be used for anything else. This might have affected the engagement of users negatively. Also, an incentive for collecting points was missing, which might have been detrimental to Players' motivation to collect points. Considering that the shared variance between predictor and criterion variable sets was moderate (34.11%), we do not recommend deriving Hexad user types based on interaction behaviour alone in Snowball Shooter. However, the amount of interaction with gamification elements might still be a useful factor for dynamic adjustments of gameful systems. Based on **R5** and **R6**, we consider **H1b: The amount of interactions with gamification elements in Snowball Shooter is correlated to the corresponding Hexad user types and thus may be used to predict them** partially supported. Although we found that there are correlations to gamification elements that match the corresponding Hexad user types, the amount of shared variance between the two sets of variables is too low to reliably predict Hexad user types.

Furthermore, our results show that both gameful applications were perceived as more enjoyable (**R7**) and more immersive (**R9**) than the traditional Hexad questionnaire. Based on these results, **H2a: Both applications are perceived as more enjoyable** and **H2d: Both applications are perceived as more immersive** are supported. We also found a significant effect on perceived pressure (**R8**), which might be related to the higher immersion and arguably a higher sense of flow [94, 148]. Based on this, **H2c: Participants feel more pressure in both applications** is supported.

We conclude that both gameful applications provide a more pleasurable gameful experience than completing the Hexad questionnaire. **H2b** is not supported because no effects were found regarding perceived competence. Contrary to previous work [303], it seems like the gamification elements did not enhance the perceived competence through feedback as much as expected. A potential explanation might be that the user interface elements (such as radio buttons) provide visual feedback on their own, potentially enhancing the perceived competence in the baseline condition.

On a more abstract level, our results show that Hexad user types can be assessed by using binary choices. These could be easily adapted to different contexts or could even be turned into concrete choices a player needs to make in a more game-like setting. Also, we show that interacting with gameful design elements is related to a user's Hexad type. This could be used to infer Hexad user types dynamically when interacting with a gameful system and provides huge potential for further research.

Limitations

This study has several limitations. First, transforming the 7-point Likert scales into binary decisions and considering only one particularly relevant item per Hexad user type as was done in Cloud Clicker unavoidably leads to a loss of information. Consequently, we recommend using Cloud Clicker as a practical tool when personalizing gameful systems while ensuring a gameful user experience and to prevent a loss of immersion. For scientific purposes, we acknowledge that Cloud Clicker cannot replace the validated Hexad questionnaire [337]. Second, we used statements in Cloud Clicker that were similar to the statements with high factor loads in the Hexad questionnaire validation study (but not the same). We decided for one statement instead of all four statements. This means that even though we used statements that had a substantial factor load, using other statements for the corresponding user types might lead to different results.

Next, regarding the Snowball Shooter application, it did not motivate a real-life behaviour but allowed users to interact with the gamification elements. Although this allows users to experience how certain gamification elements work, their perception might be different when motivating concrete real-life goals in specific domains such as physical activity. The free exploration of gamification elements within Snowball Shooter could be particularly appreciated by Free Spirits, which might explain why we could not find any correlations for this user type (because they might have tried several different gamification elements).

Also, we decided to randomize the order of the gameful applications, but not of the Hexad questionnaire. This was done to avoid detrimental effects of removing gamification on the perception of the Hexad questionnaire [146]. Since the IMI and PXI scales are the only shared dependent variables across these conditions, potential ordering effects would not primarily concern the main goal of the study, i.e. predicting Hexad user types from interaction behaviour. In addition, we assume the chance of ordering effects is low since filling out a questionnaire and interacting with gameful applications can be considered different tasks, reducing the chance of practice effects [273]. Nevertheless, the fact that participants always started by completing the Hexad questionnaire should be considered.

Since participants were asked to answer roughly 100 items in total, we cannot rule out fatigue effects. However, considering the number of items and that the duration of the study is within a maximum length of 20 minutes [284], no practically relevant effects on data quality are to be expected [153].

Last, we acknowledge that, although the design of the applications is based on previous research, certain decisions are inherently a matter of interpretation, which might affect the external validity.

Future work should investigate whether different game controls (e.g., allowing for continuous user input) will enhance the accuracy of using gameful applications to predict Hexad user types. Further research should be conducted into correlations between actual user behaviour and Hexad user types to replicate the findings of previous research. Also, our findings should be validated in different domains.

6.3.4 Contribution to Research Questions

The study revealed that Cloud Clicker could be used to assess Hexad user types in practice, as an alternative to using the Hexad questionnaire. We showed that the ranking of Hexad user types determined by Cloud Clicker is very similar to the ranking we would have obtained when using the traditional questionnaire and that the shared variance between the score of each Hexad user type determined by Cloud Clicker and the score determined by the questionnaire is substantial. For Snowball Shooter, the shared variance is lower; hence, assessing Hexad user types purely based on the interaction behavior of users in Snowball Shooter is not as reliable as with Cloud Clicker. However, our findings also show that participants interacted more with those gamification elements that matched their Hexad user types. For both applications, we were able to show that the user experience is significantly improved compared to using the Hexad questionnaire. In particular, we found that enjoyment as well as immersion are significantly higher in both gameful applications, which is conducive to our goal of assessing Hexad user types without breaking the immersion and gameful experience of a system. Overall, these findings contribute to **RQ4**, i.e. the question of how we can unobtrusively assess personal factors to personalize gamified systems without disturbing the gameful experience. In addition, the significant correlations we have found between the interaction of users with gamification elements and Hexad user types support the assumption that the perceptual differences and correlations we have found as part of Chapter 4 have actual effects on what users like and dislike in implemented gamified systems. Thus, these findings also contribute knowledge relevant to **RQ3**.

6.4 Summary

In this chapter, we investigated how Hexad user types, which can be used to personalize gamified systems, can be assessed unobtrusively, i.e. without breaking the immersion and gameful experience of a gamified system (**RQ4**). In the course of two studies, we contributed approaches to address **RQ4**.

In the first study, we were interested in exploring the potential of using smartphone data, such as installed apps and communication behavior, to predict Hexad user types. We expected to find relationships between this data and Hexad user types, since smartphone data was shown to be linked to personality traits, which in turn were shown to be correlated to Hexad user types. We implemented a smartphone application to gather the aforementioned data and to assess Hexad user types, and conducted a study with this app. We used regression analyses to find relationships between smartphone data and Hexad user types and found a regression model for each of the six Hexad user types. The amount of variance explained by these models varied between the different user types. While the regression model for the Philanthropist explained the highest amount of variance, other models for user types such as Free Spirits and Disruptors explained a considerably smaller amount of variance. Although these models need further validation and do not allow us to draw ultimate conclusions, they show the general feasibility and potential of the approach.

In the second study, we shifted our focus to predicting Hexad user types based on active user input (instead of passively assessing features to predict them, as was done in the first study) in gamified applications. Also, we were interested in analyzing the user experience of assessing Hexad user types in gameful ways and whether advantages in terms of enjoyment and immersion could be found, compared to asking users to fill out the validated Hexad questionnaire. To address these questions, we implemented two gamified applications – Cloud Clicker and Snowball Shooter – and used them in an online study. In Cloud Clicker, users were asked to decide which of two statements was more important to them, with the pairs of statements being shown in each of 15 rounds in a gameful way. In Snowball Shooter, users were asked to shoot snowballs at certain virtual items in order to make progress in gamification elements. They were free to decide which gamification elements they would like to interact with. We found that Cloud Clicker can be seen as a gameful way of assessing Hexad user types, since the shared variance between the answers in the Hexad questionnaire and the decisions in the gameful application was almost 75%. Also, we found that the immersion and enjoyment when using Cloud Clicker was significantly higher than when using the traditional Hexad questionnaire. Thus, Cloud Clicker could be a promising alternative to assess Hexad user types in practice. For Snowball Shooter, the shared variance was smaller (34%). However, this still shows that the gamification elements which users interact with are predictable by their Hexad user type, contributing knowledge to the field of gamification personalization. Moreover, this finding illustrates the general feasibility of tailoring gamified systems based on how users interact with them.

To sum up, the two studies demonstrate that personalization of gamified systems is possible without relying on scientific questionnaires and procedures. We could show that pragmatic approaches of assessing or predicting Hexad user types, which may not be suitable for scientific purposes, may be key to facilitating personalized gamification in practice. These approaches either do not need any user interaction and could be hidden from the user in the background, or could be seamlessly integrated in the gamified application without breaking its gameful experience.

Chapter 7
Conclusion

In the final chapter of this thesis, we first summarize the main aspects we investigated with regard to our research questions and outline our major contributions to understanding how gamification works, how it affects motivation, how motivation can be increased through gamification and how to facilitate this in practice. Next, we will use these contributions to create a conceptual framework for how motivation can be increased in gamified systems. This conceptual framework is based on our empirical findings and could serve as guidance for future research in the field, as will be discussed and outlined at the end of this chapter.

7.1 Summary and Major Contributions

This thesis focused on the question of how gamification affects motivation in behavior change support contexts, what effect personal factors have and how these factors can be utilized to increase motivation in gamified systems (see Section 1.3). To contribute answers to this, we conducted 18 user studies involving a total of more than 1,100 participants. We relied on a wide range of study methods including semi-structured interviews, laboratory studies, and online studies as well as in-the-wild studies. To derive answers to our main research question, we used qualitative methods such as thematic analysis [54] and content analysis [156], quantitative methods for hypothesis testing, and mixed methods combining the strengths of both approaches.

We elaborated on the important role of games and play for our culture and society in Chapter 1. We demonstrated that video games can effectively satisfy basic psychological needs and thus are powerful in creating motivational and satisfying experiences for players. We then introduced gamification as a concept to transfer these motivational experiences from purely game contexts to non-game ones by utilizing elements known from games. Besides illustrating the potential of gamification to increase user motivation in general, we also provided insights on its role to change and shape behaviors in the context of persuasive technology. At the end of Chapter 1, we demonstrated the need to better understand how and why gamification works, and which factors play a role in this regard. Based on both past empirical studies and SDT, we showed that the context and the individual person are important factors for understanding user motivation in gamification settings. This was used as a basis to formulate the research questions of this thesis, which were presented at the end of the chapter.

In Chapter 2, we introduced important theories and concepts related to motivation, especially SDT. This is important, since the ultimate goal of gamification is increasing motivation and the main research question of this thesis is focused on how motivation can be increased in gamified systems. Next, we introduced relevant theories and models and discussed their relationship to SDT. Afterwards, past research investigating the use of gamification in different behavior change contexts was presented. We learned that "one-size-fits-all" gamification can lead to positive outcomes, but may also lead to no or even adverse effects. We also saw that the presented approaches lack insights on the reasons for these differences, i.e. it remained unclear why gamification does or does not work, and which factors play a role in that regard. In the next part of the related works, we focused on past research contributing answers to the question of what determines whether gamification works. We learned that considering gamification elements instead of the whole gamified system helps to provide insights on the factors that play a role in how gamification is perceived. Also, we saw that such an approach helps to better understand how gamification elements may influence basic psychological needs. In addition, we learned that interpersonal differences exist in how gamification elements are perceived, which calls for ways to tailor gamified systems to individual users. Two approaches to this problem were presented – customization (allowing users to adjust the gamified system to their needs) and personalization (adapting the gamified system to the users' personal characteristics). In past research, personalization has been shown to be beneficial and preferred by users, which is why current gamification research investigates which factors should be considered for personalization. We learned that considering demographic factors, personality traits and user or player types yielded promising results. In the last part of the related work section, we presented relevant works from the field of predicting personal characteristics in an unobtrusive or engaging way. We learned that personality traits can be predicted from smartphone data, and that even player types and players' experience in a game can be predicted with little or no explicit user interaction.

7.1.1 Major Contributions to RQ1

In Chapter 3, we mainly investigated **RQ1**, i.e. the question of how gamification affects motivation and related behavioral as well as psychological measures in behavior change contexts. To contribute answers to this question, we implemented three gamified systems encouraging behavior change in the contexts of *physical activity*, *hand washing* and *online advertising*. Overall, we conducted six user studies in which these systems were informed based on online studies and evaluated as part of both lab and in-the-wild studies. When evaluating these systems, we relied on validated instruments to assess psychological measures related to SDT as well as behavioral measures. We also used qualitative methods to broaden the scope of potential findings. This allowed us to contribute a wide range of answers to **RQ1** and to derive assumptions on the role of the context itself. We first presented findings from an in-the-wild study in which a gamified mobile app was extended with a public display to encourage walking among people visiting a gym. We found that the satisfaction of basic needs, especially social relatedness, seemed to be the deciding cause for the significant increase in the number of steps walked. The qualitative analysis of semi-structured interviews confirmed this finding. In addition, it revealed reasons why individual participants did not increase their step count, which seemed to be related to feeling incompetent or incapable to keep up with other users, ultimately leading to amotivation.

In the second context, hand washing, we found that gamification induced positively valenced affective experiences and increased intrinsic motivation to wash one's hands. It seemed that the positive emotions evoked by the gamified system led to an increased enjoyment and motivation. Moreover, we found that participants underestimated the time they spent washing their hands when using gamification, suggesting that flow experiences might have been increased, which is another measure positively related to intrinsic motivation [95]. In the in-the-wild study, in which the gamified system was evaluated in a public bathroom, we found that the system changed people's behavior, almost doubling their hand washing duration. Also, decreases in the contamination level of the bathroom's door handle were found, further supporting the behavioral change of people. When discussing these results with regard to SDT, we could see that the context itself might thwart certain basic psychological needs, but that these could be supported with gamification. In line with the previous study, this suggests that context-inherent need satisfactions and frustrations might play an important role regarding the success of gamified systems.

In the last study of the chapter, we found support for this assumption. Here, we investigated different gamified systems to improve users' experience with online advertisements. Although we found that gamification increased intrinsic motivation to interact with and consume ads in all systems, we also found that the system which focused on gamification elements supporting autonomy was perceived best by users. This might be explainable by the fact that the context – online advertising – itself frustrates the need for autonomy, but this need can be satisfied by a corresponding autonomy-supporting gamified system. Besides these findings related to motivation and the user experience, we also found that the gamified systems increased cognitive measures such as brand or product recognition and recall, and affected attitudes towards the website hosting the ad. We summarize the aforementioned major contributions to the first research question in the following.

Major Contributions ("MC") to RQ1

MC1: Gamification can satisfy basic psychological needs in behavior change contexts and evoke positive affective experiences, leading to an increase in motivation.

MC2: The increased motivation in gamified systems can lead to changes of behavior in these contexts.

MC3: Contexts may inherently thwart basic psychological needs. Motivation in gamified systems may be increased by using gamification elements which satisfy the needs being thwarted by the context.

7.1.2 Major Contributions to RQ2

We focused on the question of how personal factors affect the perception of gamification elements (**RQ2**) in Chapter 4. To provide answers to this question, we investigated to what extent the factors age, stage of change, and Hexad user type explain inter-personal differences in the perception of gamification elements by conducting a total of seven user studies, including online, interview, and laboratory studies as well as both qualitative and quantitative methods. First, we presented results from a study with older adults (aged 75 and up) in which we used a mixed-methods approach to understand their preferences when playing games as well as to evaluate their perception of frequently used gamification elements. As a complement to this, we assessed the Hexad user types of older adults and compared them to a younger sample. Our findings demonstrate that older adults have specific preferences regarding gamification elements. We learned that their motivation to play games is focused on social aspects, altruism and relatedness needs. This is also reflected in the perception of commonly used gamification elements. Older adults liked elements such as virtual characters which they could care for and social collaboration, whereas they disliked competence-focused elements such as badges or competition. Supporting these results even further, we also saw that a huge majority of older adults scored highest on the Philanthropist Hexad user type. This shift away from performance- and competence-focused feedback towards social relatedness was discussed through the lens of SDT based on findings from psychology research, showing that the importance of social relatedness needs increases with increasing age.

After investigating age, we focused on the stage of change as a factor regulating the perception of gamification elements. We created visualizations for three types of achievement goals and evaluated their perceived persuasiveness and to what extent this was mediated by the stage of change of participants. We found that the stage of change plays a role in the perception of the three types of goals.

Task-based goals and other-based goals were found to be more persuasive in higher stages of change. Again, basic psychological needs theory played an important role to explain these findings. Since the effectiveness and capability factors showed significant correlations to the stage of change, it seemed that perceived competence, which should be lower among participants in low stages of change, was the deciding cause for the effects. We also found that self-based goals seemed to be best suited for participants, regardless of their stage of change. A potential reason might be that competence is defined based on one's own performance in the past, instead of others' performance or a static goal.

In the next user study we shifted our focus to Hexad user types. Based on storyboards illustrating commonly used gamification elements in the context of healthy eating, we were able to investigate how the perception of these elements differs between users. We found significant correlations between the perceived persuasiveness of the gamification elements and Hexad user types, which were in line with previous findings in different contexts. Again, SDT plays an important role to explain these correlations, since users who scored particularly high on a certain user type perceived gamification elements that meet the basic psychological need underlying that user type particularly well.

In the last study presented in Chapter 4, we investigated the combination of Hexad user types and stages of change. To do so, we again created storyboards illustrating gamification elements in the context of physical activity. In line with the results in the healthy eating context, we found correlations between Hexad user types and the perceived persuasiveness of gamification elements, which are explainable by considering the basic psychological needs these gamification elements satisfy. These findings strengthen the importance and the potential of Hexad user types for personalizing gamified systems. In addition, we also found support for the importance of considering participants' stages of change. Our results showed that the perception of some gamification elements differs significantly between participants in low and high stages of change. In particular, we saw that these differences concerned gamification elements focusing on performance and defining competence based on a static goal (badges, challenges) or comparison to others (social gamification elements). This is in line with the findings of the study, mentioned previously, that investigated the impact of the stage of change on the perceived persuasiveness of achievement goals. Finally, we also learned in this study that combining the stage of change and Hexad user types is a promising approach, since we found that the strength of correlations between the perceived persuasiveness of gamification elements and Hexad user types differed significantly for some gamification elements between participants in low and high stages of change. We provide a summary of the major contributions to **RQ2** in the following.

Major Contributions to RQ2

MC4: Older adults differ from younger people in how they perceive gamification elements. They value social aspects, altruism and social relatedness and prefer gamification elements related to these values and needs.

MC5: A person's stage of change might reflect the extent to which a certain behavior has been internalized and is an important factor for understanding the persuasiveness of gamification elements. Gamification elements focusing on performance, or defining competence based on a static goal or the performance of other users, are perceived as more persuasive in high stages of change.

MC6: Hexad user types might reflect motivational orientations and explain preferences for gamification elements. Users scoring particularly high on certain user types prefer gamification elements satisfying the basic psychological needs underlying the respective user types.

MC7: Combining Hexad user types and the user's stage of change is a promising approach for personalizing gamified systems. It considers both the motivational orientations and preferences of a user as well as the extent to which a behavior has been internalized.

7.1.3 Major Contributions to RQ3

Although we were able to show that factors like age, stage of change and Hexad user types have an influence on how people *perceive* gamification elements in Chapter 4, we still lacked evidence for actual *effects* on behavioral and psychological measures, when using these factors to personalize implemented gamified systems. This gap in knowledge was considered in Chapter 5, in which we presented results from two laboratory studies and one in-the-wild study. First, we evaluated whether participants' stage of change and Hexad user type has an effect on how gamification elements affect their behavior, motivation and affective experience in a laboratory study. In this study, we confronted participants with a gamified application encouraging them to cover more distance on a treadmill. We used gamification elements that should be more suitable for users in high stages of change or users scoring particularly high on the Player, Achiever or Socializer trait of the Hexad (based on the findings in Chapter 4). Indeed, we found that users in high stages of change had stronger affective experiences, and that measures related to intrinsic motivation correlated positively with the Socializer user type, while they correlated negatively with the Free Spirit user type, which is in line with what we expected based on our findings regarding the perception of gamification elements. This shows that considering the stage of change and Hexad user types to personalize gamified systems has an actual effect on users' experience when interacting with the system. In addition, these findings provide insights on the mechanisms that are affected by personalization – basic need satisfaction and consequently intrinsic motivation, as well as affective experiences eliciting emotional responses.

In a second study, we focused on the effects of dynamically tailoring and contra-tailoring the gamification elements used on an image tagging platform based on the participant's Hexad user type. We implemented an image tagging website, on which participants were shown images and asked to provide tags for them. Depending on the condition, they either received no gamification elements at all, suitable, or unsuitable gamification elements. To better understand the underlying mechanisms being affected by personalization based on Hexad user types, we used a wide range of measures, including performance measures such as the number of tags and tag quality, as well as intrinsic motivation, affective experience, and flow experience. We also complemented these by considering psychophysiological measures such as heart rate variability, skin conductance and skin temperature. Our findings support the relevance of Hexad user types for personalizing gamified systems. We found that intrinsic motivation was significantly higher in the tailored condition than in the contra-tailored condition. Also, positive affect was significantly higher in the tailored condition than in the contra-tailored condition, and even significantly lower in the contra-tailored condition than in the control condition. The same was found for flow experiences – they were significantly more prevalent in the tailored than in the contra-tailored condition. These findings demonstrate that personalization affects a wide range of psychological measures positively – it enhances intrinsic motivation, stimulates positive emotional responses and helps users to reach the state of flow. However, we were not able to demonstrate that the positive effects on psychological measures also lead to changes in the participants' behavior. As a potential reason, we assumed that the short study duration and the laboratory setting might have had an influence on participants such that they felt obligated to meet the performance goals introduced by the gamified system, regardless of whether the gamification elements were suitable or not.

Therefore, we conducted an almost-two-year-long in-the-wild study in which we evaluated a gamified fitness course booking system. In this study, we analyzed the number of courses booked in the first year to gather baseline data, and then introduced gamification elements that should be particularly suitable for Achievers, Players and Socializers for another year. We then split participants into two groups – one in which participants scored particularly high in one of the three user types and thus received suitable gamification elements, and another that did not receive suitable gamification elements. Our findings show that the group of users receiving suitable gamification elements increased their number of booked courses significantly more than users who did not receive suitable gamification elements. Also, we found that the main effect, i.e. that gamification increased the number of booked courses when considering the whole sample instead of separating the two groups, was caused by the group of users receiving suitable gamification elements. This not only shows that personalization can have an effect on the behavior of users (even in the long run) but might also serve as an explanation for the different outcomes of one-size-fits-all gamification reported in previous research: are positive, neutral and negative outcomes explainable by the suitability of the gamification elements used for the population considered? Overall, we provided evidence for the positive effects of personalization based on the stage of change and Hexad user types on a wide range of psychological outcomes and also on behavioral outcomes, as will be summarized in the following.

Major Contributions to RQ3

MC8: Personalizing gamified systems based on users' stage of change could lead to stronger affective experiences.

MC9: Personalizing gamified systems based on Hexad user types could lead to stronger affective experiences, an increased prevalence of flow experiences, and higher intrinsic motivation, and it affects the target behavior positively, even over a longer period of time.

MC10: Personalization is important for gamified systems as it has advantages over one-size-fits-all gamification concerning both psychological and behavioral outcomes.

7.1.4 Major Contributions to RQ4

In Chapter 6, we investigated ways to facilitate personalization based on Hexad user types to enable dissemination of our findings in practice. We conducted a smartphone-application-based field study as well as an online study. This contributes to **RQ4**, i.e. the question of how we can unobtrusively assess personal factors to personalize gamified systems without disturbing the gameful experience. This is important because asking users of a gamified system to complete the validated Hexad questionnaire before interacting with the system might detrimentally affect the gameful experience of such a system. To bridge this issue, we investigated whether Hexad user types could be predicted based on smartphone data such as the number and type of installed applications as well as the user's communication behavior. Also, we investigated ways to assess Hexad user types in a gameful way, and whether Hexad user types could be derived based on users' interaction behavior. First, we conducted a user study in which participants were asked to install a smartphone application, which captured smartphone data and assessed their Hexad user type. By using linear regression, we were able to find models to predict each of the six Hexad user types, with varying amounts of explained variance. While the regression models of user types such as the Philanthropist explained a relatively high amount of variance, other user types such as the Player explained less variance in the data. Therefore, we acknowledged that our results should be seen as a first exploration in this direction and could serve as a starting point for future research. This is supported by the fact that the predictors that were included in the regression models could be explained by the characteristics of the respective user types, e.g. that the percentage of rejected calls negatively predicted the Philanthropist score, or that the percentage of answered calls, the percentage of received SMS messages, the relative number of communication apps and the number of unique SMS contacts positively predicted the Socializer score.

As a second study, we presented the evaluation of two gameful applications to predict Hexad user types. We implemented Cloud Clicker, in which users had to decide which of two statements was more important to them, click on the corresponding option and receive gameful feedback (similar to popular cookie clicker games), as well as Snowball Shooter, in which users could freely choose in which gamification elements they would like to make progress by shooting snowballs at corresponding virtual items. In an online study, in which users interacted with both gameful applications in random order and were asked to fill out the validated Hexad questionnaire, we found that Cloud Clicker could be used to assess Hexad user types in a gameful way, and that the interaction with gamification elements in Snowball Shooter was related to Hexad user types. More specifically, we found that the shared variance between the answers in the validated Hexad questionnaire and the choices made in Cloud Clicker was 74%, which can be considered as substantial. Complementary to this, we also found strong correlations between the Hexad user type scores and the scores of the statements shown in Cloud Clicker. Also, benefits of Cloud Clicker regarding both enjoyment and immersion were found. Therefore, Could Clicker could be seen as an alternative to using the validated Hexad questionnaire in practice, i.e. to personalize gamified systems in non-scientific settings. For Snowball Shooter, the shared variance between the interaction in the application and the answers in the validated Hexad questionnaire was lower, reaching 34%. We again found correlations between Hexad user types and the amount of interaction with gamification elements in the application. From a scientific standpoint, this is an important finding, since it shows that which gamification elements users select and interact with is related to their Hexad user type. From a practical standpoint, the shared variance might be too low to replace the Hexad questionnaire solely with Snowball Shooter. However, which gamification elements users decide to interact with could still be helpful in adapting gamified systems and should receive further attention in gamification research. To sum up, we contributed the following major contributions to **RQ4**.

Major Contributions to RQ4

MC11: There are connections between smartphone data and Hexad user types, which could be used to predict the latter.

MC12: Hexad user types can be assessed in a gameful way, without disturbing the immersion and gameful experience of a gamified system.

MC13: There are connections between Hexad user types and gamification elements with which users interact in a gamified system, which could be leveraged to predict Hexad user types.

7.2 Towards a Conceptual Framework to Increase Motivation in Gamification

In this section, we propose a conceptual framework [3], which is based on the major contributions of this thesis through the lens of Self-Determination Theory. This conceptual framework aims to describe the considerations to be made in order to increase user motivation in gamified systems and is meant to guide future research in this domain. Here, the main factors to be considered are contextual factors (i.e. characteristics of the context and situational factors which might inherently support or thwart basic psychological needs) as well as personal factors (i.e. motivational orientations and regulatory styles applied by a person, which might moderate the functional significance of certain stimuli induced by gamification elements). We acknowledge that this conceptual framework should be seen as an initial attempt to describe how user motivation in gamified systems can be increased, and it requires empirical, systematic confirmatory investigation in future work.

7.2.1 The Context

When introducing Self-Determination Theory and its macro-theories, we have learned that contexts can be autonomy-supportive or controlling, and thus play an important role in supporting or undermining intrinsic motivation (see Section 2.1.2) as well as in organismic integration processes (see Section 2.1.2).

In Chapter 3 we focused on how gamification affected motivation in behavior change contexts. We found that gamification motivated users by satisfying basic psychological needs and inducing positive affective experiences (**MC1**). We also found that the increased motivation could manifest in changes of behavior (**MC2**), supporting the potential of gamification as an approach to help people reach their goals and guide them in behavior change processes. When recapitulating our findings from three different behavior change contexts, we observed an interesting pattern. In a context in which people who were going to the gym regularly (and thus might have a high perceived competence regarding their fitness level) were encouraged to increase their physical activity, gamification elements such as badges and points primarily satisfying competence needs [303] seemed not to affect participants' motivation or behavior. However, installing a public display in the gym facilitated social interaction and increased the satisfaction of the need for social relatedness. In contrast, a gamified system using gamification elements such as points and feedback primarily satisfying competence needs [303] increased intrinsic motivation and changed people's behavior in a context in which people did not know how to properly wash their hands. Lastly, in a context in which users were required to consume ads and not given any choice in the matter, a gamified system using gamification elements that support the need for autonomy (such as a virtual character [303]) was preferred over two other gamified systems using gamification elements primarily supporting competence needs. Based on these observations, it seems that the selection of gamification elements should be dependent on which needs are being thwarted by the context (**MC3**). This can be explained by SDT, since contextual factors can have an impact on the functional significance of certain stimuli, such as gamification elements. Therefore, to increase motivation in gamified systems, we suggest that **the context, i.e. the circumstances that form the setting of a gamified system, and any basic psychological needs that are thwarted by it, should be considered to select gamification elements which support these thwarted needs**.

7.2.2 The Person

In addition to the context and its situational factors, SDT also highlights the importance of personal factors, particularly as part of causality orientations theory (see Section 2.1.2). Our findings support the importance of considering personal factors to increase motivation in gamified systems. We have found that age has an effect on how certain gamification elements are perceived, probably due to a change in the importance of certain basic psychological needs (**MC4**). Also, we have seen that the stage of change, i.e. the intention to adopt a certain behavior, plays a significant role in what functional significance is given to stimuli evoked by gamification elements (**MC5**). For instance, we found that users in high stages of change, who thus are more likely to use rather integrated regulatory styles, perceive gamification elements endorsing competence and performance as more persuasive than people in lower stages of change. According to SDT, this suggests that people in high stages of change see the performance goals and challenges established by the gamification elements as informational, while the same gamification elements might be perceived as controlling or even amotivating (because the goals established probably seem out of reach for some people). We also found that Hexad user types explain user preferences for gamification elements (**MC6**). From a SDT perspective, this seems reasonable: the Hexad model itself is built upon SDT and the user types can be seen as personifications of different types of intrinsic and extrinsic motivation (most user types focus on basic psychological needs which are core pillars of intrinsic motivation, while one user type focuses on extrinsic motivation and rewards; see Section 2.2.3). Therefore, the concept of Hexad user types overlaps with the concept of causality orientations, describing inter-personal differences regarding people's motivational orientations (see Section 2.1.2). Lastly, we have seen that combining both Hexad user types and stages of change leads to promising results (**MC7**). For this combination, we could show that considering these personal factors to tailor the set of gamification elements to the user leads to beneficial psychological outcomes (such as increased enjoyment and intrinsic motivation, increased flow experience and stronger affective experiences) and behavioral outcomes (increased adherence to the target behavior) as was summarized by **MC8–MC10**. Based on these findings, we recommend that **the person, i.e. individual characteristics such as age, stage of change, and Hexad user type, should be considered in order to select which gamification elements to use for that person in a gamified system**. In Chapter 6, we contributed practical ways of utilizing personal factors (**MC11–MC13**), which could be used to disseminate our scientific findings in practical implementations. Figure 7.1 summarizes the contextual and personal factors that we recommend considering to increase motivation in gamified systems.

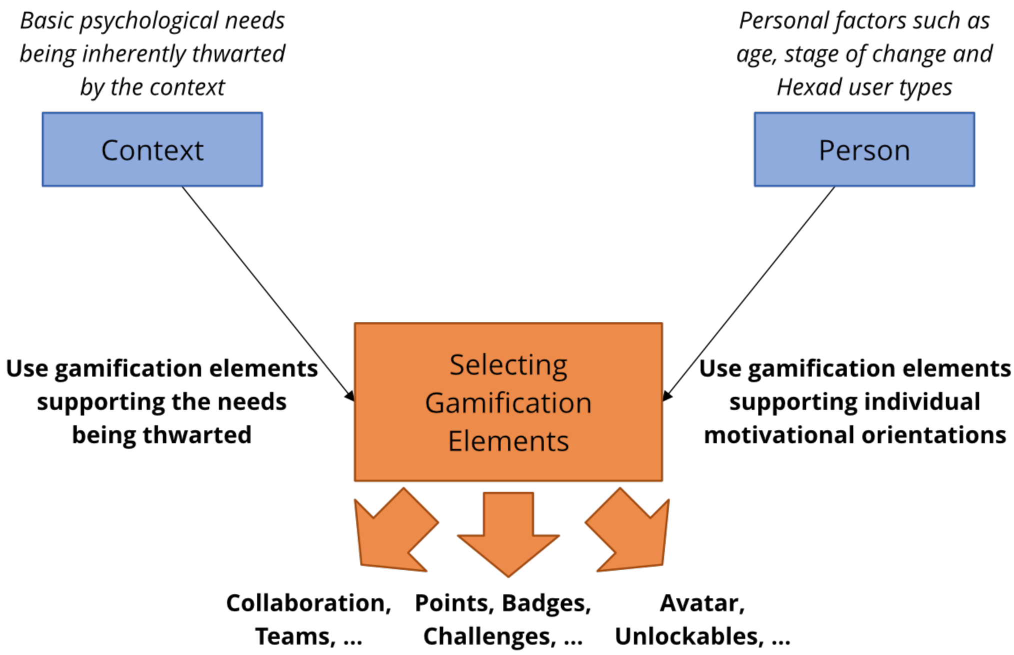


Figure 7.1: Overview of the proposed conceptual framework to increase motivation in gamified systems

7.2.3 Limitations of the Conceptual Framework

As stated before, the conceptual framework should be considered preliminary and needs further empirical investigation. Especially concerning the role of contextual factors, it must be considered that we did not conduct systematic studies to investigate whether gamification elements satisfying needs that are thwarted by the context actually increase user motivation, but rather deduced such a relationship based on our findings. Therefore, this hypothesis should be empirically investigated in controlled experiments. Regarding personal factors, we went a step further than for contextual factors and conducted not only studies on whether personal factors change the perception of gamification elements, but also confirmatory studies in which we empirically investigated whether considering personal factors to tailor gamified systems actually affects behavioral and psychological outcomes. Therefore, it should be noted that this thesis contributed primarily to understanding and utilizing personal factors to increase motivation in gamified systems, and that contextual factors should be investigated further in future work. Since we found empirical support for the importance of the personal factors considered in this thesis, we conclude that personal factors do play an important role to increase user motivation in gamified systems. However, it should be kept in mind that we covered only a subset of potential personal factors, including age, stage of change, and Hexad user types. Thus, there could be other personal factors explaining user preferences in gamified systems, beyond those covered within this conceptual framework.

7.3 Future Work

In this section, guided by the conceptual framework introduced previously, we highlight opportunities for future research and describe how future work could replicate, confirm and build upon our findings.

Empirical validation of the proposed conceptual framework

Considering that the conceptual framework presented in the previous section was deduced from the main contributions of this thesis, a systematic evaluation of its factors and their impact on motivation in gamified systems should follow as a next step. This could be done by studying the impact of contextual factors in a more controlled and isolated way (as described in the following section), considering other personal factors (as described in a subsequent section), and by studying discrete combinations of contextual and personal factors.

Systematic investigation on the influence of the context

Based on our findings from Chapter 3, it seemed that gamification worked best when using gamification elements supporting basic psychological needs that were thwarted by the context. However, we did not systematically investigate whether such a relationship exists. In fact, we did not have an a-priori hypothesis about such a relationship, but observed that there *might* be a connection. Therefore, future work should empirically investigate if there is a connection between needs being thwarted by the context and gamification elements supporting these thwarted needs. This could be done in controlled laboratory experiments, in which basic needs are systematically thwarted. For example, one could systematically craft a context in which the need for autonomy is thwarted and investigate whether gamification elements such as a virtual character or exploration-supportive gamification elements like unlockables work particularly well, compared to other gamification elements. This approach could be repeated for other needs being thwarted and other gamification elements. Depending on the context in which such a study is conducted, online studies could also be done.

Other (behavior change support) contexts

We primarily targeted gamification in health-related contexts such as physical activity, public health or healthy eating, but also covered marketing and advertising. To better understand the context and the intertwined relationships of situational factors, other contexts should be considered in future work to get a more holistic picture of contextual factors in gamified systems. It is important to note that gamification research should rely on theories like SDT to understand these factors; investigating *if* gamification works in specific contexts may support gamification as an approach, but does not provide insights on *how* gamification affects motivation and related measures. Therefore, we suggest tightly coupling the research questions being posed and the hypotheses being formulated with SDT and its macro-theories. Investigating theory-driven research questions in different contexts is an important direction for future work and has great potential to help us understand how gamification affects motivation.

Long-term studies on the effects of personalized gamification

In general, we have seen that there is a lack of long-term studies on the effect of gamification and personalization in gamified systems (see Section 5.4). In Chapter 5, we have investigated the effects of personalized gamification. Here, we could show that personalized gamification positively impacts motivation and related psychological measures in laboratory studies. However, in these laboratory studies, we did not find effects on behavioral measures. In a long-term study over a duration of almost two years, we found that using gamification elements which are suitable for certain Hexad user types led to an increase in course participation in a gym. However, we did not collect any psychological measures nor qualitative data, so we could not investigate further how personalized gamification was able to increase course bookings. Therefore, to better understand the effects of tailoring gamification elements to the user in the long run, long-term studies should be conducted in the future which complement behavioral with psychological measures. In addition, future long-term studies should combine quantitative and qualitative methods to get a better understanding of how personalized gamification impacts users.

Other personal factors

In this thesis, we focused on age, stage of change, and Hexad user types as personal factors. These factors represent demographic data and intentions to change behavior as well as motivational orientations, and all have been shown to play a role in the macro-theories of SDT (see Section 2.2). However, we acknowledge that there are other personal factors which might be relevant to understanding inter-personal differences in the perception and effectiveness of gamification elements. These factors, which of them may be most important and which ones could be combined to explain higher amounts of variance should be studied in future work. In particular, past research has shown that personality traits, demographic data such as gender or cultural factors, and pedagogical profiles play a role in how gamification elements are perceived, just to name a few [182]. However, we would like to emphasize that the selection of which factors should be studied should be theory-driven, i.e. it is important to provide reasons and deduce hypotheses for why a certain factor should play a role in the perception or effectiveness of gamification elements. Thus, when selecting personal factors to study, theories like SDT should be considered (as we did in Section 2.2).

Combining personalization and customization

In Section 2.4 we introduced two approaches toward accounting for inter-personal differences in the perception and effectiveness of gamification elements: customization, i.e. allowing users to adapt the gamified systems to their needs and preferences, and personalization, i.e. adapting the gamified system to the user. While this thesis contributed to the latter, we consider combining both approaches as important future work. Customization, from the perspective of SDT, satisfies autonomy needs by allowing users to freely choose which gamification elements they would like to use. However, past research has shown that users prefer personalization over customization, because customization was seen as difficult and complex, too time-consuming and distracting [260]. However, the complexity of customization could be decreased when combining it with personalization: a system suggesting a small set of suitable gamification elements and allowing users to select which one of these elements they would like to use could have benefits both in terms of providing users a high degree of autonomy and using gamification elements that are suitable for them, without being too complex and time-consuming.

Deriving contextual and personal factors automatically

In Chapter 6, we found that Hexad user types can be predicted based on smartphone data and interaction behavior, to a certain extent. However, we did not investigate whether contextual factors or other personal factors could be automatically derived. Also, we did not focus on the question of what "automatically" means in this context and which level of automation is technically possible and accepted by users. These questions and research directions should be investigated in future work. In particular, past research in the field of context awareness in HCI (such as research by Schmidt et al. [308, 309]) should be considered to inform the design of personalized gamified systems tailored to the application context.

A taxonomy of gamification elements

Lastly, we would like to emphasize the importance of studying the gamification elements themselves. Although there is past work suggesting certain sets of gamification elements and clustering them (see e.g. Tondello et al. [338, 342] or the Octalysis framework by Chou[38]), the concrete realization of these gamification elements might have an impact on the motivation of users. In the end, according to SDT, it is the *functional significance* which makes the difference in whether external stimuli, such as gamification elements, are perceived as informational, controlling or amotivating. Consequently, it can be very little things, such as the wording of a badge or the way the performance of other users is visualized in a system, that can decide whether a gamification element is perceived as informational (supporting intrinsic motivation) or controlling (undermining intrinsic motivation). As an example, a wording such as "you should" or "you must" is very likely to be perceived as controlling [223]. Also, the goal itself, e.g. to unlock a badge, can be a deciding cause – if its requirements are set too high for users and it seems out of reach for them, the gamification element can be perceived as amotivating. However, these little nuances are currently mostly not reflected in gamification research: a badge is badge, independent of how it was realized in the concrete study and system. Therefore, future work should be conducted on building a taxonomy, classifying gamification elements in a way that makes it possible to better compare study results related to a certain gamification element. This would increase the external validity of study results in which atomic gamification elements are investigated.

[38] Octalysis: *Complete Gamification Framework*, `https://bit.ly/3mJ2EdD` (last accessed: 2021-12-01)

List of Figures

List of Tables

Bibliography

[1] Vero Vanden Abeele, Katta Spiel, Lennart Nacke, Daniel Johnson and Kathrin Gerling. 2020. Development and Validation of the Player Experience Inventory: A Scale to Measure Player Experiences at the Level of Functional and Psychosocial Consequences. *International Journal of Human Computer Studies* 135. Elsevier, 1–47. DOI:http://dx.doi.org/10.1016/j.ijhcs.2019.102370

[2] Clark C. Abt. 1987. *Serious Games*. University Press of America.

[3] Dickson. Adom, Emad Kamil Hussain and Adu Agyem. Joe. 2018. Theoretical and Conceptual Framework : Mandatory Ingredients of a Quality Research. *International Journal of Scientific Research* 7 (1). World Wide Journals, 438–441.

[4] Allison E Aiello, Rebecca M Coulborn, Vanessa Perez and Elaine L Larson. 2008. Effect of Hand Hygiene on Infectious Disease Risk in the Community Setting: A Meta-Analysis. *American Journal of Public Health* 98 (8). American Public Health Association, 1372–1381. DOI:http://dx.doi.org/10.2105/AJPH.2007.124610

[5] Hina Akasaki, Shoko Suzuki, Kanako Nakajima, Koko Yamabe, Mizuki Sakamoto, Todorka Alexandrova and Tatsuo Nakajima. 2016. One Size Does Not Fit All: Applying the Right Game Concepts for the Right Persons to Encourage Non-game Activities. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 9735. Springer, 103–114. DOI:http://dx.doi.org/10.1007/978-3-319-40397-7_11

[6] William Albert and E. Dixon. 2003. Is This What You Expected? The Use of Expectation Measures in Usability Testing. *Proceedings of Usability Professionals Association Conference*. Usability Professionals Association.

[7] Noora Aldenaini, Felwah Alqahtani, Rita Orji and Srinivas Sampalli. 2020. Trends in Persuasive Technologies for Physical Activity and Sedentary Behavior: A Systematic Review. *Frontiers in Artificial Intelligence* 3. Frontiers, 7. DOI:http://dx.doi.org/10.3389/frai.2020.00007

[8] Maximilian Altmeyer, Kathrin Dernbecher, Vladislav Hnatovskiy, Marc Schubhan, Pascal Lessel and Antonio Krüger. 2019. Gamified Ads: Bridging the Gap Between User Enjoyment and the Effectiveness of Online Ads. *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, 1–12. DOI:http://dx.doi.org/10.1145/3290605.3300412

[9] Maximilian Altmeyer and Pascal Lessel. 2017. The Importance of Social Relations for Well-Being Change in Old Age - Do Game Preferences Change As Well? *Proceedings of the Positive Gaming: Workshop on Gamification and Games for Wellbeing (PG '17)*. CEUR-WS.

[10] Maximilian Altmeyer, Pascal Lessel, Seyedmostafa Hosseini and Antonio Krueger. 2018. SilverCycling: Evaluating Persuasive Strategies to Promote Physical Activity among Older Adults. *Adjunct Proceedings of the Designing Interactive Systems Conference (DIS '18)*. ACM, 45–50. DOI:http://dx.doi.org/10.1145/3197391.3205410

[11] Maximilian Altmeyer, Pascal Lessel, Subhashini Jantwal, Linda Muller, Florian Daiber and Antonio Krüger. 2021. Potential and Effects of Personalizing Gameful Fitness Applications Using Behavior Change Intentions and Hexad User Types. *User Modeling and User-Adapted Interaction* 31 (1). Springer, 675–712. DOI:http://dx.doi.org/10.1007/s11257-021-09288-6

[12] Maximilian Altmeyer, Pascal Lessel and Antonio Krüger. 2018. Investigating Gamification for Seniors Aged 75 +. In *Proceedings of the 2018 Designing Interactive Systems Conference (DIS '18)*. ACM, 453–458. DOI:http://dx.doi.org/10.1145/3196709.3196799

[13] Maximilian Altmeyer, Pascal Lessel, Linda Muller and Antonio Krüger. 2019. Combining Behavior Change Intentions and User Types to Select Suitable Gamification Elements for Persuasive Fitness Systems. In *Proceedings of the International Conference on Persuasive Technology (PERSUASIVE '19)*. Springer, 337–349.

[14] Maximilian Altmeyer, Pascal Lessel, Tobias Sander and Antonio Krüger. 2018. Extending a Gamified Mobile App with a Public Display to Encourage Walking. *Proceedings of the International Academic Mindtrek Conference (Mindtrek '18)*. ACM, 20–29. DOI:http://dx.doi.org/10.1145/3275116.3275135

[15] Maximilian Altmeyer, Pascal Lessel, Marc Schubhan, Vladislav Hnatovskiy and Antonio Krüger. 2019. Germ Destroyer – A Gamified System to Increase the Hand Washing Duration in Shared Bathrooms. In *Proceedings of the the Annual Symposium on Computer-Human Interaction in Play (CHI Play '19)*. ACM, 509–519. DOI:http://dx.doi.org/10.1145/3311350.3347157

[16] Maximilian Altmeyer, Pascal Lessel, Rehman Ur Atiq Waqar and Antonio Krüger. 2021. Design Guidelines to Increase the Persuasiveness of Achievement Goals for Physical Activity. *Proceedings of the International GamiFIN Conference (GamiFIN '21)*. CEUR-WS, 40–49.

[17] Maximilian Altmeyer, Marc Schubhan, Antonio Krüger and Pascal Lessel. 2021. A Long-Term Investigation on the Effects of (Personalized) Gamification on Course Participation in a Gym. *Proceedings of the International GamiFIN Conference (GamiFIN '21)*. CEUR-WS, 60–69.

[18] Maximilian Altmeyer, Marc Schubhan, Pascal Lessel and Antonio Krüger. 2019. Towards Predicting Hexad User Types from Smartphone Data. In *Extended Abstracts Proceedings of the Annual Symposium on Computer-Human Interaction in Play (CHI Play '19)*. ACM, 315–322. DOI:http://dx.doi.org/10.1145/3341215.3356266

[19] Maximilian Altmeyer, Marc Schubhan, Pascal Lessel, Linda Muller and Antonio Krüger. 2020. Using Hexad User Types to Select Suitable Gamification Elements to Encourage Healthy Eating. *Extended Abstracts Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '20)*. ACM, 1–8. DOI:http://dx.doi.org/10.1145/3334480.3383011

[20] Maximilian Altmeyer, Gustavo F. Tondello, Antonio Krüger and Lennart E. Nacke. 2020. HexArcade: Predicting Hexad User Types By Using Gameful Applications. *Proceedings of the the Annual Symposium on Computer-Human Interaction in Play (CHI Play '20)*. ACM, 219–230.

[21] Maximilian Altmeyer, Berina Zenuni, Hanne Spelt, Theierry Jegen, Pascal Lessel and Antonio Krüger. 2022. Do Hexad User Types Matter? Effects of (Non-) Personalized Gamification on Task Performance and User Experience in an Image Tagging Task. *Under Review at the Annual Symposium on Computer-Human Interaction in Play (CHI Play '22)*. UNDER REVIEW.

[22] Carole Ames. 1984. Competitive, Cooperative, and Individualistic Goal Structures: A Cognitive-Motivational Analysis. *Research on Motivation in Education* 1. Academic Press, 177–207.

[23] Mimi An. 2016. Why People Block Ads and What It Means for Marketers and Advertisers. *Hubspot Research*. https://blog.hubspot.com/marketing/why-people-block-ads-and-what-it-means-for-marketers-and-advertisers

[24] Evy L Ansems, Elcin Hanci, Peter Ruijten and Wijnand Ijsselsteijn. 2019. I Focus on Improvement: Effects of Type of Mastery Feedback on Motivational Experiences. *International Conference on Persuasive Technology Proceedings (PERSUASIVE '19)* 11433. Springer, 213–224. DOI:http://dx.doi.org/10.1007/978-3-030-17287-9

[25] Christopher J. Armitage, Paschal Sheeran, Mark Conner and Madelynne A. Arden. 2004. Stages of Change or Changes of Stage? Predicting Transitions in Transtheoretical Model Stages in Relation to Healthy Food Choice. *Journal of Consulting and Clinical Psychology* 72 (3). American Psychological Association, 491–499. DOI:http://dx.doi.org/10.1037/0022-006X.72.3.491

[26] Ernesto Arroyo, Leonardo Bonanni and Ted Selker. 2005. Waterbot: Exploring Feedback and Persuasive Techniques at the Sink. *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '05)*. ACM, 631–639. DOI:http://dx.doi.org/10.1145/1054972.1055059

[27] Toshiko Asai, Akinori Kanazawa, Hidehiko Hayashi and Akinori Minazuki. 2013. Development of a System to Raise Awareness of Hand Hygiene in Various Environments. *International Conference on Signal-Image Technology and Internet-Based Systems Proceedings (SITIS '13)*. IEEE, 924–931. DOI:http://dx.doi.org/10.1109/SITIS.2013.150

[28] Stefanie Ashford, Jemma Edmunds and David P French. 2010. What is the Best Way to Change Self-Efficacy to Promote Lifestyle and Recreational Physical Activity? A Systematic Review with Meta-Analysis. *British Journal of Health Psychology* 15 (2). Wiley, 265–288. DOI:http://dx.doi.org/10.1348/135910709X461752

[29] Milad Babolhavaeji, Mahnaz Asefpour Vakilian and Alireza Slambolchi. 2015. Color Preferences Based On Gender As a New Approach In Marketing. *Advanced Social Humanities and Management* 2 (1). E.M.T Publishing Company, 35–44.

[30] Rebecca Balebako, Jaeyeon Jung, Wei Lu, Lorrie Faith Cranor and Carolyn Nguyen. 2013. "Little Brothers Watching You": Raising Awareness of Data Leaks on Smartphones. *Symposium on Usable Privacy and Security Proceedings (SOUPS '13)*. ACM, 1–11. DOI:http://dx.doi.org/10.1145/2501604.2501616

[31] Dubravka Balen-Letunić. 2014. An Overview of Prehistoric Toys. *Etnološka istraživanja* 18/19. Etnografski muzej, 11–17.

[32] Albert Bandura. 1978. Self-Efficacy: Toward a Unifying Theory of Behavioral Change. *Advances in Behaviour Research and Therapy* 84 (3). American Psychological Association, 139–161. DOI:http://dx.doi.org/10.1016/0146-6402(78)90002-4

[33] Aaron Bangor, Philip Kortum and James Miller. 2009. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. *Journal of Usability Studies* 4 (3). Citeseer, 114–123.

[34] Richard Bartle. 1996. Hearts, Clubs, Diamonds, Spades: Players Who Suit MUDs. *Journal of MUD Research* 1 (1). 19.

[35] David R. Bassett, Holly R. Wyatt, Helen Thompson, John C. Peters and James O. Hill. 2010. Pedometer-Measured Physical Activity and Health Behaviors in U.S. Adults. *Medicine and Science in Sports and Exercise* 42 (10). American College of Sports Medicine, 1819–1825. DOI:http://dx.doi.org/10.1249/MSS.0b013e3181dc2e54

[36] Chris Bateman, Rebecca Lowenhaupt and Lennart E. Nacke. 2011. Player Typology in Theory and Practice. *DiGRA Conference Proceedings (DiGRA '11)*. Digital Games Research Association.

[37] Adrian E Bauman, Cora L Craig, Alison L Marshall, Michael Sjo, Michael L Booth, Barbara E Ainsworth, Michael Pratt, U L F Ekelund, Agneta Yngve, James F Sallis and Pekka Oja. 2003. International Physical Activity Questionnaire : 12-Country Reliability and Validity. *Medicine and Science in Sports and Exercise* 35 (8). American College of Sports Medicine, 1381–1395. DOI:http://dx.doi.org/10.1249/01.MSS.0000078924.61453.FB

[38] Michelle E Bayles. 2002. Designing Online Banner Advertisements: Should We Animate?. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '02)*. ACM, 363–366. DOI:http://dx.doi.org/10.1145/503376.503441

[39] Raoul Bell and Axel Buchner. 2018. Positive Effects of Disruptive Advertising on Consumer Preferences. *Journal of Interactive Marketing* 41. Elsevier, 1–13. DOI:http://dx.doi.org/10.1016/j.intmar.2017.09.002

[40] Yoav Benjamini and Yosef Hochberg. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society* 57 (1). Wiley, 289–300.

[41] John R. Best. 2013. Exergaming in Youth. *Zeitschrift fur Psychologie / Journal of Psychology* 221 (2). Hogrefe Publishing, 72–78. DOI:http://dx.doi.org/10.1027/2151-2604/a000137

[42] Max V. Birk, Cheralyn Atkins, Jason T. Bowey and Regan L. Mandryk. 2016. Fostering Intrinsic Motivation through Avatar Identification in Digital Games. *CHI Conference on Human Factors in Computing Systems Proceedings (CHI '16)*. ACM, 2982–2995. DOI:http://dx.doi.org/10.1145/2858036.2858062

[43] Max V. Birk, Maximilian A. Friehs and Regan L. Mandryk. 2017. Age-Based Preferences and Player Experience: A Crowdsourced Cross-sectional Study. *Proceedings of the the Annual Symposium on Computer-Human Interaction in Play (CHI Play '17)*. ACM, 157–170. DOI:http://dx.doi.org/10.1145/3116595.3116608

[44] Anthony J. Bishara and James B. Hittner. 2012. Testing the Significance of a Correlation with Nonnormal Data: Comparison of Pearson, Spearman, Transformation, and Resampling Approaches. *Psychological Methods* 17 (3). American Psychological Association, 399–417. DOI:http://dx.doi.org/10.1037/a0028087

[45] Jenny V. Bittner, Jeffrey Shipper and Jeffrey Schipper. 2014. Motivational Effects and Age Differences of Gamification in Product Advertising. *Journal of Consumer Marketing* 31 (5). Emerald, 391–400. DOI:http://dx.doi.org/10.1108/JCM-04-2014-0945

[46] Ian Bogost. 2005. The Rhetoric of Exergaming. *Proceedings of the Digital Arts and Cultures (DAC '05)*. Citeseer, 51–60.

[47] Marcela C.C. Bomfim, Sharon I. Kirkpatrick, Lennart E. Nacke and James R. Wallace. 2020. Food Literacy while Shopping: Motivating Informed Food Purchasing Behaviour with a Situated Gameful App. *Proceedings of the CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI '20)*. ACM, 1–13. DOI:http://dx.doi.org/10.1145/3313831.3376801

[48] Julia Ayumi Bopp, Elisa D. Mekler and Klaus Opwis. 2016. Negative Emotion, Positive Experience? *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, 2996–3006. DOI:http://dx.doi.org/10.1145/2858036.2858227

[49] Carl P Borchgrevink, JaeMin Cha and SeungHyun Kim. 2013. Hand Washing Practices in a College Town Environment. *Journal of Environmental Health* 75 (8). JSTOR, 18–25.

[50] Gunnar Borg. 1970. Perceived Exertion as an Indicator of Somatic Stress. *Scandinavian Journal of Rehabilitation Medicine* 2 (2). Almqvist & Wiksell Periodical Co., 92–98.

[51] Graciela Bouciguez, María José Santos and María José Abásolo. 2014. Towards the Use of Video Games for Learning: A Survey about Video Games Preferences of Engineering Students. *Journal of Computer Science and Technology* 14 (1). Springer, 25–31. http://sedici.unlp.edu.ar/handle/10915/34540

[52] Wolfram Boucsein. 2012. *Electrodermal Activity*. Springer Science & Business Media.

[53] Giorgio Brajnik and Silvia Gabrielli. 2010. A Review of Online Advertising Effects on the User Experience. *International Journal of Human-Computer Interaction* 26 (10). Taylor and Francis, 971–997. DOI:http://dx.doi.org/10.1080/10447318.2010.502100

[54] Virginia Braun and Victoria Clarke. 2012. Thematic Analysis. *APA Handbook of Research Methods in Psychology* 2. American Psychological Association, 57–71. DOI:http://dx.doi.org/10.1037/13620-004

[55] Philipp Brauner, Andre Valdez, Ulrik Schroeder and Martina Ziefle. 2013. Increase Physical Fitness and Create Health Awareness through Exergames and Gamification – The Role of Individual Factors, Motivation and Acceptance. *Human Factors in Computing and Informatics (SouthCHI '13)*. Springer, 349–362. DOI:http://dx.doi.org/10.1007/978-3-642-39062-3_22

[56] Dena Bravata, Crystal Smith-Spangler, Vandana Sundaram, Allison Gienger, Nancy Lin, Robyn Lewis, Christopher Stave, Ingram Olkin and John Sirard. 2007. Using Pedometers to Increase Physical Activity and Improve Health: A Systematic Review. *JAMA* 298 (19). American Medical Association, 2296–2304. DOI:http://dx.doi.org/10.1001/jama.298.19.2296

[57] Chris Bray. 2009. *Backgammon for Dummies*. Wiley.

[58] Whitney A Brechwald and Mitchell J Prinstein. 2011. Beyond Homophily: A Decade of Advances in Understanding Peer Influence Processes. *Journal of Research on Adolescence* 21 (1). Society for Research on Adolescence, 166–179. DOI:http://dx.doi.org/10.1111/j.1532-7795.2010.00721.x

[59] B. Breyer and M. Bluemke. 2016. Deutsche Version der Positive and Negative Affect Schedule PANAS. *GESIS Panel (GESIS '16)*. Zusammenstellung Sozialwissenschaftlicher Items und Skalen. DOI:http://dx.doi.org/10.6102/zis242

[60] John Brooke. 1996. *SUS - A Quick and Dirty Usability Scale*. Vol. 189. Taylor and Francis. 4–7 pages.

[61] M Brown. 2002. The Use of Banner Advertisements with Pull-Down Menus-A Copy Testing Approach. *Journal of Interactive Advertising* 2 (2). Taylor and Francis, 57–65. DOI:http://dx.doi.org/10.1080/15252019.2002.10722062

[62] Patrick Buckley and Elaine Doyle. 2017. Individualising Gamification: An Investigation of the Impact of Learning Styles and Personality Traits on the Efficacy of Gamification using a Prediction Market. *Computers and Education* 106. Elsevier, 43–55. DOI:http://dx.doi.org/10.1016/j.compedu.2016.11.009

[63] Kathryn L. Burgio, Bernard T. Engel and Julie L. Locher. 1991. Normative Patterns of Diurnal Urination Across 6 Age Decades. *The Journal of Urology* 145. American Urological Association, 728–731.

[64] Marc Busch, Elke Mattheiss, Wolfgang Hochleitner, Christina Hochleitner, Michael Lankes, Peter Fröhlich, Rita Orji and Manfred Tscheligi. 2016. Using Player Type Models for Personalized Game Design - An Empirical Investigation. *Interaction Design and Architectures* 28 (1). IxD&A, 145–163.

[65] Marc Busch, Elke Mattheiss, Rita Orji, Peter Fröhlich, Michael Lankes and Manfred Tscheligi. 2016. Player Type Models - Towards Empirical Validation. *Extended Abstracts Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, 1835–1841. DOI: http://dx.doi.org/10.1145/2851581.2892399

[66] Marc Busch, Elke Mattheiss, Michaela Reisinger, Rita Orji, Peter Fröhlich and Manfred Tscheligi. 2016. More than Sex: The Role of Femininity and Masculinity in the Design of Personalized Persuasive Games. *International Conference on Persuasive Technology (PERSUASIVE '16)*. Springer, 219–229. DOI:http://dx.doi.org/10.1007/978-3-319-31510-2_19

[67] Ruth Butler. 1992. What Young People want to Know When: Effects of Mastery and Ability Goals on Interest in Different Kinds of Social Comparisons. *Journal of Personality and Social Psychology* 62 (6). American Psychological Association, 934–943. DOI:http://dx.doi.org/10.1037/0022-3514.62.6.934

[68] Fabio Buttussi, Luca Chittaro, Roberto Ranon and Alessandro Verona. 2007. Adaptation of Graphics and Gameplay in Fitness Games by Exploiting Motion and Physiological Sensors. *Proceedings of the International Symposium on Smart Graphics (SG '07)*. Springer, 85–96. DOI:http://dx.doi.org/10.1007/978-3-540-73214-3_8

[69] John Cacioppo, Louis G. Tassinary and Gary G. Berntson. 2007. *The Handbook of Psychophysiology*. Cambridge University Press.

[70] Lisa A. Cadmus-Bertram, Bess H. Marcus, Ruth E. Patterson, Barbara A. Parker and Brittany L. Morey. 2015. Randomized Trial of a Fitbit-Based Physical Activity Intervention for Women. *American Journal of Preventive Medicine* 49 (3). Elsevier, 414–418. DOI:http://dx.doi.org/10.1016/j.amepre.2015.01.020

[71] Damon E Campbell and Ryan T Wright. 2008. Shut-Up I Don't Care: Understanding the Role of Relevance and Interactivity on Customer Attitudes Toward Repetitive Online Advertising. *Journal of Electronic Commerce Research* 9 (1). 62–76.

[72] Centers for Disease Control and Prevention. 2018. How to Wash Your Hands. https://www.cdc.gov/features/handwashing/index.html

[73] Christopher P. Cerasoli, Jessica M. Nicklin and Michael T. Ford. 2014. Intrinsic Motivation and Extrinsic Incentives Jointly Predict Performance: A 40-Year Meta-Analysis. *Psychological Bulletin* 140 (4). American Psychological Association, 980–1008. DOI:http://dx.doi.org/10.1037/a0035661

[74] Robert Cercos and Florian Floyd Mueller. 2013. Watch your Steps: Designing a Semi-Public Display to Promote Physical Activity. *Proceedings of the Australasian Conference on Interactive Entertainment (IE '13)*. ACM, 1–6. DOI:http://dx.doi.org/10.1145/2513002.2513016

[75] Sainabou Cham, Abdullah Algashami, John McAlaney, Angelos Stefanidis, Keith Phalp and Raian Ali. 2019. Goal Setting for Persuasive Information Systems: Five Reference Checklists. *Proceedings of the International Conference on Persuasive Technology (PERSUASIVE '19)*. Springer, 237–253. DOI:http://dx.doi.org/10.1007/978-3-030-17287-9_20

[76] Kerry Shih Ping Chang, Catalina M. Danis and Robert G. Farrell. 2014. Lunch Line: Using Public Displays and Mobile Devices to Encourage Healthy Eating in an Organization. *Proceedings of the International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '14)*. ACM, 823–834. DOI:http://dx.doi.org/10.1145/2632048.2636086

[77] Patrali Chatterjee. 2008. Are Unclicked Ads Wasted? Enduring Effects of Banner and Pop-Up Ad Exposures on Brand Memory and Attitudes. *Journal of Electronic Commerce Research* 9 (1). College of Commerce, 51–61.

[78] Qimei Chen and William D. Wells. 1999. Attitude Toward the Site. *Journal of Advertising Research* 39 (5). World Advertising Research Center Ltd., 27–38.

[79] Yu Chen and Pearl Pu. 2014. HealthyTogether: Exploring Social Incentives for Mobile Fitness Applications. *Proceedings of the Second International Symposium of Chinese CHI (Chinese CHI '14)*. ACM, 25–34. DOI:http://dx.doi.org/10.1145/2592235.2592240

[80] Gokul Chittaranjan, Jan Blom and Daniel Gatica-Perez. 2011. Who's Who with Big-Five: Analyzing and Classifying Personality Traits with Smartphones. *Proceedings of the Annual International Symposium on Wearable Computers (ISWC '11)*. IEEE, 29–36. DOI:http://dx.doi.org/10.1109/ISWC.2011.29

[81] Luca Chittaro and Riccardo Sioni. 2012. Turning the Classic Snake Mobile Game into a Location – Based Exergame that Encourages Walking. *Proceedings of the International Conference on Persuasive Technology (PERSUASIVE '12)*. Springer, 43–54. DOI:http://dx.doi.org/10.1007/978-3-642-31037-9_4

[82] Meng-chieh Chiu, Shih-ping Chang, Yu-chen Chang, Hao-hua Chu, Cheryl Chia-hui Chen, Fei-hsiu Hsiao and Ju-chun Ko. 2009. Playful Bottle: A Mobile Social Persuasion System to Motivate Healthy Water Intake. *Proceedings of the International Conference on Ubiquitous Computing (UbiComp '09)*. ACM, 185–194.

[83] Chang Hoan Cho and Hongsik John Cheon. 2004. Why Do People Avoid Advertising on the Internet? *Journal of Advertising* 33 (4). Taylor and Francis, 89–97. DOI:http://dx.doi.org/10.1080/00913367.2004.10639175

[84] Yu-kai Chou. 2019. *Actionable Gamification: Beyond Points, Badges, and Leaderboards*. Packt Publishing.

[85] Ching Yue Chow, Reisya Rizki Riantiningtyas, Mie Bojer Kanstrup, Maria Papavasileiou, Gie Djin Liem and Annemarie Olsen. 2020. Can Games Change Children's Eating Behaviour? A Review of Gamification and Serious Games. *Food Quality and Preference* 80. Elsevier, 1–16. DOI:http://dx.doi.org/10.1016/j.foodqual.2019.103823

[86] Jacob Cohen. 1988. *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates.

[87] Sunny Consolvo, Katherine Everitt, Ian Smith and James A Landay. 2006. Design Requirements for Technologies that Encourage Physical Activity. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '06)*. ACM, 457–466. DOI:http://dx.doi.org/10.1145/1124772.1124840

[88] Sunny Consolvo, Predrag Klasnja, David W Mcdonald, Daniel Avrahami, Jon Froehlich, Louis Legrand, Ryan Libby, Keith Mosher and James A Landay. 2008. Flowers or a Robot Army? Encouraging Awareness and Activity with Personal, Mobile Displays. *Proceedings of the International Conference on Ubiquitous Computing (Ubicomp '08)*. ACM, 54–63. DOI:http://dx.doi.org/10.1145/1409635.1409644

[89] Sunny Consolvo, David W Mcdonald, Tammy Toscos, Mike Y Chen, Jon Froehlich, Beverly Harrison, Predrag Klasnja, Anthony Lamarca, Louis Legrand, Ryan Libby, Ian Smith and James A Landay. 2008. Activity Sensing in the Wild: A Field Trial of UbiFit Garden. *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '08)*. ACM, 1797–1806. DOI:http://dx.doi.org/10.1145/1357054.1357335

[90] Francesco Corato, Maria Frucci and Gabriella Sanniti Di Baja. 2012. Virtual Training of Surgery Staff for Hand Washing Procedure. *Proceedings of the International Working Conference on Advanced Visual Interfaces (AVI '12)*. ACM, 274–277. DOI:http://dx.doi.org/10.1145/2254556.2254608

[91] Benjamin Cornwell, Edward O Laumann and L Philip Schumm. 2008. The Social Connectedness of Older Adults: A National Profile. *American Sociological Review* 73. Sage, 185–203.

[92] Henriette Cramer. 2015. Effects of Ad Quality and Content-Relevance on Perceived Content Quality. *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '15)* 1. ACM, 2231–2234. DOI:http://dx.doi.org/10.1145/2702123.2702360

[93] Mihaly Csikszentmihalyi. 1975. *Beyond Boredom and Anxiety*. Jossey-Bass.

[94] Mihaly Csikszentmihalyi. 1997. *Finding Flow: The Psychology of Engagement with Everyday Life*. Basic Books.

[95] Mihaly Csikszentmihalyi, Sami Abuhamdeh and Jeanne Nakamura. 2014. Flow. In *Flow and the Foundations of Positive Psychology: The Collected Works of Mihaly Csikszentmihalyi*. Springer, 227–238. DOI:http://dx.doi.org/10.1007/978-94-017-9088-8

[96] Randall J Curtis, Marjorie D Wenrich, Jan D Carline, Sarah E Shannon, Donna M Ambrozy and Paul G Ramsey. 2001. Understanding Physicians' Skills at Providing End-of-Life Care: Perspectives of Patients, Families, and Health Care Workers. *Journal of General Internal Medicine* 16 (1). Springer, 41–49. DOI:http://dx.doi.org/10.1046/j.1525-1497.2001.00333.x

[97] Val Curtis, Wolf Schmidt, Stephen Luby, Rocio Florez, Ousmane Touré and Adam Biran. 2011. Hygiene: New Hopes, New Horizons. *The Lancet Infectious Diseases* 11 (4). Elsevier, 312–321. DOI:http://dx.doi.org/10.1016/S1473-3099(10)70224-3

[98] Stephen Cuzzort and Thad Starner. 2008. AstroWheelie: A Wheelchair Based Exercise Game. *Proceedings of the International Symposium on Wearable Computers (ISWC '08)*. IEEE, 113–114. DOI:http://dx.doi.org/10.1109/ISWC.2008.4911599

[99] Marco de Sa, Vidhya Navalpakkam and Elizabeth F. Churchill. 2013. Mobile Advertising: Evaluating the Effects of Animation, User and Content Relevance. *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, 2487–2496. DOI:http://dx.doi.org/10.1145/2470654.2481344

[100] Bob De Schutter and Vero Vanden Abeele. 2010. Designing Meaningful Play Within the Psycho-Social Context of Older Adults. *Proceedings of the International Conference on Fun and Games (Fun and Games '10)*. ACM, 84–93. DOI:http://dx.doi.org/10.1145/1823818.1823827

[101] Edward L. Deci. 1971. Effects of Externally Mediated Rewards on Intrinsic Motivation. *Journal of Personality and Social Psychology* 18 (1). American Psychological Association, 105–115. DOI:http://dx.doi.org/10.1037/h0030644

[102] Mouna Denden, Ahmed Tlili, Fathi Essalmi and Mohamed Jemni. 2017. An Investigation of the Factors Affecting the Perception of Gamification and Game Elements. *Proceedings of the International Conference on Information and Communication Technology and Accessbility (ICTA '17)*. IEEE, 1–5. DOI:http://dx.doi.org/10.1109/ICTA.2017.8336019

[103] Sameer Deshpande, Michael D. Basil and Debra Z. Basil. 2009. Factors Influencing Healthy Eating Habits Among College Students: An Application of the Health Belief Model. *Health Marketing Quarterly* 26 (2). Taylor and Francis, 145–164. DOI:http://dx.doi.org/10.1080/07359680802619834

[104] Sebastian Deterding. 2012. Gamification: Designing for Motivation. *Interactions* 19. ACM. DOI:http://dx.doi.org/10.1145/2212877.2212883

[105] Sebastian Deterding, Dan Dixon, Rilla Khaled and Lennart Nacke. 2011. From Game Design Elements to Gamefulness: Defining Gamification. *Proceedings of the International Academic MindTrek Conference (Mindtrek '11)*. ACM, 9–15. DOI:http://dx.doi.org/10.1145/2181037.2181040

[106] Sebastian Deterding, Miguel Sicart, Lennart E. Nacke, Kenton O'Hara and Dan Dixon. 2011. Gamification: Using Game Design Elements in Non-Gaming Contexts. *Extended Abstracts Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, 2425–2428.

[107] Dan Dixon. 2011. Player Types and Gamification. *Proceedings of the CHI 2011 Workshop on Gamification*. Gamification Research, 7–12. http://gamification-research.org/wp-content/uploads/2011/04/11-Dixon.pdf

[108] Julie Doyle, Daniel Kelly and Brian Caulfield. 2011. Design Considerations in Therapeutic Exergaming. *Proceedings of the International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth '11)*. IEEE, 389–393. DOI:http://dx.doi.org/10.4108/icst.pervasivehealth.2011.246115

[109] Julie Doyle, Daniel Kelly, Matt Patterson and Brian Caulfield. 2011. The Effects of Visual Feedback in Therapeutic Exergaming on Motor Task Accuracy. *Proceedings of the International Conference on Virtual Rehabilitation (ICVR '11)*. IEEE. DOI:http://dx.doi.org/10.1017/CBO9781107415324.004

[110] Anders Drachen, Georgios Yannakakis, Lennart E. Nacke and Anja Lee Pedersen. 2010. Correlation Between Heart Rate, Electrodermal Activity and Player Experience in First-Person Shooter Games. *Proceedings of the SIGGRAPH Symposium on Video Games (Sandbox '10)*. ACM, 49–54. DOI: http://dx.doi.org/10.1145/1836135.1836143

[111] Filip Drozd, Tuomas Lehto and Harri Oinas-Kukkonen. 2012. Exploring Perceived Persuasiveness of a Behavior Change Support System: A Structural Model. *Proceedings of the International Conference on Persuasive Technology (PERSUASIVE '12)*. Springer, 157–168. DOI:http://dx.doi.org/10.1007/978-3-642-31037-9_14

[112] Carol S. Dweck. 1986. Motivational Processes Affecting Learning. *American Psychologist* 41 (10). American Psychological Association, 1040–1048. DOI: http://dx.doi.org/10.1037//0003-066x.41.10.1040

[113] Patricia E. Easterling. 1997. *The Cambridge Companion to Greek Tragedy*. Cambridge University Press.

[114] Regina I Ejemot, John E Ehiri, Martin M Meremikwu and Julia A Critchley. 2008. Hand Washing for Preventing Diarrhoea. *Cochrane Database of Systematic Reviews 2008*, 4 (2). Wiley, 893–939. DOI:http://dx.doi.org/10.1002/14651858.CD004265.pub2

[115] Fatema El-Amrawy, B Pharm and Mohamed Ismail Nounou. 2015. Are Currently Available Wearable Devices for Activity Tracking and Heart Rate Monitoring Accurate, Precise, and Medically Beneficial? *Healthcare Informatics Research* 21 (4). The Korean Society of Medical Informatics, 315–320. DOI: http://dx.doi.org/10.4258/hir.2015.21.4.315

[116] Andrew J. Elliot and Judith M. Harackiewicz. 1996. Approach and Avoidance Achievement Goals and Intrinsic Motivation: A Mediational Analysis. *Journal of Personality and Social Psychology* 70 (3). American Psychological Association, 461–475.

[117] Andrew J. Elliot and H.A. McGregor. 2001. A 2x2 Achievement Goal Framework. *Journal of Personality and Social Psychology* 80 (3). American Psychological Association, 501–519. DOI:http://dx.doi.org/10.1037//0022-3514.80.3.501

[118] Andrew J. Elliot, Kou Murayama and Reinhard Pekrun. 2011. A 3x2 Achievement Goal Model. *Journal of Educational Psychology* 103 (3). American Psychological Association, 632–648. DOI:http://dx.doi.org/10.1037/a0023952

[119] Daniel Epstein, An Ping, James Fogarty and Sean A Munson. 2015. A Lived Informatics Model of Personal Informatics. *Proceedings of the International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. ACM, 731–742. DOI:http://dx.doi.org/10.1145/2750858.2804250

[120] Richard M Felder and Linda K. Silverman. 1988. Learning and Teaching Styles In Engineering Education. *Engineering Education* 78 (7). Citeseer, 674–681.

[121] Christopher J. Ferguson. 2009. An Effect Size Primer: A Guide for Clinicians and Researchers. *Professional Psychology: Research and Practice* 40 (5). American Psychological Association, 532–538. DOI:http://dx.doi.org/10.1037/a0015808

[122] Fernando T Fernandes and Plinio T. A. Junior. 2016. Gamification Aspects in the Context of Electronic Government and Education: A Case Study. *Proceedings of the International Conference on HCI in Business, Government, and Organizations (HCIBGO '16)* 9752. Springer. DOI:http://dx.doi.org/10.1007/978-3-319-39399-5

[123] Andy Field. 2013. *Discovering Statistics Using IBM SPSS Statistics*. Sage Publications.

[124] Samantha L. Finkelstein, Andrea Nickel, Tiffany Barnes and Evan A. Suma. 2010. Astrojumper: Designing a Virtual Reality Exergame to Motivate Children with Autism to Exercise. *Proceedings of the Virtual Reality Conference (VR '10)*. IEEE, 267–268. DOI:http://dx.doi.org/10.1109/VR.2010.5444770

[125] William Flores, Jeng Chung Victor Chen and William H. Ross. 2014. The Effect of Variations in Banner Ad, Type of Product, Website Context, and Language of Advertising on Internet Users' Attitudes. *Computers in Human Behavior* 31 (1). Elsevier Ltd, 37–47. DOI:http://dx.doi.org/10.1016/j.chb.2013.10.006

[126] B J Fogg. 2002. *Persuasive Technology: Using Computers to Change What We Think and Do*. Morgan Kaufmann Publishers.

[127] Derek Foster, Conor Linehan, Ben Kirman, Shaun Lawson, Gary James and Bracebridge Heath. 2010. Motivating Physical Activity at Work: Using Persuasive Social Media for Competitive Step Counting. *Proceedings of the International Academic MindTrek Conference (Mindtrek '10)*. ACM, 111–116. DOI:http://dx.doi.org/10.1145/1930488.1930510

[128] Sigmund Freud. 1915. Triebe und Triebschicksale. *Internationale Zeitschrift für Psychoanalyse* 3 (2). Internationaler Psychoanalytischer Verlag, 84–100.

[129] Marco Furini. 2007. Mobile Games: What to Expect in the Near Future. *Proceedings of the International Conference on Intelligent Games and Simulation (GAMEON '07)*. Elsevier, 93–95.

[130] Carsten Fuß, Tim Steuer, Kevin Noll and André Miede. 2014. Teaching the Achiever, Explorer, Socializer, and Killer - Gamification in University Education. *Proceedings of the International Conference on Serious Games (Gamedays '14)*. Springer, 92–99. DOI:http://dx.doi.org/10.1007/978-3-319-05972-3_11

[131] Luciano Gamberini, Mariano Alcaniz, Giacinto Barresi, Malena Fabregat, Francisco Ibanez and Lisa Prontu. 2006. Cognition , Technology and Games for the Elderly: An Introduction to the ELDERGAMES Project. *PsychNology Journal* 4 (3). PNJ, 285–308.

[132] Gartner. 2012. Gartner says by 2014, 80 percent of current gamified applications will fail to meet business objectives primarily due to poor design. http://www.gartner.com/newsroom/id/2251015

[133] Kathrin Maria Gerling and Maic Masuch. 2011. Exploring the Potential of Gamification Among Frail Elderly Persons. *Proceedings of the CHI 2011 Workshop on Gamification*. Gamification Research, 1–4.

[134] Kathrin Maria Gerling, Frank Paul Schulte, Jan Smeddinck and Maic Masuch. 2011. Designing and Evaluating Digital Games for Frail Elderly Persons. *Proceedings of the International Conference on Advances In Computer Entertainment Technology (ACE '11)*. ACM, 1–8.

[135] Kathrin Maria Gerling, Frank Paul Schulte, Jan Smeddinck and Maic Masuch. 2012. Game Design for Older Adults: Effects of Age-Related Changes on Structural Elements of Digital Games. *International Conference on Entertainment Computing (ICEC'12)*. Springer, 235–242.

[136] Nicolas Gillet, Marc André K. Lafrenière, Tiphaine Huyghebaert and Evelyne Fouquereau. 2015. Autonomous and Controlled Reasons Underlying Achievement Goals: Implications for the 3x2 Achievement Goal Model in Educational and Work Settings. *Motivation and Emotion* 39 (6). Springer, 858–875. DOI:http://dx.doi.org/10.1007/s11031-015-9505-y

[137] Xinning Gui, Yu Chen, Clara Caldeira, Dan Xiao and Yunan Chen. 2017. When Fitness Meets Social Networks: Investigating Fitness Tracking and Social Practices on WeRun. *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, 1647–1659. DOI:http://dx.doi.org/10.1145/3025453.3025654

[138] Maryellen E Guinan, Maryanne Mcguckin-guinan and M T Ascp. 1997. Who Washes Hands After Using the Bathroom? *American Journal of Infection Control* 25. Association for Professionals in Infection Control and Epidemiology, 424–425.

[139] Michael Gurr, Wim Saris, Eric Jéquier, Peter Zock, Anthony Diplock, Nils-Georg Asp, Stylianos Nicolaidis, David Conning, Adrianne Hardman and Klaus König. 1998. *Healthy Lifestyles: Nutrition and Physical Activity*. ILSI Europe.

[140] Fabian Hadiji, Rafet Sifa, Anders Drachen, Christian Thurau, Kristian Kersting and Christian Bauckhage. 2014. Predicting Player Churn in the Wild. *Proceedings of the Conference on Computatonal Intelligence and Games (CIG '14)*. IEEE, 1–8. DOI:http://dx.doi.org/10.1109/CIG.2014.6932876

[141] Sajanee Halko and Julie A Kientz. 2010. Personality and Persuasive Technology: An Exploratory Study on Health-Promoting Mobile Applications. *International Conference on Persuasive Technology Proceedings (PERSUASIVE '10)*. Springer, 150–161. DOI:http://dx.doi.org/10.1007/978-3-642-13226-1

[142] Kara L. Hall and Joseph S. Rossi. 2008. Meta-Analytic Examination of the Strong and Weak Principles across 48 Health Behaviors. *Preventive Medicine* 46 (3). Elsevier, 266–274. DOI:http://dx.doi.org/10.1016/j.ypmed.2007.11.006

[143] Stuart Hallifax, Elise Lavoué and Audrey Serna. 2020. To Tailor or not to Tailor Gamification? An Analysis of the Impact of Tailored Game Elements on Learners' Behaviours and Motivation. *Proceedings of the International Conference on Artificial Intelligence in Education (AIED '20)*. Springer, 216–227. DOI:http://dx.doi.org/10.1007/978-3-030-52237-7_18

[144] Stuart Hallifax, Audrey Serna, Jean-charles Marty, Guillaume Lavoué and Elise Lavoué. 2019. Factors to Consider for Tailored Gamification. *Proceedings of the Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '19)*. ACM, 559–572.

[145] Juho Hamari, Jonna Koivisto and Tuomas Pakkanen. 2014. Do Persuasive Technologies Persuade? - A Review of Empirical Studies. *Proceedings of the International Conference on Persuasive Technology (PERSUASIVE '14)*. Springer, 118–136. DOI:http://dx.doi.org/10.1007/978-3-319-07127-5_11

[146] Juho Hamari, Jonna Koivisto and Harri Sarsa. 2014. Does Gamification Work? - A Literature Review of Empirical Studies on Gamification. *Proceedings of the Hawaii International Conference on System Sciences (HICSS '14)*. Association for Information Systems, 3025–3034. DOI:http://dx.doi.org/10.1109/HICSS.2014.377

[147] Michael D. Hanus and Jesse Fox. 2015. Assessing the Effects of Gamification in the Classroom: A Longitudinal Study on Intrinsic Motivation, Social comparison, Satisfaction, Effort, and Academic Performance. *Computers and Education* 80. Elsevier Ltd, 152–161. DOI:http://dx.doi.org/10.1016/j.compedu.2014.08.019

[148] Johannes Harms, Stefan Biegler, Christoph Wimmer, Karin Kappel and Thomas Grechenig. 2015. Gamification of Online Surveys: Design Process, Case Study, and Evaluation. *Human-Computer Interaction (INTERACT '15)*. Springer, 219–236. DOI:http://dx.doi.org/10.1007/978-3-319-22701-6_16

[149] Johannes Harms, Dominik Seitz, Christoph Wimmer, Karin Kappel and Thomas Grechenig. 2015. Low-Cost Gamification of Online Surveys: Improving the User Experience through Achievement Badges. *Proceedings of the Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '15)*. ACM, 109–114. DOI:http://dx.doi.org/10.1145/2793107.2793146

[150] Johannes Harms, Christoph Wimmer, Karin Kappel and Thomas Grechenig. 2014. Gamification of Online Surveys: Conceptual Foundations and a Design Process based on the MDA Framework. *Proceedings of the Nordic Conference on Human-Computer Interaction (NordiCHI '14)*. ACM, 565–568. DOI:http://dx.doi.org/10.1145/2639189.2639230

[151] Daniel Harrison and Paul Marshall. 2014. Tracking Physical Activity: Problems Related to Running Longitudinal Studies with Commercial Devices. *Adjunct Proceedings of the International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '14)*. ACM, 699–702. DOI:http://dx.doi.org/10.1145/2638728.2641320

[152] Jutta Heckhausen and Richard Schulz. 1995. A Life-Span Theory of Control. *Psychological Review* 102 (2). American Psychological Association, 284–304. DOI:http://dx.doi.org/10.1037/0033-295X.102.2.284

[153] A. Regula Herzog and Jerald G. Bachman. 1981. Effects of Questionnaire Length on Response Quality. *Public Opinion Quarterly* 45 (4). Elsevier, 549–559. DOI:http://dx.doi.org/10.1086/268687

[154] Gail D. Heyman and Carol S. Dweck. 1992. Achievement Goals and Intrinsic Motivation: Their Relation and their Role in Adaptive Motivation. *Motivation and Emotion* 16 (3). Springer, 231–247. DOI:http://dx.doi.org/10.1007/BF00991653

[155] David Howell. 2002. *Statistical Methods For Psychology*. Duxbury. DOI:http://dx.doi.org/10.2307/2348956

[156] Hsiu-Fang Hsieh and Sarah Shannon. 2005. Three Approaches to Qualitative Content Analysis. *Qualitative Health Research* 15 (9). Sage, 1277–1288. DOI:http://dx.doi.org/10.1177/1049732305276687

[157] Mark J. Huiskes and Michael S. Lew. 2008. The MIR Flickr Retrieval Evaluation. *Proceedings of the International Conference on Multimedia Information Retrieval (MIR '08)*. ACM, 39–43. DOI:http://dx.doi.org/10.1145/1460096.1460104

[158] J. Huizinga. 1949. *Homo Ludens: A Study of the Play-Element in Culture*. ROUTLEDGE & KEGAN PAUL.

[159] Clark L. Hull. 1943. *Principles of Behavior. An Introduction to Behavior Theory*. Appleton-Century Company Inc. DOI:http://dx.doi.org/10.2307/2019960

[160] Clark L. Hull. 1952. *A Behavior System: An Introduction to Behavior Theory Concerning the Individual Organism*. Yale University Press. DOI:http://dx.doi.org/10.2307/1418653

[161] Kai Huotari and Juho Hamari. 2012. Defining Gamification. *Proceedings of the International Academic Mindtrek Conference (Mindtrek '12)*. ACM, 17–22. DOI:http://dx.doi.org/10.1145/2393132.2393137

[162] Alice M. Isen and Johnmarshall Reeve. 2005. The Influence of Positive Affect on Intrinsic and Extrinsic Motivation: Facilitating Enjoyment of Play, Responsible Work Behavior, and Self-Control. *Motivation and Emotion* 29 (4). Springer, 297–325. DOI:http://dx.doi.org/10.1007/s11031-006-9019-8

[163] Rosemary J Thomas, Judith Masthoff and Nir Oren. 2019. Can I Influence You? Development of a Scale to Measure Perceived Persuasiveness and two Studies showing the Use of the Scale. *Frontiers in Artificial Intelligence* 2. Frontiers, 1–14. DOI:http://dx.doi.org/10.3389/FRAI.2019.00024

[164] William James. 1890. *The Principles of Psychology, Volume 2*. Macmillan.

[165] Wilfrid Jänig. 2003. The Autonomic Nervous System and its Coordination by the brain. In *Handbook of Affective Sciences*, Richard J Davidson, Klaus R Scherer, and H Goldsmith (Eds.). Oxford University Press.

[166] Dane A. Jensen, Michelle D. Danyluk, Linda J. Harris and Donald W. Schaffner. 2015. Quantifying the Effect of Hand Wash Duration, Soap Use, Ground Beef Debris, and Drying Methods on the Removal of Enterobacter aerogenes on Hands. *Journal of Food Protection* 78 (4). International Association for Food Protection, 685–690. DOI:http://dx.doi.org/10.4315/0362-028x.jfp-14-245

[167] Yuan Jia, Yikun Liu, Xing Yu and Stephen Voida. 2017. Designing Leaderboards for Gamification: Perceived Differences based on User Ranking, Application domain, and Personality traits. *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, 1949–1960. DOI:http://dx.doi.org/10.1145/3025453.3025826

[168] Yuan Jia, Bin Xu, Yamini Karanam and Stephen Voida. 2016. Personality-Targeted Gamification: A Survey Study on Personality Traits and Motivational Affordances. *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, 2001–2013. DOI:http://dx.doi.org/10.1145/2858036.2858515

[169] Durell Johnson, Danielle Sholcosky, Karen Gabello, Robert Ragni and Nicole Ogonosky. 2003. Sex Differences in Public Restroom Handwashing Behavior Associated with Visual Behavior Prompts. *Perceptual and Motor Skills* 97 (7). Sage, 805. DOI:http://dx.doi.org/10.2466/pms.97.7.805-810

[170] Brooke A. Jones, Gregory J. Madden and Heidi J. Wengreen. 2014. The FIT Game: Preliminary Evaluation of a Gamification Approach to Increasing Fruit and Vegetable Consumption in School. *Preventive Medicine* 68. Elsevier, 76–79. DOI:http://dx.doi.org/10.1016/j.ypmed.2014.04.015

[171] Barry H. Kantowitz, Henry L. Roediger and David G. Elmes. 2009. *Experimental Psychology* (9 ed.). Wadsworth Publishing Co Inc.

[172] Dennis L Kappen, Lennart E Nacke, Kathrin M Gerling and Lia E Tsotsos. 2016. Design Strategies for Gamified Physical Activity Applications for Older Adults. *Proceedings of the Hawaii International Conference on System Sciences (HICSS '16)*. Association for Information Systems, 1309–1318. DOI:http://dx.doi.org/10.1109/HICSS.2016.166

[173] Johannes Keller, Herbert Bless, Frederik Blomann and Dieter Kleinböhl. 2011. Physiological Aspects of Flow Experiences: Skills-Demand-Compatibility Effects on Heart Rate Variability and Salivary Cortisol. *Journal of Experimental Social Psychology* 47 (4). Elsevier, 849–852. DOI:http://dx.doi.org/10.1016/j.jesp.2011.02.004

[174] Kerry Kennedy and Thomas K. Gregoire. 2009. Theories of Motivation in Addiction Treatment: Testing the Relationship of the Transtheoretical Model of Change and Self-Determination Theory. *Journal of Social Work Practice in the Addictions* 9 (2). Taylor and Francis, 163–183. DOI:http://dx.doi.org/10.1080/15332560902852052

[175] Florian Keusch and Chan Zhang. 2017. A Review of Issues in Gamified Surveys. *Social Science Computer Review* 35 (2). Sage, 147–166. DOI:http://dx.doi.org/10.1177/0894439315608451

[176] Jong Uk Kim, Woong Jin Kim and Sang Cheol Park. 2010. Consumer Perceptions on Web Advertisements and Motivation Factors to Purchase in the Online Shopping. *Computers in Human Behavior* 26 (5). Elsevier, 1208–1222. DOI:http://dx.doi.org/10.1016/j.chb.2010.03.032

[177] Jeffrey Y Kim, Jonathan P Allen and Elan Lee. 2008. Alternate Realtity Gaming. *Commun. ACM* 51 (2). ACM, 36–42. DOI:http://dx.doi.org/10.1145/1314215.1340912

[178] Andrea Kinnison, Randal R. Cottrell and Keith A. Ki. 2004. Proper Hand-Washing Techniques in Public Restrooms: Differences in Gender, Race, Signage, and Time of Day. *American Journal of Health Education* 35 (2). Taylor and Francis, 86–89. DOI:http://dx.doi.org/10.1080/19325037.2004.10603614

[179] Michel C.A. Klein, Adnan Manzoor and Julia S. Mollee. 2017. Active2Gether: A Personalized M-Health Intervention to Encourage Physical Activity. *Sensors* 17 (6). MDPI, 1436–1452. DOI:http://dx.doi.org/10.3390/s17061436

[180] Paul R. Kleinginna and Anne M. Kleinginna. 1981. A Categorized List of Motivation Definitions, with a Suggestion for a Consensual Definition. *Motivation and Emotion* 5 (3). Springer, 263–291. DOI:http://dx.doi.org/10.1007/BF00993889

[181] Rex B. Kline. 2011. *Principles and Practice of Strctural Equation Modeling*. The Guilford Press.

[182] Ana Carolina Tomé Klock, Isabela Gasparini, Marcelo Soares Pimenta and Juho Hamari. 2020. Tailored Gamification: A Review of Literature. *International Journal of Human Computer Studies* 144. Elsevier, 1–22. DOI:http://dx.doi.org/10.1016/j.ijhcs.2020.102495

[183] Michael T. Knierim, Raphael Rissler, Verena Dorner, Alexander Maedche and Christof Weinhardt. 2017. The Psychophysiology of Flow: A Systematic Review of Peripheral Nervous System Features. *Lecture Notes in Information Systems and Organisation* 25. Springer, 109–120. DOI:http://dx.doi.org/10.1007/978-3-319-67431-5_13

[184] Henry C Koerper and Nancy a Whitney-Desautels. 1999. Astragalus Bones: Artifacts Or Ecofacts? *Pacific Coast Archaeological Society Quarterly* 35 (2-3). PCAS, 69–80.

[185] Jonna Koivisto and Juho Hamari. 2019. Gamification of Physical Activity: A Systematic Literature Review of Comparison Studies. *Proceedings of the International GamiFIN Conference (GamiFIN '19)*. CEUR-WS, 106–117.

[186] Jonna Koivisto and Juho Hamari. 2019. The Rise of Motivational Information Systems: A Review of Gamification Research. *International Journal of Information Management* 45. Elsevier, 191–210. DOI:http://dx.doi.org/10.1016/j.ijinfomgt.2018.10.013

[187] Sebastian Korbas. 2015. Gamification User Types zur Unterstützung der Konzeption in der Softwareentwicklung. *Proceedings of the Mensch und Computer Conference (MuC '15)*. De Gruyter, 367–376.

[188] Felix Kosmalla, Christian Murlowski, Florian Daiber and Antonio Krüger. 2018. Slackliner – An Interactive Slackline Training Assistant. *Proceedings of the International Conference on Multimedia (MM '18)*. ACM, 154–162. DOI: http://dx.doi.org/10.1145/3240508.3240537

[189] Dimosthenis Kotsopoulos, Cleopatra Bardaki, Stavros Lounis and Katerina Pramatari. 2018. Employee Profiles and Preferences towards IoT-enabled Gamification for Energy Conservation. *International Journal of Serious Games* 5 (2). IJSG, 65–85. DOI:http://dx.doi.org/10.17083/ijsg.v5i2.225

[190] Sylvia D Kreibig. 2010. Autonomic Nervous System Activity in Emotion: A Review. *Biological Psychology*. Elsevier, 394–421. DOI:http://dx.doi.org/10.1016/j.biopsycho.2010.03.010

[191] Vladyslav Kushnir, Alexandra Godinho, David C. Hodgins, Christian S. Hendershot and John A. Cunningham. 2016. Motivation to Quit or Reduce Gambling: Associations Between Self-Determination Theory and the Transtheoretical Model of Change. *Journal of Addictive Diseases* 35 (1). Taylor and Francis, 58–65. DOI:http://dx.doi.org/10.1080/10550887.2016.1107315

[192] Wilburn Lane and Chris Manner. 2012. The Influence of Personality Traits on Mobile Phone Application Preferences. *Journal of Economics and Behavioral Studies* 4 (5). JEBS, 252–260. DOI:http://dx.doi.org/10.22610/jebs.v4i5.325

[193] Élise Lavoué, Baptiste Monterrat, Michel Desmarais and Sébastien George. 2018. Adaptive Gamification for Learning Environments. *IEEE Transactions on Learning Technologies* 12 (1). IEEE, 16–28. DOI:http://dx.doi.org/10.1109/TLT.2018.2823710

[194] Guy Le-Masurier, Cara Sidman and Charles Corbin. 2003. Accumulating 10,000 Steps: Does this Meet Current Physical Activity Guidelines? *Research Quarterly for Exercise and Sport* 74 (4). Taylor and Francis, 389–394. DOI: http://dx.doi.org/10.1080/02701367.2003.10609109

[195] Bruce W Lemon, Vern L Bengtson and James A Peterson. 1972. An Exploration of the Activity Theory of Aging: Activity Types and Life Satisfaction among In-Movers to a Retirement Community. *Journal of Gerontology* 27 (4). Oxford, 511–523.

[196] Pascal Lessel, Maximilian Altmeyer, Frederic Kerber, Michael Barz, Cornelius. Leidinger and Antonio Krüger. 2016. WaterCoaster: A Device to Encourage People in a Playful Fashion to Reach Their Daily Water Intake Level. *Extended Abstracts Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, 1813–1820. DOI:http://dx.doi.org/10.1145/2851581.2892498

[197] Pascal Lessel, Maximilian Altmeyer and Antonio Krüger. 2018. Users As Game Designers : Analyzing Gamification Concepts in a "Bottom-Up" Setting. *Proceedings of the International Academic Mindtrek Conference (Mindtrek '18)*. ACM, 1–10. DOI:http://dx.doi.org/10.1145/3275116.3275118

[198] Pascal Lessel, Maximilian Altmeyer, Marc Müller, Christian Wolff and Antonio Krüger. 2016. Don't Whip Me With Your Games - Investigating Bottom-Up Gamification. *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, 2026–2037. DOI:http://dx.doi.org/10.1145/2858036.2858463

[199] Pascal Lessel, Maximilian Altmeyer, Marc Müller, Christian Wolff and Antonio Krüger. 2017. Measuring the Effect of "Bottom-Up" Gamification in a Microtask Setting. *Proceedings of the 21st International Academic Mindtrek Conference*. 63–72. DOI:http://dx.doi.org/10.1145/3131085.3131086

[200] Pascal Lessel, Maximilian Altmeyer, Lea Verena Schmeer and Antonio Krüger. 2019. "Enable or Disable Gamification?" – Analyzing the Impact of Choice in a Gamified Image Tagging Task. *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM.

[201] Ian Li, Anind Dey and Jodi Forlizzi. 2010. A Stage-Based Model of Personal Informatics Systems. *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, 557–566. DOI:http://dx.doi.org/10.1145/1753326.1753409

[202] Xiaozhou Li, Chien Lu, Jaakko Peltonen and Zheying Zhang. 2019. A Statistical Analysis of Steam User Profiles Towards Personalized Gamification. *Proceedings of the International GamiFIN Conference (GamiFIN '19)*. CEUR-WS, 217–228.

[203] Feng Lin, Aosen Wang, Chen Song and Wenyao Xu. 2015. A Comparative Study of Smart Insole on Real-world Step Count. *Signal Processing in Medicine and Biology Symposium (SPMB '15)*. IEEE, 1–6. DOI:http://dx.doi.org/10.1109/SPMB.2015.7405425

[204] James J Lin, Lena Mamykina, Silvia Lindtner, Gregory Delajoux and Henry B Strub. 2006. Fish'n'Steps: Encouraging Physical Activity with an Interactive Computer Game. *Proceedings of the International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '06)*. Springer, 261–278. DOI:http://dx.doi.org/10.1007/11853565_16

[205] Yuping Liu and L. J. Shrum. 2002. What is Interactivity and Is It Always Such a Good Thing? Implications of Definition, Person, and Situation for the Influence of Interactivity on Advertising Effectiveness. *Journal of Advertising* 31 (4). Taylor and Francis, 53–64. DOI:http://dx.doi.org/10.1080/00913367.2002.10673685

[206] Edwin A. Locke and Gary P. Latham. 2002. Building a Practically Useful Theory of Goal Setting and Task Motivation: A 35-Year Odyssey. *American Psychologist* 57 (9). American Psychological Association, 705–717. DOI:http://dx.doi.org/10.1037/0003-066X.57.9.705

[207] Paul M. Lukacs, Kenneth P. Burnham and David R. Anderson. 2010. Model Selection Bias and Freedman's Paradox. *Annals of the Institute of Statistical Mathematics* 62 (1). Springer, 117–125. DOI:http://dx.doi.org/10.1007/s10463-009-0234-4

[208] U. R. S. Mäder, B. W. Martin, Y. Schutz and B. Marti. 2006. Validity of Four Short Physical Activity Questionnaires in Middle-Aged Persons. *Medicine and Science in Sports and Exercise* 38 (7). American College of Sports Medicine, 1255–1266. DOI:http://dx.doi.org/10.1249/01.mss.0000227310.18902.28

[209] Martin L Maehr and J.G Nicholls. 1980. Culture and Achievement Motivation: A Second Look. *Studies in Cross-Cultural Psychology* 2. 221–267.

[210] Tobias Mahlmann, Anders Drachen, Julian Togelius, Alessandro Canossa and Georgios N. Yannakakis. 2010. Predicting Player Behavior in Tomb Raider: Underworld. *Proceedings of the Conference on Computational Intelligence and Games (CIG '10)*. IEEE, 178–185. DOI:http://dx.doi.org/10.1109/ITW.2010.5593355

[211] Naomi Mandel and Eric Johnson. 2002. When Web Pages Influence Choice: Effects of Visual Primes on Experts and Novices. *Journal of Consumer Research* 29 (September 2002). The University of Chicago Press, 235–245.

[212] B H Marcus, V C Selby, R S Niaura, J S Rossi, David Lavallee, Jean M Williams and Marc V Jones. 2008. Self-Efficacy and the Stages of Exercise Behavior Change. *Research Quarterly for Exercise and Sport* 63 (1). Taylor and Francis, 60–66. DOI:http://dx.doi.org/10.1080/02701367.1992.10607557

[213] Andrzej Marczewski. 2015. *Even Ninja Monkeys Like to Play: Gamification, Game Thinking and Motivational Design*. CreateSpace Independent Publishing Platform.

[214] Anna Lisa Martin-Niedecken, Katja Rogers, Laia Turmo Vidal, Elisa D Mekler and Elena Márquez Segura. 2019. ExerCube vs. Personal Trainer: Evaluating a Holistic, Immersive, and Adaptive Fitness Game Setup. *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, 1–15. DOI:http://dx.doi.org/10.1145/3290605.3300318

[215] A.H. Maslow. 1943. A Theory of Human Motivation. *Psychological Review* 50 (4). American Psychological Association, 370–396. DOI:http://dx.doi.org/10.1037/h0054346

[216] Edward D. McAuley, Terry Duncan and Vance V. Tammen. 1989. Psychometric Properties of the Intrinsic Motivation Inventoiy in a Competitive Sport Setting: A Confirmatory Factor Analysis. *Research Quarterly for Exercise and Sport* 60 (1). Taylor and Francis, 48–58. DOI:http://dx.doi.org/10.1080/02701367.1989.10607413

[217] Cameron McCarthy, Nikhilesh Pradhan, Calum Redpath and Andy Adler. 2016. Validation of the Empatica E4 Wristband. *Proceedings of the EMBS International Student Conference (ISC '16)*. IEEE, 4–7. DOI:http://dx.doi.org/10.1109/EMBSISC.2016.7508621

[218] Scott McCoy, Andrea Everard, Peter Polak and Dennis F. Galletta. 2007. The Effects of Online Advertising. *Commun. ACM* 50 (3). ACM, 84–88. DOI:http://dx.doi.org/10.1145/1226736.1226740

[219] Robert R. McCrae and Oliver P. John. 1992. An Introduction to the Five-Factor Model and Its Applications. *Journal of Personality* 60 (2). Wiley, 175–215. DOI:http://dx.doi.org/10.1111/j.1467-6494.1992.tb00970.x

[220] William McDougall. 1908. *An Introduction to Social Psychology*. Methuen & Co. Ltd.

[221] Mary L McHugh. 2012. Interrater Reliability: The Kappa Statistic. *Biochemia Medica* 22 (3). Medicinska Naklada, 276–282.

[222] Anne Collins McLaughlin and Maribeth Gandy. 2013. Successful Aging Through Digital Games: Socioemotional Differences Between Older Adult Gamers and Non-Gamers. *Computers In Human Behavior* 29 (4). Elsevier, 1302–1306. DOI:http://dx.doi.org/10.1016/j.chb.2013.01.014

[223] Elisa D. Mekler, Florian Brühlmann, Klaus Opwis and Alexandre N. Tuch. 2013. Do Points, Levels and Leaderboards Harm Intrinsic Motivation? An Empirical Analysis of Common Gamification Elements. *Proceedings of the International Conference on Gameful Design, Research, and Applications (Gamification '13)*. ACM, 66–73. DOI:http://dx.doi.org/10.1145/2583008.2583017

[224] Elisa D. Mekler, Florian Brühlmann, Alexandre N. Tuch and Klaus Opwis. 2017. Towards Understanding the Effects of Individual Gamification Elements on Intrinsic Motivation and Performance. *Computers in Human Behavior* 71. Elsevier, 525–534. DOI:http://dx.doi.org/10.1016/j.chb.2015.08.048

[225] Elisa D. Mekler, Alexandre N. Tuch, Florian Brühlmann and Klaus Opwis. 2013. Disassembling Gamification: The Effects of Points and Meaning on User Motivation and Performance. *Extended Abstracts Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, 1137–1142. DOI:http://dx.doi.org/10.1145/2468356.2468559

[226] Anne-Sophie Melenhorst, Wendy Rogers and Evan Caylor. 2001. The Use of Communication Technologies by Older Adults: Exploring The Benefits From The User's Perspective. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 45 (3). Sage, 221–225.

[227] Satya Menon and Dilip Soman. 2002. Managing the Power of Curiosity for Effective Web Advertising Strategies. *Journal of Advertising* 31 (3). Taylor and Francis, 1–14. DOI:http://dx.doi.org/10.1080/00913367.2002.10673672

[228] Jochen Meyer, Elke Beck, Kai Von Holdt, Dirk Gansefort, Tilman Brand, Hajo Zeeb and Susanne Boll. 2018. Actistairs: Design and Acceptance of a Technology-based Intervention to Advocate Stair-Climbing in Public Spaces. *Proceedings of the International Workshop on Multimedia for Personal Health and Health Care (HealthMedia '18)*. ACM, 59–66. DOI:http://dx.doi.org/10.1145/3264996.3264997

[229] Anouk Middelweerd, Julia Mollee, Michel Klein, Adnan Manzoor, Johannes Brug and Saskia J te Velde. 2020. The Use and Effects of an App-Based Physical Activity Intervention "Active2Gether" in Young Adults: Quasi-Experimental Trial. *JMIR Formative Research* 4 (1). JMIR, 1–19. DOI:http://dx.doi.org/10.2196/12538

[230] Andrew D Miller and Elizabeth D Mynatt. 2014. StepStream: A School-based Pervasive Social Fitness System for Everyday Adolescent Health. *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, 1–10. DOI:http://dx.doi.org/10.1145/2556288.2557190

[231] Ben Miroglio, David Zeber, Jofish Kaye and Rebecca Weiss. 2018. The Effect of Ad Blocking on User Engagement with the Web. *Proceedings of the International World Wide Web Conference (WWW '18)*. ACM, 813–821. DOI: http://dx.doi.org/10.1145/3178876.3186162

[232] Torin Monahan and Jill A Fisher. 2010. Benefits of Observer Effects: Lessons from the Field. *Qualitative Research* 10 (3). Sage, 357–376. DOI: http://dx.doi.org/10.1177/1468794110362874

[233] Md Abu Sayeed Mondol and John A. Stankovic. 2015. Harmony: A Hand Wash Monitoring and Reminder System using Smart Watches. *EAI Endorsed Transactions on Ambient Systems* 2 (6). EAI, 1–10. DOI:http://dx.doi.org/10.4108/eai.22-7-2015.2260042

[234] Elizabeth Monk-Turner, Stephen Warren, Donald Edwards, Maria Rushing, Steve Poorman and Jaritha Willie. 2002. Predictors of Hand-Washing Behavior. *Social Behavior and Personality* 30 (8). Scientific Journal Publishers, 751–756. DOI:http://dx.doi.org/10.2224/sbp.2002.30.8.751

[235] Baptiste Monterrat, Michel Desmarais, Élise Lavoué and Sébastien George. 2015. A Player Model for Adaptive Gamification. *Proceedings of the International Conference on Artificial Intelligence in Education (AIED '15)*. Springer, 297–306. DOI:http://dx.doi.org/10.1007/978-3-319-19773-9

[236] Alberto Mora, Gustavo F. Tondello, Lennart E. Nacke and Joan Arnedo-Moreno. 2018. Effect of Personalized Gameful Design on Student Engagement. *Proceedings of the Global Engineering Education Conference (EDUCON '18)*. IEEE, 1925–1933. DOI:http://dx.doi.org/10.1109/EDUCON.2018.8363471

[237] Elaine Mullan and David Markland. 1997. Variations in Self-Determination Across the Stages of Change for Exercise in Adults. *Motivation and Emotion* 21 (4). Springer, 349–362. DOI:http://dx.doi.org/10.1023/A:1024436423492

[238] Lennart E. Nacke, Chris Bateman and Regan L. Mandryk. 2014. BrainHex: A Neurobiological Gamer Typology Survey. *Entertainment Computing* 5 (1). Elsevier, 55–62. DOI:http://dx.doi.org/10.1016/j.entcom.2013.06.002

[239] Lennart E. Nacke and Sebastian Deterding. 2017. The Maturing of Gamification Research. *Computers in Human Behavior* 71. 450–454. DOI: http://dx.doi.org/10.1016/j.chb.2016.11.062

[240] Lennart E. Nacke, Mark N. Grimshaw and Craig A. Lindley. 2010. More than a Feeling: Measurement of Sonic User Experience and Psychophysiology in a First-Person Shooter Game. *Interacting with Computers* 22 (5). Elsevier, 336–343. DOI:http://dx.doi.org/10.1016/j.intcom.2010.04.005

[241] Tatsuo Nakajima and Vili Lehdonvirta. 2013. Designing Motivation Using Persuasive Ambient Mirrors. *Personal and Ubiquitous Computing* 17 (1). Springer, 107–126. DOI:http://dx.doi.org/10.1007/s00779-011-0469-y

[242] David P Nalbone, Kevin P. Lee, Amanda R. Suroviak and Jennifer M. Lannon. 2005. The Effects of Social Norms on Male Hygiene. *Individual Differences Research* 3 (3). IDR Publishing, 171–176.

[243] Henk Herman Nap, Yvonne A.W De Kort and Wijnand A. IJsselsteijn. 2009. Senior Gamers: Preferences, Motivations and Needs. *Gerontechnology* 8 (4). International Society for Gerontechnology, 247–262. DOI:http://dx.doi.org/10.4017/gt.2009.08.04.003.00

[244] Elnaz Nasirzadeh and Mohammad Fathian. 2020. Investigating the Effect of Gamification Elements on Bank Customers to Personalize Gamified Systems. *International Journal of Human Computer Studies* 143. Elsevier, 102469. DOI:http://dx.doi.org/10.1016/j.ijhcs.2020.102469

[245] Michelle R. Nelson. 2002. Recall of Brand Placements in Computer/Video Games. *Journal of Advertising Research* 42 (2). JAR, 80–92. DOI:http://dx.doi.org/10.2501/JAR-42-2-80-92

[246] Andreas B Neubauer and Andreas Voss. 2016. Validation and Revision of a German Version of the Balanced Measure of Psychological Needs Scale. *Journal of Individual Differences* 37 (1). Hogrefe Publishing, 56–72. DOI:http://dx.doi.org/10.1027/1614-0001/a000188

[247] Aljoscha C. Neubauer and Bertram F. Malle. 1997. Questionnaire Response Latencies: Implications for Personality Assessment and Self-Schema Theory. *European Journal of Psychological Assessment* 13 (2). Hogrefe & Huber Publishers, 109–117. DOI:http://dx.doi.org/10.1027/1015-5759.13.2.109

[248] Eric J. Newman, Donald E. Stem and David E. Sprott. 2004. Banner Advertisement and Web Site Congruity Effects on Consumer Web Site Perceptions. *Industrial Management and Data Systems* 104 (3). Emerald, 273–281. DOI:http://dx.doi.org/10.1108/02635570410525816

[249] Newzoo. 2020. *Global Games Market Report*. Newzoo. https://resources.newzoo.com/hubfs/2019

[250] John G. Nicholls. 1984. Achievement Motivation: Conceptions of Ability, Subjective Experience, Task Choice, and Performance. *Psychological Review* 91 (3). American Psychological Association, 328–346. DOI:http://dx.doi.org/10.1037/0033-295X.91.3.328

[251] Scott Nicholson. 2012. A User-Centered Theoretical Framework for Meaningful Gamification. *Proceedings of the Games+Learning+Society Conference (GLS '12)*. ETC Press, 223–229.

[252] Jakob Nielsen. 1992. Evaluating the Thinking-Aloud Technique for Use by Computer Scientists. In *Advances in Human-Computer Interaction*. Vol. 3. Ablex Publishing, 69–82.

[253] Jakob Nielsen. 2000. Why You Only Need to Test with 5 Users. https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/

[254] Jasmin Niess and Paweł W. Woźniak. 2018. Supporting Meaningful Personal Fitness: The Tracker Goal Evolution Model. *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, 1–12. DOI:http://dx.doi.org/10.1145/3173574.3173745

[255] Kim Nimon, Robin K. Henson and Michael S. Gates. 2010. Revisiting Interpretation of Canonical Correlation Analysis: A Tutorial and Demonstration of Canonical Commonality Analysis. *Multivariate Behavioral Research* 45 (4). Taylor and Francis, 702–724. DOI:http://dx.doi.org/10.1080/00273171.2010.498293

[256] Claire E. Norris and Andrew M. Colman. 1992. Context Effects on Recall and Recognition of Magazine Advertisements. *Journal of Advertising* 21 (3). Taylor and Francis, 37–46. DOI:http://dx.doi.org/10.1080/00913367.1992.10673374

[257] Rita Orji. 2014. Exploring the Persuasiveness of Behavior Change Support Strategies and Possible Gender Differences. *International Workshop on Behavior Change Support Systems (BCSS '14)* BCSS. 41–57.

[258] Rita Orji, Regan L Mandryk and Julita Vassileva. 2015. Gender, Age, and Responsiveness to Cialdini's Persuasion Strategies. *Proceedings of the International Conference on Persuasive Technology (PERSUASIVE '15)*. Springer, 147–159. DOI:http://dx.doi.org/10.1007/978-3-319-20306-5_14

[259] Rita Orji, Lennart E. Nacke and Chrysanne Di Marco. 2017. Towards Personality-driven Persuasive Health Games and Gamified Systems. *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, 1015–1027. DOI:http://dx.doi.org/10.1145/3025453.3025577

[260] Rita Orji, Kiemute Oyibo and Gustavo F. Tondello. 2017. A Comparison of System-Controlled and User-Controlled Personalization Approaches. *Adjunct Publication of the Conference on User Modeling, Adaptation and Personalization (UMAP '17)*. ACM, 413–418. DOI:http://dx.doi.org/10.1145/3099023.3099116

[261] Rita Orji, Gustavo F. Tondello and Lennart E. Nacke. 2018. Personalizing Persuasive Strategies in Gameful Systems to Gamification User Types. *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, 1–14. DOI:http://dx.doi.org/doi.org/10.1145/3173574.3174009

[262] Rita Orji, Julita Vassileva and Regan L Mandryk. 2014. Modeling the Efficacy of Persuasive Strategies for Different Gamer Types in Serious Games for Health. *User Modeling and User-Adapted Interaction* 24. Springer, 453–498. DOI:http://dx.doi.org/10.1007/s11257-014-9149-8

[263] Thérèse J.M. Overbeek, Anton van Boxtel and Joyce H.D.M. Westerink. 2012. Respiratory Sinus Arrhythmia Responses to Induced Emotional States: Effects of RSA Indices, Emotion Induction Method, Age, and Sex. *Biological Psychology* 91 (1). Elsevier, 128–141. DOI:http://dx.doi.org/10.1016/j.biopsycho.2012.05.011

[264] Kiemute Oyibo, Rita Orji and Julita Vassileva. 2017. Investigation of the Persuasiveness of Social Influence in Persuasive Technology and the Effect of Age and Gender. *Proceedings of the Personalization in Persuasive Technology Workshop (PPT'17)*. 32–44.

[265] Allan Paivio and Kalman Csapo. 1973. Picture Superiority in Free Recall: Imagery or Dual Coding? *Cognitive Psychology* 5 (2). Academic Press, 176–206. DOI:http://dx.doi.org/10.1016/0010-0285(73)90032-7

[266] RR Pate, M Pratt, SN Blair, WL Haskell, CA Macera, C Bouchard, D Buchner, W Ettinger, GW Heath, AC King, A Kriska, A Leon, B Marcus, J Morris, R Paffenberger, K Patrick, M Pollock, J Rippe, J Sallis and J Wilmore. 1995. Physical Activity and Public Health: A Recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine. *Jama* 273 (5). American Medical Association. DOI:http://dx.doi.org/10.1001/jama.273.5.402

[267] B. R. Payne, J. J. Jackson, S. R. Noh and E. A. L. Stine-Morrow. 2011. Activity Flow State Scale. DOI:http://dx.doi.org/10.1037/t06855-000

[268] B. R. Payne, J. J. Jackson, S. R. Noh and E. A. L. Stine-Morrow. 2011. In The Zone: Flow State and Cognition in Older Adults Brennan. *Psychology and Aging* 26 (3). American Psychological Association, 738–743. DOI:http://dx.doi.org/10.1037/a0022359

[269] James G. Phillips, Sarah Butt and Alex Blaszczynski. 2006. Personality and Self-Reported Use of Mobile Phones for Games. *CyberPsychology and Behavior* 9 (6). Mary Ann Liebert, 753–758. DOI:http://dx.doi.org/10.1089/cpb.2006.9.753

[270] Peter A. Piccione. 1980. In Search of the Meaning of Senet. *Archaeology* 33. Archaeological Institute of America, 55–58.

[271] Ganna Pogrebna and Alexander A Kharlamov. 2020. The Impact of Cross-Cultural Differences in Handwashing Patterns on the COVID-19 Outbreak Magnitude. *Research Gate (Preprint)* March. 10. DOI:http://dx.doi.org/10.13140/RG.2.2.23764.96649

[272] J Posner, James A. Russell and Bradley S. Peterson. 2005. The Circumplex Model of Affect: An Integrative Approach to Affective Neuroscience, Cognitive Development, and Psychopathology. *Development and Psychopathology* 17 (3). Cambridge University Press, 715–734. DOI:http://dx.doi.org/10.1017/S0954579405050340

[273] Paul Christopher Price, Rajiv Jhangiani and I-Chant A. Chiang. 2015. *Research Methods in Psychology* (2nd ed.). Bcampus. https://openlibrary-repo.ecampusontario.ca/jspui/handle/123456789/246

[274] Paul C Price, Rajiv S Jhangiani, I-Chant A. Chiang, Dana C. Leighton and Carrie Cuttler. 2017. *Research Methods in Psychology* (3rd ed.). Bcampus. https://opentext.wsu.edu/carriecuttler/

[275] James O. Prochaska and Carlo C. DiClemente. 1983. Stages and Processes of Self-Change of Smoking: Toward an Integrative Model of Change. *Journal of Consulting and Clinical Psychology* 51 (3). American Psychological Association, 390–395. DOI:http://dx.doi.org/10.1037/0022-006X.51.3.390

[276] James O. Prochaska and Wayne F. Velicer. 1997. The Transtheoretical Change Model of Health Behavior. *American Journal of Health Promotion* 12 (1). Sage, 38–48. DOI:http://dx.doi.org/10.4278/0890-1171-12.1.38

[277] Enric Pujol, Oliver Hohlfeld and Anja Feldmann. 2015. Annoyed Users: Ads and Ad-Block Usage in the Wild. *Proceedings of the Internet Measurement Conference (IMC '15)*. ACM, 93–106. DOI:http://dx.doi.org/10.1145/2815675.2815705

[278] Rachael Purta, Stephen Mattingly, Lixing Song, Omar Lizardo, David Hachen, Christian Poellabauer and Aaron Striegel. 2016. Experiences Measuring Sleep and Physical Activity Patterns Across a Large College Cohort with Fitbits. *Proceedings of the International Symposium on Wearable Computers (ISWC '16)*. ACM, 28–35. DOI:http://dx.doi.org/10.1145/2971763.2971767

[279] Shantha M W Rajaratnam and Josephine Arendt. 2001. Health in a 24-h Society. *The Lancet* 358. Elsevier, 999–1005. DOI:http://dx.doi.org/10.1016/S0140-6736(01)06108-6

[280] Beatrice Rammstedt and Oliver P. John. 2007. Measuring Personality in One Minute or Less: A 10-Item Short Version of the Big Five Inventory in English and German. *Journal of Research in Personality* 41 (1). Elsevier, 203–212. DOI:http://dx.doi.org/10.1016/j.jrp.2006.02.001

[281] Amon Rapp, Frank Hopfgartner, Juho Hamari, Conor Linehan and Federica Cena. 2019. Strengthening Gamification Studies: Current Trends and Future Opportunities of Gamification Research. *International Journal of Human Computer Studies* 127. Elsevier, 1–6. DOI:http://dx.doi.org/10.1016/j.ijhcs.2018.11.007

[282] Johnmarshall Reeve. 2009. *Understanding Motivation and Emotion* (5th ed.). Wiley.

[283] Gary Reker, Edward Peacock and Paul Wong. 1987. Meaning and Purpose in Life and Well-Being: A Life-Span Perspective. *Journal of Gerontology* 42 (1). Oxford, 44–49.

[284] Melanie Revilla and Carlos Ochoa. 2017. Ideal and Maximum Length for a Web Survey. *International Journal of Market Research* 59 (5). Sage, 557–565. DOI:http://dx.doi.org/10.2501/IJMR-2017-039

[285] Chad Richards, Craig W Thompson and T C Nicholas Graham. 2014. Beyond Designing for Motivation: The Importance of Context in Gamification. *Proceedings of the Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '14)*. ACM, 217–226. DOI:http://dx.doi.org/10.1145/2658537.2658683

[286] Scott Rigby and Richard M. Ryan. 2011. *Glued to Games - How Video Games Draw Us In and Hold Us Spellbound*. ABC-CLIO.

[287] Sara E. Rimm-Kaufman and Jerome Kagan. 1996. The Psychological Significance of Changes in Skin Temperature. *Motivation and Emotion* 20 (1). Springer, 63–78. DOI:http://dx.doi.org/10.1007/BF02251007

[288] Kirsten Risden, Mary Czerwinski, Stephanie Worley, Lynda Hamilton, Joe Kubiniec, Hunter Hoffman, Nancy Mickel and Elizabeth Loftus. 1998. Interactive Advertising: Patterns of Use and Effectiveness. *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '98)*. ACM, 219–224. DOI:http://dx.doi.org/10.1145/274644.274676

[289] Ute Ritterfeld, Michael Cody and Peter Vorderer. 2009. *Serious Games: Mechanisms and Effects*. Taylor and Francis. DOI:http://dx.doi.org/10.4324/9780203891650

[290] Richard W Robins, Chris R. Fraley and Robert F. Krueger. 2007. *Handbook of Research Methods in Personality Psychology*. The Guilford Press.

[291] Richard W Robins, Kali H Trzesniewski, Jessica L Tracy, Samuel D Gosling and Jeff Potter. 2002. Global Self-Esteem Across the Life Span. *Psychology and Aging* 17 (3). American Psychological Association, 423–434. DOI:http://dx.doi.org/10.1037//0882-7974.17.3.423

[292] Christian Rohrer and John Boyd. 2004. The Rise of Intrusive Online Advertising and the Response of User Experience Research at Yahoo!. In *Extended Abstracts Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '04)*. ACM, 1085–1086. DOI:http://dx.doi.org/10.1145/985921.985992

[293] Natalia Romero, Janienke Sturm, Tilde Bekker, Linda De Valk and Sander Kruitwagen. 2010. Playful Persuasion to Support Older Adults' Social and Physical Activities. *Interacting with Computers* 22 (6). Elsevier, 485–495. DOI:http://dx.doi.org/10.1016/j.intcom.2010.08.006

[294] John Rooksby, Mattias Rost, Alistair Morrison and Matthew Chalmers. 2014. Personal Tracking as Lived Informatics. *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, 9016–9016.

[295] John F. Runcie. 1980. By Days I Make the Cars. *Harvard Business Review* 58 (3). Harvard Business Publishing, 106–115.

[296] James A Russell and James M Carroll. 1999. On the Bipolarity of Positive and Negative Affect. *Psychological Bulletin* 125 (1). American Psychological Association, 3–30. DOI:http://dx.doi.org/10.1037/0033-2909.125.1.3

[297] Richard M. Ryan. 1982. Control and Information in the Intrapersonal Sphere: An Extension of Cognitive Evaluation Theory. *Journal of Personality and Social Psychology* 43 (3). American Psychological Association, 450–461. DOI:http://dx.doi.org/10.1037/0022-3514.43.3.450

[298] Richard M Ryan and Edward L Deci. 2000. Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being. *American Psychologist* 55 (1). American Medical Association, 68–78. DOI:http://dx.doi.org/10.1037/0003-066X.55.1.68

[299] Richard M. Ryan and Edward L Deci. 2017. *Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness*. The Guilford Press.

[300] Richard M. Ryan, C. Scott Rigby and Andrew Przybylski. 2006. The Motivational Pull of Video Games: A Self-Determination Theory Approach. *Motivation and Emotion* 30 (4). Springer, 347–363. DOI:http://dx.doi.org/10.1007/s11031-006-9051-8

[301] Carol D Ryff. 1989. Happiness is Everything, or Is It? Explorations on the Meaning of Psychological Well-Being. *Journal of Personality and Social Psychology* 57 (6). American Psychological Association, 1069–1081. DOI:http://dx.doi.org/10.1037/0022-3514.57.6.1069

[302] Carol D Ryff and Corey Lee M Keyes. 1995. The Structure of Psychological Well-Being Revisited. *Journal of Personality and Social Psychology* 69 (4). American Psychological Association, 719–727. DOI:http://dx.doi.org/10.1037/0022-3514.69.4.719

[303] Michael Sailer, Jan Ulrich Hense, Sarah Katharina Mayr and Heinz Mandl. 2017. How Gamification Motivates: An Experimental Study of the Effects of Specific Game Design Elements on Psychological Need Satisfaction. *Computers in Human Behavior* 69. Elsevier, 371–380. DOI:http://dx.doi.org/10.1016/j.chb.2016.12.033

[304] Herman Saksono, Ashwini Ranade, Geeta Kamarthi, Carmen Castaneda-Sceppa, Jessica A. Hoffman, Cathy Wirth and Andrea G. Parker. 2015. Spaceship Launch: Designing a Collaborative Exergame for Families. *Proceedings of the International Conference on Computer-Supported Cooperative Work and Social Computing (CSCW '15)*. ACM, 1776–1787. DOI:http://dx.doi.org/10.1145/2675133.2675159

[305] Carol Sansone, Charlene Weir, Lora Harpster and Carolyn Morgan. 1992. Once a Boring Task Always a Boring Task?: Interest as a Self-Regulatory Mechanism. *Journal of Personality and Social Psychology* 63 (3). American Psychological Association, 379–390. DOI:http://dx.doi.org/10.1037/0022-3514.63.3.379

[306] Chris Schaefbauer, Danish Kahn, Amy Le, Garrett Sczechowski and Katie Siek. 2015. Snack Buddy: Supporting Healthy Snacking in Low Socioeconomic Status Families. *Proceedings of the Conference on Computer Supported Cooperative Work (CSCW '15)*. ACM, 1045–1057.

[307] Jesse Schell. 2014. *The Art of Game Design - A Book of Lenses* (2 ed.). Taylor and Francis. DOI:http://dx.doi.org/10.1201/b17723

[308] Albrecht Schmidt. 2000. Implicit Human Computer Interaction through Context. *Personal Technologies* 4. Springer, 191–199. DOI:http://dx.doi.org/10.1007/BF01324126

[309] Albrecht Schmidt, Michael Beigl and Hans W. Gellersen. 1999. There is More to Context than Location. *Computers and Graphics* 23 (6). Elsevier, 893–901. DOI:http://dx.doi.org/10.1016/S0097-8493(99)00120-X

[310] Martin Schrepp, Andreas Hinderks and Jörg Thomaschewski. 2017. Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S). *International Journal of Interactive Multimedia and Artificial Intelligence* 4 (6). IJIMAI, 103–108. DOI:http://dx.doi.org/10.9781/ijimai.2017.09.001

[311] Marc Schubhan, Maximilian Altmeyer, Dominic Buchheit and Pascal Lessel. 2020. Investigating User-Created Gamification in an Image Tagging Task. *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '20)*. ACM, 1–12. DOI:http://dx.doi.org/10.1145/3313831.3376360

[312] Diane P. Schultz and Sydney Ellen Schultz. 2011. *A History of Modern Psychology* (10 ed.). Wadsworth.

[313] Bob De Schutter. 2010. Never Too Old to Play: The Appeal of Digital Games to an Older Audience. *Games and Culture* 1. Sage, 1–16. DOI:http://dx.doi.org/10.1177/1555412010364978

[314] Christiane Schwieren and Doris Weichselbaumer. 2010. Does Competition Enhance Performance or Cheating? A Laboratory Experiment. *Journal of Economic Psychology* 31 (3). Elsevier, 241–253. DOI:http://dx.doi.org/10.1016/j.joep.2009.02.005

[315] Katie Seaborn and Deborah Fels. 2015. Gamification in Theory and Action: A Survey. *International Journal of Human-Computer Studies* 74. Elsevier, 14–31. DOI:http://dx.doi.org/10.1016/j.ijhcs.2014.09.006

[316] Suranga Seneviratne, Aruna Seneviratne, Prasant Mohapatra and Anirban Mahanti. 2014. Predicting User Traits from a Snapshot of Apps Installed on a Smartphone. *Mobile Computing and Communications Review* 18 (2). ACM, 1–8. DOI:http://dx.doi.org/10.1145/2636242.2636244

[317] Kennon M Sheldon and Jonathan C Hilpert. 2012. The Balanced Measure of Psychological Needs (BMPN) Scale: An Alternative Domain General Measure of Need Satisfaction. *Motivation and Emotion* 36. Springer, 439–451. DOI:http://dx.doi.org/10.1007/s11031-012-9279-4

[318] Alissa Sherry and Robin K. Henson. 2005. Conducting and Interpreting Canonical Correlation Analysis in Personality Research: A User-Friendly Primer. *Journal of Personality Assessment* 84 (1). Taylor and Francis, 37–48. DOI:http://dx.doi.org/10.1207/s15327752jpa8401_09

[319] Ho Shu Hsun, Lin Yu Ling and Yang Yu Ting. 2011. In-Game Advertising: Consumers' Attitude and the Effect of Product Placements on Memory. *African Journal of Business Management* 5 (24). Academic Journals, 10117–10127. DOI:http://dx.doi.org/10.5897/AJBM11.377

[320] B.F. Skinner. 1938. *The Behavior of Organisms*. BF Skinner Foundation.

[321] Jan Smeddinck, M Herrlich and M Krause. 2012. Did They Really Like the Game?–Challenges in Evaluating Exergames with Older Adults. *CHI 2012 Workshop on Game User Research: Exploring Methodologies*. 1–4.

[322] Debra Smith, Barry S. Michaels, Charles A. Bartleson, Judy D. Greid and Ewen C. D. Todd. 2016. Outbreaks Where Food Workers Have Been Implicated in the Spread of Foodborne Disease. Part 9. Washing and Drying of Hands To Reduce Microbial Contamination. *Journal of Food Protection* 73 (10). International Association for Food Protection, 1937–1955. DOI: http://dx.doi.org/10.4315/0362-028x-73.10.1937

[323] Hanne Spelt, Elisabeth Kersten-van Dijk, Jaap Ham, Joyce Westerink and Wijnand IJsselsteijn. 2019. Psychophysiological Measures of Reactance to Persuasive Messages Advocating Limited Meat Consumption. *Information* 10 (320). MDPI, 1–12. DOI:http://dx.doi.org/10.3390/info10100320

[324] Hanne Spelt, Joyce Westerink, Jaap Ham and Wijnand Ijsselsteijn. 2018. Cardiovascular Reactions During Exposure to Persuasion Principles. *Proceedings of the International Conference on Persuasive Technology (PERSUASIVE '18)*. Springer, 267–278. DOI:http://dx.doi.org/10.1007/978-3-319-78978-1

[325] Hanne A. A. Spelt, Jaap Westerink, Joyce H. D. M. Ham and Wijnand A. IJsselsteijn. 2019. Psychophysiological Reactions to Persuasive Messages Deploying Persuasion Principles. *IEEE Transactions on Affective Computing*. IEEE. DOI:http://dx.doi.org/10.1109/TAFFC.2019.2931689

[326] Clemens Stachl, Quay Au, Ramona Schoedel, Samuel D. Gosling, Gabriella M. Harari, Daniel Buschek, Sarah Theres Völkel, Tobias Schuwerk, Michelle Oldemeier, Theresa Ullmann, Heinrich Hussmann, Bernd Bischl and Markus Bühner. 2020. Predicting Personality from Patterns of Behavior Collected with Smartphones. *Proceedings of the National Academy of Sciences of the United States of America* 117 (30). National Academy of Sciences, 17680–17687. DOI:http://dx.doi.org/10.1073/pnas.1920484117

[327] James R. Stellar and Eliot Stellar. 1985. *The Neurobiology of Motivation and Reward* (1 ed.). Springer. DOI:http://dx.doi.org/10.1007/978-1-4615-8032-4

[328] S. Shyam Sundar. 1999. Exploring Receivers' Criteria for Perception of Print and Online News. *Journalism and Mass Communication Quaterly* 76 (2). Sage, 373–386. DOI:http://dx.doi.org/10.1177/107769909907600213

[329] S. Shyam Sundar and Jinhee Kim. 2005. Interactivity and Persuasion. *Journal of Interactive Advertising* 5 (2). Taylor and Francis, 5–18. DOI:http://dx.doi.org/10.1080/15252019.2005.10722097

[330] Barabara G. Tabachnick and Linda S. Fidell. 2013. *Using Multivariate Statistics* (6th ed.). Pearson.

[331] Ralf Terlutter and Michael L. Capella. 2013. The Gamification of Advertising: Analysis and Research Directions of In-Game Advertising, Advergames, and Advertising in Social Network Games. *Journal of Advertising* 42 (2-3). Taylor and Francis, 95–112. DOI:http://dx.doi.org/10.1080/00913367.2013.774610

[332] Bruce Thompson. 1984. *Canonical Correlation Analysis: Uses and Interpretation*. Sage.

[333] Simon Tobin and Simon Grondin. 2009. Video Games and the Perception of Very Long Durations by Adolescents. *Computers in Human Behavior* 25 (2). Elsevier, 554–559. DOI:http://dx.doi.org/10.1016/j.chb.2008.12.002

[334] Dereck Toker, Ben Steichen and Max Birk. 2014. Predicting Player Type in Social Network Games. *Proceedings of UMAP Posters, Demonstrations and Late-Breaking Results*. CEUR-WS, 37–40.

[335] Gustavo F. Tondello. 2019. *Dynamic Personalization of Gameful Interactive Systems*. Ph.D. Dissertation. University of Waterloo. http://hdl.handle.net/10012/14807

[336] Gustavo F. Tondello, Karina Arrambide, Giovanni Ribeiro, Andrew Jian lan Cen and Lennart E. Nacke. 2019. "I don't fit into a Single Type": A Trait Model and Scale of Game Playing Preferences. *Proceedings of the International Conference on Human-Computer Interaction (INTERACT '19)*. Springer, 375–395. DOI:http://dx.doi.org/10.1007/978-3-030-29384-0_23

[337] Gustavo F. Tondello, Alberto Mora, Andrzej Marczewski and Lennart E. Nacke. 2019. Empirical Validation of the Gamification User Types Hexad Scale in English and Spanish. *International Journal of Human-Computer Studies* 127. Elsevier, 95–111. DOI:http://dx.doi.org/10.1016/j.ijhcs.2018.10.002

[338] Gustavo F. Tondello, Alberto Mora and Lennart E. Nacke. 2017. Elements of Gameful Design Emerging from User Preferences. *Proceedings of the Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '17)*. ACM, 129–142. DOI:http://dx.doi.org/10.1145/3116595.3116627

[339] Gustavo F. Tondello and Lennart E. Nacke. 2020. Validation of User Preferences and Effects of Personalized Gamification on Task Performance. *Frontiers in Computer Science* 2 (August). Frontiers, 1–23. DOI:http://dx.doi.org/10.3389/fcomp.2020.00029

[340] Gustavo F. Tondello, Hardy Premsukh and Lennart Nacke. 2018. A Theory of Gamification Principles Through Goal-Setting Theory. *Proceedings of the Hawaii International Conference on System Sciences (HICSS '18)*. Association for Information Systems. DOI:http://dx.doi.org/10.24251/HICSS.2018.140

[341] Gustavo F. Tondello, Deltcho Valtchanov, Adrian Reetz, Rina R. Wehbe, Rita Orji and Lennart E. Nacke. 2018. Towards a Trait Model of Video Game Preferences. *International Journal of Human-Computer Interaction* 34 (8). Taylor and Francis, 732–748. DOI:http://dx.doi.org/10.1080/10447318.2018.1461765

[342] Gustavo F. Tondello, Rina R Wehbe, Lisa Diamond, Marc Busch, Andrzej Marczewski and Lennart E Nacke. 2016. The Gamification User Types Hexad Scale. *Proceedings of the Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '16)*. ACM, 229–243. DOI:http://dx.doi.org/10.1145/2967934.2968082

[343] Tammy Toscos, Anne Faber, Shunying An and Mona P. Gandhi. 2006. Chick Clique: Persuasive Technology to Motivate Teenage Girls to Exercise. *Extended Abstracts Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '06)*. ACM, 1873–1878. DOI:http://dx.doi.org/10.1145/1125451.1125805

[344] Tamilla Triantoro, Ram Gopal, Raquel Benbunan-Fich and Guido Lang. 2019. Would You Like to Play? A Comparison of a Gamified Survey with a Traditional Online Survey Method. *International Journal of Information Management* 49. Elsevier, 242–252. DOI:http://dx.doi.org/10.1016/j.ijinfomgt.2019.06.001

[345] Khai N. Truong, Gillian R. Hayes and Gregory Abowd. 2006. Storyboarding: An Empirical Determination of Best Practices and Effective Guidelines. In *Proceedings of the ACM Conference on Designing Interactive Systems (DIS '06)*. ACM, 12–21. DOI:http://dx.doi.org/10.1145/1142405.1142410

[346] Catrine Tudor-Locke, Yoshiro Hatano, Robert P. Pangrazi and Minsoo Kang. 2008. Revisiting "How Many Steps Are Enough?". *Medicine and Science in Sports and Exercise* 40 (7). American College of Sports Medicine, 537–543. DOI:http://dx.doi.org/10.1249/MSS.0b013e31817c7133

[347] Catrine E Tudor-Locke. 2002. Taking Steps Towards Increased Physical Activity: Using Pedometers to Measure and Motivate. *President's Council on Physical Fitness & Sports Research Digest* 3 (17). ERIC, 3–10.

[348] Selen Turkay and Charles K. Kinzer. 2014. The Effects of Avatar-Based Customization on Player Identification. *International Journal of Gaming and Computer-Mediated Simulations* 6 (1). IGI Global, 1–25. DOI:http://dx.doi.org/10.4018/ijgcms.2014010101

[349] Rebecca M. Turner, Sheila M. Bird and Julian P T Higgins. 2013. The Impact of Study Size on Meta-Analyses: Examination of Underpowered Studies in Cochrane Reviews. *PLoS ONE* 8 (3). PLOS, 1–8. DOI:http://dx.doi.org/10.1371/journal.pone.0059202

[350] April Tyack and Elisa D. Mekler. 2020. Self-Determination Theory in HCI Games Research: Current Uses and Open Questions. *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '20)*. ACM, 1–22. DOI:http://dx.doi.org/10.1145/3313831.3376723

[351] Carien M. Van Reekum, Tom Johnstone, Rainer Banse, Alexander Etter, Thomas Wehrle and Klaus R. Scherer. 2004. Psychophysiological Responses to Appraisal Dimensions in a Computer Game. *Cognition and Emotion* 18 (5). Taylor and Francis, 663–688. DOI:http://dx.doi.org/10.1080/02699930341000167

[352] Roumen Vesselinov and John Grego. 2012. *Duolingo Effectiveness Study*. Technical Report December 2012. 1–25 pages. http://static.duolingo.com/s3/DuolingoReport

[353] Aku Visuri, Simo Hosio and Denzil Ferreira. 2017. Exploring Mobile Ad Formats to Increase Brand Recollection and Enhance User Experience. *Proceedings of the International Conference on Mobile and Ubiquitous Multimedia (MUM '17)*. ACM, 311–319. DOI:http://dx.doi.org/10.1145/3152832.3152834

[354] Luis Von Ahn and Laura Dabbish. 2008. Designing Games with a Purpose. *Commun. ACM* 51 (8). ACM, 58–67. DOI:http://dx.doi.org/10.1145/1378704.1378719

[355] David Walker. 2003. Converting Kendall's Tau For Correlational Or Meta-Analytic Analyses. *Journal of Modern Applied Statistical Methods* 2 (2). JMASM, 525–530. DOI:http://dx.doi.org/10.22237/jmasm/1067646360

[356] Hartmut Wandke, Michael Sengpiel and Malte Sönksen. 2012. Myths About Older People's Use of Information and Communication. *Gerontology* 58 (6). Karger, 564–570. DOI:http://dx.doi.org/10.1159/000339104

[357] Hao-Ting Wang, Jonathan Smallwood, Janaina Mourao-Miranda, Cedric Huchuan Xia, Theodore D. Satterthwaite, Danielle S. Bassett and Danilo Bzdok. 2018. Finding the Needle in High-Dimensional Haystack: A Tutorial on Canonical Correlation Analysis. *arXiv preprint arXiv:1812.02598*. arXiv. http://arxiv.org/abs/1812.02598

[358] David Watson, Lee Anna Clark and Auke Tellegen. 1988. Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales. *Journal of Personality and Social Psychology* 54 (6). American Psychological Association, 1063–1070. DOI:http://dx.doi.org/10.1037/0022-3514.54.6.1063

[359] Kevin Werbach. 2014. (Re) Defining Gamification: A Process Approach. *International Conference on Persuasive Technology (PERSUASIVE '14)*. Springer, 266–272. DOI:http://dx.doi.org/10.1007/978-3-319-07127-5_23

[360] Kevin Werbach and Dan Hunter. 2012. *For the Win: How Game Thinking can Revolutionize your Business*. Wharton Digital Press.

[361] M. Wilde, K. Bätz, A. Kovaleva and D. Urhahne. 2009. Testing a Short Scale of Intrinsic Motivation. *Zeitschrift für Didaktik der Naturwissenschaften* 15. Springer, 31–45.

[362] Sally Wilkins. 2002. *Sports and Games of Medieval Cultures*. Greenwood Press.

[363] Wirthlin Worldwide Research and Bayer Pharmaceutical Division. 1996. Handwashing Survey. Fact Sheet. Distributed at American Society for Microbiology Annual Meeting, New Orleans.

[364] Robert Sessions Woodworth. 1918. *Dynamic Psychology*. Columbia University Press.

[365] Nannan Xi and Juho Hamari. 2019. Does Gamification Satisfy Needs? A Study on the Relationship between Gamification Features and Intrinsic Need Satisfaction. *International Journal of Information Management* 46. Elsevier, 210–221. DOI:http://dx.doi.org/10.1016/j.ijinfomgt.2018.12.002

[366] Jing Jian Xiao, Barbara O'Neill, Janice M. Prochaska, Claudia M. Kerbel, Patricia Brennan and Barbara J. Bristow. 2004. A Consumer Education Programme Based on the Transtheoretical Model of Change. *International Journal of Consumer Studies* 28 (1). Wiley, 55–65. DOI:http://dx.doi.org/10.1111/j.1470-6431.2004.00334.x

[367] Carmen Ximénez. 2007. Effect of Variable and Subject Sampling on Recovery of Weak Factors in CFA. *Methodology* 3 (2). Hogrefe & Huber Publishers, 67–80. DOI:http://dx.doi.org/10.1027/1614-2241.3.2.67

[368] Hyeseung Yang and Mary Beth Oliver. 2004. Exploring the Effects of Online Advertising on Readers' Perceptions of Online News. *Journalism and Mass Communication Quarterly* 81 (4). Sage, 733–749. DOI:http://dx.doi.org/10.1177/107769900408100402

[369] Moonhee Yang, David R. Roskos-Ewoldsen, Lucian Dinu and Laura M. Arpan. 2006. The Effectiveness of In-Game Advertising: Comparing College Students' Explicit and Implicit Memory for Brand Names. *Journal of Advertising* 35 (4). Taylor and Francis, 143–152. DOI:http://dx.doi.org/10.2753/JOA0091-3367350410

[370] Zamzami Zainuddin, Samuel Kai Wah Chu, Muhammad Shujahat and Corinne Jacqueline Perera. 2020. The Impact of Gamification on Learning and Instruction: A Systematic Review of Empirical Evidence. *Educational Research Review* 30. Elsevier, 1–23. DOI:http://dx.doi.org/10.1016/j.edurev.2020.100326

[371] Oren Zuckerman and Ayelet Gal-Oz. 2014. Deconstructing Gamification: Evaluating the Effectiveness of Continuous Measurement, Virtual Rewards, and Social Comparison for Promoting Physical Activity. *Personal and Ubiquitous Computing* 18 (7). Springer, 1705–1719. DOI:http://dx.doi.org/10.1007/s00779-014-0783-2