Investigating Avatar Customization as a Motivational Design Strategy for Improving Engagement with Technology-Enabled Services for Health

A Thesis Submitted to the
College of Graduate and Postdoctoral Studies
in Partial Fulfillment of the Requirements
for the degree of Doctor of Philosophy
in the Department of Computer Science
University of Saskatchewan
Saskatoon

By
Max Valentin Birk

Permission to Use

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Computer Science
176 Thorvaldson Building
110 Science Place
University of Saskatchewan
Saskatoon, Saskatchewan S7N 5C9
Canada

Or

Dean
College of Graduate and Postdoctoral Studies
University of Saskatchewan
116 Thorvaldson Building, 110 Science Place
Saskatoon, Saskatchewan S7N 5C9
Canada

Abstract

Technology-enabled services for physical and mental health are a promising approach to improve healthcare globally. Unfortunately, the largest barrier for effective technology-based treatment is participants' gradually fading engagement with effective novel training applications, such as exercise apps or online mental health training programs. Engaging users through design presents an elegant solution to the problem; however, research on technology-enabled services is primarily focused on the efficacy of novel interventions and not on improving adherence through engaging interaction design. As a result, motivational design strategies to improve engagement—both in the moment of use and over time—are underutilized. Drawing from game-design, I investigate avatar customization as a game-based motivational design strategy in four studies. In Study 1, I examine the effect of avatar customization on experience and behaviour in an infinite runner game. In Study 2, I induce different levels of motivation to research the effects of financial rewards on self-reported motivation and performance in a gamified training task over 11 days. In Study 3, I apply avatar customization to investigate the effects of attrition in an intervention context using a breathing exercise over three weeks. In Study 4, I investigate the immediate effects of avatar customization on the efficacy of an anxiety reducing attentional retraining task. My results show that avatar customization increases motivation over time and in the moment of use, suggesting that avatar customization is a viable strategy to address the engagement barrier that thwarts the efficacy of technology-enabled services for health.

Acknowledgements

First and foremost, I want to thank my supervisor, Dr. Regan Mandryk. I'm grateful that I had the chance to work on your side for so many years and I hope that there are many more to come. Thank you for giving me guidance, inspiration, and friendship. You made me who I am today. Thank you for hearing my voice so loud and clear and believing in me every step of our journey. You are a force of nature and a revelation to work with—I could not have dreamed of a more dynamic, enriching, and rewarding training. You have my eternal gratitude.

I want to thank my thesis committee, Dr. Carl Gutwin, Dr. Ian Stavness, Dr. John Howland, and Dr. Gavin Doherty, for elevating my work and pushing me to excellence. I could not have asked for a better group of world-class experts to critique my work and challenge me where challenge was due.

I also want to thank the wonderful people I had a chance to work with over the years. By year of publication: Brett Taylor, Zenja Ivkovic, Dr. Kathrin Gerling, Dr. Andre Doucette, Diane Watson, Dr. Mark Hancock, Dereck Toker, Dr. Ben Steichen, Ian Livingston, Dr. Carl Gutwin, Matthew Miller, Dr. Rita Orji, Dr. Jan Smeddinck, Mike Kalyn, Dr. Christina Conati, Dr. Christoph Lürig, Jason Bowey, Dr. Ioanna Iacovides, Dr. Daniel Johnson, Dietrich Barsilowski, Dr. Rainer Malaka, Cher Atkins, Thomas Muender, Benjamin Buttlar, Susanne Poeller, Shelby Thomson, Dr. Nicola Baumann, Dr. Elisa Mekler, Dr. Stefan Rank, Sharon Steinemann, Dr. Ansgar Depping, Tushita Patel, Dr. Adam Lobel, Dr. Marieke von Rooij, Dr. Isabela Granic, Dr. Vero Vanden Abeele, Maximillian Friehs, Cale Passmore, Rowan Yates, Dr. Greg Wadley, Dr. John Torous, Jason Wuertz, and Dr. Scott Bateman. I sincerely hope that the future holds many opportunities to share ideas and realize great research——it was a pleasure working with every single one of you.

A special thanks goes to Jared Cechanowicz, Kristen Dergousoff, Roxanne Dowd, Greg Oster, Colby Johanson, Benj Hingston, and Ashley Coveney. Your support was essential to breathe life into the many shared projects we worked on.

I want to thank the HCI Lab and its many members for creating a space where research can thrive, ideas are turned into insights and knowledge, while laughter and creativity are around every corner. Thank you for creating a home away from home.

I want to thank my friends in Saskatoon for the many shared laughs, drinks, and for welcoming me as an equal. You will always have a special place in my heart and as much as I have longed to be back home in Germany, I will long for being back home in Saskatoon from now on——thank you for being amazing.

I want to thank my friends in Germany for staying in touch over long distance, making time and space for me when I was around, and welcoming me back home like I have never been away. I hope we can grow old together.

A special thank you to Jörg Stölting, who set my head straight and pointed me in the right direction.

Finally, I want to thank my family: my parents, who had my back throughout all the difficulties that living abroad entails, my sisters, who are magical creatures, my aunt and uncle, for their excitement and support, and my daughter Mia, the most wonderful child I could have wished for—I hope that my efforts and the generosity that I have received will pave your way to a bright future.

Contents

List of Tables

List of Figures

List of Abbreviations

ABMT	Attention Bias Modification Training
ADHD	Attention Deficit Hyperactivity Disorder
AA	Alcoholics Anonymous
ANOVA	Analysis of Variance
ANCOVA	Analysis of Covariance
ARPU	Average Revenue per Paying User
BPNS	Basic Psychological Need Satisfaction
BFI	Big Five Personality
CBT	Cognitive Behavioural Therapy
CCT	Computerized Cognitive Training
CL	Bias
dL	Sensitivity
Dota	Defense of the Ancients
FACS	Facial Action Coding System
GUR	Games User Research
HIT	Human Intelligence Task
HLM	Hierarchical Linear Modeling
HUD	Heads-up display
IAPS	International Affective Picture System
IMI	Intrinsic Motivation Inventory
JSON	JavaScript Object Notation
LoL	League of Legends
MMORPG	Massively Multiplayer Online Roleplaying Game
MOBA	Multiplayer Online Battle Arena
MTurk	Amazon Mechanical Turk
PANAS	Positive Affect/Negative Affect Scale
PENS	Player Experience of Need Satisfaction
PIS	Player Identification Scale
PTSD	Post-Traumatic Stress Disorder
RCTs	Randomized Control Trials
RQ	Research Question
SAM	Self-Assessment Manikin
SDT	Self-Determination Theory
SIMS	Situational Intrinsic and Extrinsic Motivation Scale
SCT	Social Cognitive Theory
STAI	State Trait Anxiety Inventory
SPSS	Statistical Package for Social Sciences
SQL	Structured Query Language
TIPI	Ten Item Personality Inventory
TES	Technology-Enabled Services
TAM	Technology Acceptance Model
UMA	Unity Multipurpose Avatar
WHO	World Health Organization
WoW	World of Warcraft

Chapter 1

Introduction

In its constitution, the World Health Organization (WHO), envisions "...the highest attainable standard of health as a fundamental right of every human being". As of today, the global community has focused massive resources into attaining this standard, utilizing established healthcare implementation patterns (Warner & Obrecht, 2015; Grol & Grimshaw, 1999), new technologies (e.g., drone drug delivery[1]), and by pushing the boundaries of future inventions (e.g., precision medicine; see Collins & Varmus, 2015; Insel, 2014). In the past, health care of any form required at least one medically-trained person to be physically accessible for patients, making accessibility of location and human resources—next to drug availability—the main determinants of a steady health care supply. However, population growth and a shortage of trained personnel have created a massive gap in mental and physical health care, displayed in low income countries as declining health care provision—and for mental health issues a lack of treatment altogether (World Health Organization, 2017). The World Health Organization Mental Health Atlas 2017 shows for mid to high income countries an increase in prolonged waiting times, shortened doctoral attention, and an increase in treatment errors. The costs of the health care gap are estimated to be in the trillion dollars globally (Chisholm et al., 2016), but even more alerting is the unnecessary suffering of millions due to a lack of solutions able to effectively expand current health care services to meet the global need for treatment—a situation that justifies novel solutions and unconventional approaches.

Global connectivity and health services enabled by technology have drastically shifted the paradigm of global health care. Technology-enabled services (TES) for physical and mental health have shown huge potential in unburdening existing health care infrastructures (Bennett & Glasgow, 2009; Chisholm et al., 2016; Cuijpers et al., 2017; Mohr et al., 2014) and in creating novel infrastructures (Dinakar et al., 2014; Morris et al., 2015) to close the health care gap. MoodGYM[2], for example, offers multiple training modules to address anxiety and depression. RCTs have shown that MoodGYM is effective when used correctly and when all training modules are completed (Christensen et al., 2002). IntelliCare (Mohr, Tomasino, et al., 2017), is another example, a service that provides a framework of 13 different mobile apps to provide support such as a positive message for stressful situations or sleep tracking applications. IntelliCare shows promising results to support general mental health (Mohr, Tomasino, et al., 2017).

[1] http://www.flyzipline.com/
[2] https://moodgym.anu.edu.au/

While the prospects of TES to revolutionize the health care system gives reason for optimism and excitement, the implementation reality of these services looks different. The largest barrier for an effective treatment through TES is patients losing interest in using the technology over time. The loss of motivation results in a situation in which novel treatments are effective, but the efficacy is thwarted because applications are simply not used, or not used as required (Christensen et al., 2002; Eysenbach, 2005). High attrition (i.e., the gradual decline of attendance) diminishes even the most effective treatment—a situation that is known from drug treatment protocols where the drug itself has shown to reduce symptoms, but the unsteady intake of patients negatively affects treatment efficacy (Vermeire et al., 2001). While attrition is a threat to treatment efficacy in general, it is the primary threat to the efficacy of technology-enabled services for health and the main reason why TES—which have shown good efficacy in clinical studies (Königbauer et al., 2018; Heber et al., 2016; Webb et al., 2010)—fail to show the same levels of efficacy "in-the-wild" (Eysenbach, 2005; Firth et al., 2017, 2018).

Most treatments, especially when they rely on conveying complex information or attention-based training, require patients' engagement in the moment of use—to maximize the efficacy of a program, patients need to focus on the task at hand, remember information, and invest effort (Zimmerman, 2000). To secure that patients are invested, traditional health programs rely on external regulation such as controlled environments (e.g., clinics) or direct supervision (e.g., in a doctor's office). However, technology-enabled services for health lack such levels of external regulation; health applications designed for the use during commutes or in the privacy of our homes are specifically designed for unsupervised use, but that does not free them from the same attentional requirements that are applicable in conventional approaches. For example, learning about the symptoms of a panic attack, and breathing techniques to keep an episode under control, requires a user's attention for maximal efficacy (Grossberg, 1999)—a demand of TES that is particularly difficult to adhere to. Without any external control, it is difficult to actively and repeatedly focus our attention on often boring tasks. Imagine engaging with a service at home, where a text message might cause distraction, the TV is delivering a constant stream of information, or checking Facebook divides our attention every few minutes—a situation that makes it challenging to engage with a service and experience its full benefits. While TES could be used in a clinical context, the required resources diminish the advantages of TES in increasing accessibility and decreasing human support. Requiring user engagement without external regulation is a crucial vulnerability of TES and thwarts the clinically-shown efficacy of TES "in-the-wild".

Research on TES shows that treatment efficacy, in the short and long-term, matches conventional approaches (Christensen, Griffiths, & Jorm, 2004; Richards et al., 2015), but when transitioning from controlled studies to implementations in the wild, the lacking external regulation in TES requires careful consideration—attrition and reduced in-the-moment engagement thwart the efficacy of TES (Mohr, Weingardt, et al., 2017).

To address the lack of external regulation in TES, researchers have successfully integrated external regulation such as human support (Christensen et al., 2009; Tate et al., 2006; Andersson & Cuijpers, 2009), notifications (Kelders et al., 2012), or nudging (Thaler & Sunstein, 2008). Baumeister et al. (2014) reviewed

14 articles comparing human supported or guided internet-based interventions with unsupported or unguided interventions. The results suggest that guided therapy forms have better effects on symptom severity, general completion rate, and a higher number of completed training modules. The effect of notifications is more complex, because while notifications show an increase in attendance in the short-term (Kaplowitz et al., 2004), they come with attentional cost (Horvitz et al., 1999) caused by a steady stream of notifications from multiple apps, resulting in disabled or ignored notifications in the long-term (Almuhimedi et al., 2015; Iqbal & Horvitz, 2007; Weber et al., 2016). Nudging—or architecting choice—has been very successful in areas like food choice (Hanks et al., 2012), mobile security (Almuhimedi et al., 2015), and water consumption (Kenney et al., 2015). The critique towards nudges develops around the argument that active choice is avoided, which is an important factor for long-term behaviour change (Ryan & Lynch, 2011; Ryan & Deci, 2008). Motivational theories such as Self-Determination Theory, suggest that volitionally engaging in a task is predictive of intrinsic motivation and therefore long-term behaviour change.

While these solutions based on external regulation provide promising results, they are either resource intensive, e.g., human support requires a trained expert, or they leverage methodologies that have been shown to be effective in the short-term, but do not result in long-term engagement (i.e., notifications). Hence, integrating external regulation into TES is not a viable solution when aiming to create services that are effective and show sustained use. Other technologies have successfully deployed motivational design strategies—educational services (e.g., Duolingo[3]), entertainment software (e.g., video games), or knowledge-sharing platforms (e.g., Quora[4]) leverage motivational design strategies to keep users engaged over time and in the moment. While other areas have shown that motivational design is effective, these strategies are currently underutilized in TES for health.

As such, the problem I address in my thesis is *that current technology-enabled services for health underutilize inherently motivating design strategies, thwarting the sustainable and effective implementation of such services in health care systems.*

1.1 Proposed Solution

Motivational strategies have mainly been implemented in commercial applications, such as clearly communicated goals for advancing grammar and vocabulary in language learning software (i.e., Duolingo) or providing positive feedback to correct answers to complex questions (i.e., Quora). Video games are a rich resource for motivational strategies, for example by seamlessly combining goal-setting techniques (e.g., through quest progression) with staged reward mechanism (e.g., receiving small reward consecutively, but a large fixed reward at the end of multiple quests). The ability to combine motivational approaches seamlessly in the game context allows designers to create novel approaches to motivate players and is core to the powerful

[3]https://www.duolingo.com/
[4]https://www.quora.com/

motivational engine of video games. Video games have been shown to foster engagement in the moment—e.g., a person who can barely speak because they are deeply engaged with their favorite game (Sweetser & Wyeth, 2005)—and sustainably over time—e.g., some World of Warcraft players, for example, play their characters for years (Rapp, 2017; Braithwaite, 2018). Game-based motivational design strategies (Deterding, 2016), applied outside of the game context, have been shown to foster intrinsic motivation by enhancing user experience and increasing task engagement (Trepte & Reinecke, 2010).

Fostering intrinsic motivation through game-based motivational design is a promising solution to overcome shortages of in-the-moment efficacy and sustainable usage of technology-enabled services over time and surpasses the limitations of external regulation. Game-based motivational design strategies such as goal-setting (Locke & Latham, 2006) (e.g., quests in World of Warcraft), social comparison (Festinger, 1954) (e.g., leaderboards in Fortnite), or customization (e.g., the customization of the main protagonist in Mass Effect), are crucial components of successful games and allow players to experience mastery over difficult challenges, connect to other players, or feel ownership over a central aspect of the game.

Customization plays a central role when designing for motivation in games. Creating customizable experiences serves several purposes, such the experience of choice, ownership, and identification. Specifically, avatar customization is a game-based motivational design strategy that has been shown to increase task engagement, enjoyment, and intrinsic motivation (Trepte & Reinecke, 2010), is versatile and applicable across contexts (Vasalou & Joinson, 2009), and is commonly used by game-designers (Ducheneaut et al., 2009), and therefore well suited to be investigated as a potential solution for the shortcomings of technology-enabled services.

In my dissertation, I investigate whether a game-based motivational design strategy (i.e., avatar customization) fosters inherent motivation to engage with technology-enabled services for health in the moment and over time, with the goals to increase immediate task efficacy and facilitate sustainable engagement.

1.2 Steps to the Solution

To accomplish the aims of my research, I conducted four studies[5]: In a first study, I investigated if design techniques drawn from games (i.e., avatar customization) can help to facilitate voluntary engagement, measured both subjectively through validated scales, and objectively through measuring motivated behaviour. In a second study, I investigated if voluntary engagement is responsive to known manipulations of motivation (i.e., overjustification effect) in the medium-term, measured as logging behaviour. In a third study, I investigated the effectiveness of motivational design to combat waning motivation over a 3-week long study

[5] All manuscripts and attached studies that I present as part of my dissertation are a team effort and would not have been possible without the work of my colleagues. I will acknowledge the individual contribution at the end of each Chapter. However, to distinguish between the collectively created manuscripts and the enclosing dissertation sections, and to highlight my intellectual contribution in designing the research in coordination with my supervisor, my efforts in contextualizing the contribution as part of my dissertation and reporting the design rationale behind each contribution I will use "I", in the dissertation sections and "we" in the manuscripts.

conducted in the context of evaluating an online health intervention. And in a fourth and final study, I applied the gained insights to investigate if volitional engagement can improve in-the-moment engagement of participants, and subsequently affect training efficacy of computerized attention-bias retraining.

The four manuscripts presented in this dissertation address these individual facets of avatar customization and its potential effects on task efficacy and sustainable engagement: Manuscript A investigates whether avatar customization is suited as a design strategy to increase motivation, operationalized as increased user experience and sustained short-term behaviour; Manuscript B investigates if the effects on intrinsic motivation induced through avatar customization as found in Manuscript A are exhibited in the medium term and are reliable (i.e., responsive to known manipulations of intrinsic motivation). Building on the insights gained in Manuscripts A and B, Manuscript C investigates the effects of avatar customization on user retention over a 3-week period, while Manuscript D investigates the effects of avatar customization on in-the-moment engagement and task efficacy. Manuscripts A, B, and C have been published and Manuscript D is in press at the time of writing this dissertation. The citations are:

Manuscript A: Birk, M. V., Atkins, C., Bowey, J. T., & Mandryk, R. L. (2016, May). Fostering intrinsic motivation through avatar identification in digital games. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 2982-2995). ACM.

Manuscript B: Birk, M. V., Mandryk, R. L., & Atkins, C. (2016, October). The Motivational Push of Games: The Interplay of Intrinsic Motivation and External Rewards in Games for Training. *In Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play* (pp. 291-303). ACM.

Manuscript C: Birk, M. V., & Mandryk, R. L. (2018, April). Combating Attrition in Digital Self-Improvement Programs using Avatar Customization. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (p. 660). ACM.

Manuscript D: Birk, M. V., & Mandryk, R. L. (in press). How Customization Improves the Efficacy of an Online Digital Intervention for Mental Health. *Journal of Medical Internet Research.*

1.3 Design Rationale

I chose video games as a source for motivational design strategies, because the medium itself has shown substantial effects on motivation across platforms (i.e., from desktop to mobile play). Games are comprised of individual design strategies—e.g., social comparison through leaderboards (Bowey et al., 2015), goal-setting through quests (Chen et al., 2012), or personalization through avatar customization (Trepte & Reinecke, 2010)—which when combined, create engaging experiences with application potential beyond games. Researching the effects on experience and behaviour of these strategies has the potential to affect design decisions and the uptake and long-term use of technology-enabled services.

Challenges such as reaction time tasks, design strategies such as avatar customization, or moral decisions such as found in the Walking Dead by Telltale (2012), make games sophisticated sandboxes in which to study human behaviour—games allow researchers and designers to create, manipulate, and measure human behaviour in an interactive context. Games allow researchers to answer complex applied questions—i.e., can retention be increased, or can the reaction time be reduced in a shooter game by changing the target appearance—but also questions about experience and behaviour in general—e.g., the effect of individual differences on experience (Birk, Mandryk, et al., 2015) or how previous in-game experiences, for example in a tutorial, affect future behaviour. The number of players who can play a game in parallel (i.e., single player vs. multiplayer), and how the game is played (i.e., cooperative or competitive), varies—enabling games to be a rich research platform for answering social research questions such as those on remote collaboration or on how to design for friendship (Bell, 2012). Games provide a rich and versatile source of design strategies, while their interactive nature allows for experimentally complex designs (Blascovich et al., 2002)—properties that fit well within the current socio-technological state of society in which interactive technology is ubiquitous in our daily life. The entertaining nature of games make them appealing to many, independent of age, sex, or race (Birk et al., 2017; Passmore et al., 2018; Entertainment Software Association, 2017). Therefore, design strategies drawn from games have a high chance to be already familiar to the user, and hence more likely to be accepted.

Digital games produce large amounts of data—technically, every bullet fired, every time the mouse was clicked, and every second of idle time can be measured. This large amount of data has advantages from a research perspective because behavioral outcomes can be triangulated in data—i.e., multiple data sources such as time spent running, number of coins collected, and time spent to defeat a boss—and can be used to change a player's play experience, for example by balancing skill levels between players (Vicencio-Moreira et al., 2015). Furthermore, data resulting from behaviour can be used as an outcome measure for experimental manipulation, e.g., does a player run more or less after being exposed to severe social stress? Finally, data reported back to the player, e.g., leaderboard statistics or in-game item stats, can be easily manipulated, allowing researches to create research paradigms that manipulate the player's experience (Bowey et al., 2015).

1.3.1 Advantages of Engaging Study Designs

The interactive and complex nature of games makes it challenging to balance outcome validity and generalizability (Blascovich et al., 2002). Interactive systems geared for maximizing internal validity, i.e., minimizing systematic error in favour of a causal conclusion, such as the CANTAB[6] (Cambridge Cognition) tests are able to tease out minimal differences in cognitive abilities, but the results might be difficult to replicate under less controlled conditions, and the applicability outside of the lab might be minimal. But using a complex commercial game might be an overwhelming experience for participants and any difference in experience or behaviour could be attributed to multiple sources (Depping et al., 2016). With respect to these challenges, my research environments are built from scratch, allowing me to manipulate and control every aspect of the game to prioritize internal validity while maximizing external validity.

[6] www.cambridgecognition.com/cantab/

Building research environments instead of using or modifying off-the-shelf games has several advantages. First, researchers can create standardized and controlled systems, which increase the internal validity and reliability of the results. For example, in my research environments, secondary stimuli, e.g., audio or Heads-up display (HUD) elements, are removed to increase the focus of participants on the task at hand. Elements that are not task relevant might be distracting and introduce another source of information for participants that needs to be controlled. Second, while creating controlled environments, which are comparable across conditions except for the experimental manipulation of interest, custom-built games allow researchers to create immersive game experiences that are well received by participants, as shown by positive written feedback (e.g., "I just wanted to drop you a message letting you know that your hit was AWESOME. First one I've had in over a year that was actually engaging, paid well, and was fun to do because it was about something that I could relate to.") and experience scores (e.g., Manuscript C). Third, custom-built games allow researchers to create specific logging events a prioi, which increases meaningfully gathered data, reduces analysis time, minimizes processing time after the data collection phase, and allows researchers to use additional metrics in real-time during a study. For example, the game used in Manuscript A allowed me not only to log time spent and coins collected, but in addition every keypress made during the study, which I used to assure that participants actively engaged in the task.

1.3.2 Assuring Engagement

Computerized studies have the advantages of being highly standardized, not requiring human interference, and producing already digitized data logs; conversely, they are less controlled than traditional experimental settings. Especially when disseminated in an online context, computerized interactive studies require the implementation of strategies to minimize the impact of distractions, attentional decline over time, or task interruptions.

Design strategies would require to either assure that the participant shows continuous task engagement (e.g., by measuring continuous input such as key presses), or measures to be aware of a user's disengagement (e.g., a count of window changes during the experiment), or ways of observing participants and their environment, using a participant's camera to capture distracting events (e.g., relatives showing up, a phone ringing, or a cat walking across the keyboard).

1.3.3 Designing for Experimental Manipulation

Not all games are equally suited to be used in an experimental context. For example, creating a fully-functional massively-multiplayer role-playing game such as World of Warcraft to manipulate social dynamics in digital societies might be a bit too ambitious, whereas coming up with entirely new game mechanics to answer specific research questions is not out of the question. However, it is often more efficient to build on pre-existing game mechanics and adjust the game according to the research question. Comparing, for example, the logic of Match-3 games and infinite runner games, we find different properties, which have different advantages for addressing different research questions.

Match-3 games require players to swap two adjacent items per move to match at least 3-items of the same property—often color; once matched, the items disappear, and three new items take the place of the removed items. The playtime is usually restricted, either by time directly or by performance, operationalized as number of points at a given time. The random nature of the board allows players to easily and believably manipulate outcome—players can't keep track of their exact score, and when compared to others they have a difficult time to judge how good their relative performance was. Beyond the properties of randomness, performance in Match-3 games depends on an understanding of the scoring system (i.e., a cascade of multiple matches is worth more than a sequence of individual matches) and the abilities of recognizing patterns and correctly predicting how the board will change after a move.

Infinite runner games usually start out with the player's character in the middle lane of a three-lane track. The player constantly moves forward with no means to slow down, while obstacles appear in different lanes. The player can only change lanes (i.e., left and right), crouch, jump, or use special items such as shields or boosts. The genre name comes from the fact that infinite runners can technically be played infinitely, because the obstacles are procedurally generated. Infinite runner games are well suited to test the time players will stay engaged in the game after manipulating certain design elements, such as the fidelity of the game—a manipulation that has been shown before to affect the experience of casual game play (Gerling et al., 2013). Infinite runner games can also be used in stop-signal tasks (van Muijden et al., 2012)—an attentionally-demanding task that requires participants to stop a continuous activity immediately as indicated by a stop-signal. The procedurally-generated nature of infinite runner games also allows for other manipulations, such as spacing out rewards or manipulating difficulty through the numbers of obstacles presented. Infinite runner games can also be used to test changes affecting continuous behaviour adjustments using previous behaviour (i.e., feedforwarding) by, for example, keeping the level layout constant and presenting a "ghost" of the player from a prior play through.

The selection of a game and a thorough understanding of the underlying principles is crucial to select a game that has the potential to test a hypothesis rigorously and to avoid unnecessary development time. Justifications for the games chosen to answer each research question are presented in the individual manuscripts.

1.4 Theoretical Rationale

To investigate whether a game-based motivational design strategy (i.e., avatar customization) fosters inherent motivation to engage with technology-enabled services for health in the moment and over time requires a theoretical framework that allows for the definition of the motivational properties of game-based strategies (i.e., a theoretical understanding of why a strategy increases engagement in the moment and/or over time).

In my comprehensive exam depth paper, I surveyed several motivational theories that could capture the dynamic, complex, and fast-paced context of video games, with an already existing body of literature to inform and motivate design decisions. Theories such as the Technology Acceptance Model (TAM) (Venkatesh et al., 2003), Goal Setting Theory (Locke & Latham, 2006), or Social Cognitive Theory (SCT) (Bandura, 2001), offer interesting insights into why users accept technology, how goal setting increases the likelihood of achieving a task, and the importance of the social context and our perceived self-efficacy for task performance. However, most theories are either too broad or too narrow and do not capture well the socio-contextual variables required to increase user engagement.

1.4.1 Self-Determination Theory

The best theoretical fit to describe the motivational pull of video games is Self-Determination Theory (SDT) (Ryan & Deci, 2002), which proposes that people are more likely to engage in a task when the socio-contextual variables fulfill their inherent need for competence—mastery of a task; autonomy—acting under their own volition; and relatedness—feeling connected to others. The theory has been applied in many different areas, such as work (Amabile, 1993), health (Seaverson et al., 2009), and sport (McAuley et al., 1990), and has been applied in the context of games research before (Przybylski et al., 2012; Johnson et al., 2016; Birk, Mandryk, et al., 2015).

While I applied and measured need satisfaction as defined by SDT in Manuscript A, B, and C, my experience with SDT in the context of games goes beyond the work presented in my dissertation. I contributed to the interpretation of core SDT constructs in the context of research on the experience of touch interfaces (Watson et al., 2013); the effects of controller type on player experience (Birk & Mandryk, 2013); the effect of self-esteem on player experience (Birk, Mandryk, et al., 2015); the manipulation of leaderboards for player experience (Bowey et al., 2015); and when investigating the effects of age differences on player experience (Birk et al., 2017).

Self-Determination Theory as a measure of player experience has also been picked up by the games industry, which applies the Player Experience of Need Satisfaction (PENS) Scale in user studies to evaluate player experience. The company Immersyve[7] commercialized Self-Determination Theory in the context of games and offers their services to game companies.

1.4.2 Avatar Customization and Identification

Games have been shown to motivate players to stay engaged with a single game for hours, independent of whether the game is an immersive AAA title like God of War, a social network game like Farmville, or a casual mobile game like the infinite runner Alto's Adventure.

In the book "Glued to Games: How Video Games Draw Us in and Hold Us Spellbound", Rigby and Ryan (2011) discuss a conceptual perspective on motivation through the lens of Self-Determination Theory (Ryan & Deci, 2002). The general idea is that people inherently like to grow, and seek out situations in

[7] http://immersyve.com/

which they can experience personal growth. Situations that allow us to experience competence, autonomy, and relatedness are most likely to facilitate personal growth either by applying our skills (competence), by experiencing the consequences of our own decisions (autonomy), or by growing through interactions with others (relatedness). The book argues and presents evidence that within video games, people can experience situations of choice, challenging situations, and social encounters that are in line with their needs, resulting in increased motivation and engagement in the moment and over time.

The game context offers plenty of opportunities to experience situations and challenges that satisfy our need for personal growth; when well designed, games facilitate motivation to play and high levels of enjoyment and invested time and effort. Designs like leaderboards, for example, allow for social comparison and allow players to understand their performance in context. The autonomy to make decisions and experience consequences can be elevated to an overall design principle, as for example done in many open world games. Zelda Breath of the Wild (2016, Nintendo), for example, restricts the number of weapons a player can carry and weapons also break after a certain number of uses. Following Pears (2017) the design intentions are: first that players need to decide which weapon to use when; and second, that players need to explore and find new weapons, affecting how they interact with their environment. Designing for relatedness can include the design of interfaces to support communication—be it as text chats, predefined commands, or gestures; however, designing for relatedness can also be expressed through the design of mechanisms against toxicity, e.g., by restricting communication or nudging players to make different communication choices. Both approaches are in favour of increasing social cohesion and an experience of connectedness to others.

In line with Self-Determination Theory, avatar customization should increase volitional engagement by giving participants the means to decide how their experience, at least visually through the representation of their avatar, will be. Research has shown that we experience increased ownership over our own creations (Franke et al., 2010), which should positively affect engagement. People identify with their avatar on different levels (Van Looy et al., 2012)—how similar they are, if they can see themselves being the avatar, or perceive an avatar as an ideal.

1.4.2.1 Identification

An important construct in my thesis is identification with an avatar. People identify with representations for different reasons. Van Looy et al. (2012) suggest that there are three different ways we identify with a representation: similarity identification, wishful identification, and embodied identification.

First, research has shown that people identify with an avatar that looks similar to ourselves in physical appearance or because of personality aspects (Van Looy et al., 2012; Trepte & Reinecke, 2010). Second, people identify with avatars that have ideal characteristics; for example, because they are particularly strong or beautiful or have other characteristics that we might wish to possess, e.g., being able to fly. Third, people experience presence in a virtual environment because the environment can be experienced through an avatar's body—an example question from the Player Identification Scale (Van Looy et al., 2012) (i.e., the scale used to measure the three constructs of identification) would be “In the game, it is as if I act directly through my character”.

Because of the benefits of identifying with an avatar, various methods for facilitating this relationship have been suggested. Trepte and Reinecke (2010) showed that customizing an avatar increases identification and leads to higher enjoyment, and also that creating an avatar that has high similarity to the player increases identification, and ultimately game enjoyment. Schneider et al. (2004) showed that participants who played a first-person shooter game with a narrative experienced more identification with their avatar than those who played a non-narrative version.

In my research, identification was measured after participants customized an avatar—similar to the work done by Trepte and Reinecke (2010)—and was investigated as a variable explanatory for user engagement (see Manuscript A, and B).

1.4.3 Measuring Motivation

When measuring motivation, we need to distinguish between lasting motivation and motivation in the moment of use. Motivation in the moment is the joy we experience engaging in a task, and the effort invested, reflected for example in performance.

We show lasting motivation when we repeatedly come back to a task, e.g., a training exercise. Self-Determination Theory distinguishes different levels of motivation (Deci et al., 2001): intrinsic motivation—engaging because we inherently enjoy the situation, e.g., enjoying ourselves while playing beach volleyball during a vacation; introjected regulation—engaging because we believe that doing so has positive implications, e.g., running on a treadmill; and external regulation—engaging because an external source (e.g., an authority) tells us that it would be beneficial, e.g., participating in school sports not because we like sports, but because otherwise we would receive a bad mark. And lastly, sometimes we are not motivated to engage at all, i.e., being amotivated. Distinguishing why someone engages in general is important, because it allows us to predict if they will show the same behaviour in the future, and allows us to understand the circumstances under which motivation would wane, e.g., when removing an external source of regulation, such as removing mandatory sports classes.

While the theoretical underpinnings of motivation over time are relevant to understand and explain engagement over time, the measures are self-reflective and affected by expectations, the experimental situation, incentives, and in lab experiments by the demeanor of the experimenter. Not asking explicitly about a participant’s motivation, but using implicit, projective tests such as the Thematic Apperception Test (Murray, 1943) or the Operant Motive Test (Kuhl, 2013), offer measures less prone to situational effects. Unfortunately, these measures have to be evaluated by hand and require very specific expertise to be used correctly, which renders them difficult to use in larger studies.

One of the best measures of motivation over time is actual behaviour, for example, login behaviour or time spent with a service. Especially in an online context, in which participation is subject to few social demands and engagement cannot be affected by the experimenter's expectations, returning to a task on a daily basis or engaging in an attentionally-demanding task under our own volition are very good indicators of motivation. Displayed increases in behaviour can be explained through subjective motivational scales and would be in support of the external validity of a study.

1.4.4 Individual Differences

People are different from each other when it comes to how they experience different situations—games are no exception. Gained expertise, cultural influences, preferences, and personality traits affect how games are experienced. A professional guitar player will experience guitar hero differently, and a player of colour might experience the predominantly white culture in games differently than a white player (Passmore et al., 2018). Research has shown that individual differences explain important differences in play experience. Although not included in this dissertation, I have published several studies exploring individual differences in the context of play.

In early work, I investigated how self-congruence (i.e., congruency between who a person is and who they want to be) and personality affect play experience using different controller types. The research demonstrated that player experience is not driven by a player's overall personality, but by the personality experienced within the context of the game (Birk & Mandryk, 2013).

Further, I have investigated how self-esteem affects play experience (Birk, Mandryk, et al., 2015), and showed that players with different levels of self-esteem experience games differently; specifically, people with low self-esteem experienced lower satisfaction of needs (e.g., autonomy), which potentially can be influenced through game-design (e.g., creating choice).

Furthermore, I have investigated how personality traits affect play experience and subsequent behaviour (Birk, Toker, et al., 2015); the results suggested that individual differences alter how players experience a social network game, which is reflected in players' in-game behaviour.

Finally, I have investigated how age affects game preferences and experience (Birk et al., 2017). The results suggest that with increasing age, preferences, play motive, play style, identification as a gamer, and player experience shift away from a focus on performance and towards a focus on completion, choice, and enjoyment.

1.5 Methodological Rationale

To investigate whether avatar customization fosters inherent motivation to engage with technology-enabled services for health in the moment and over time, with the goal to increase immediate task efficacy and facilitate sustainable engagement, my research required different methodologies.

It was important to work in a context in which quick iterations and easy adaptation of methodology would be possible, with the ability to rapidly collect behavioural and experiential data.

1.5.1 Experimental Platform

To gather participants' data in different settings, I designed a computerized experimental delivery platform that would allow for easy experimental flow control and data collection, and would allow me to seamlessly present questionnaires and experimental stimuli, such as games or negative mood inductions. The platform was developed using Python, the microframework Bottle, and the lightweight SQL database SQLite.

I chose to build the experimental platform from scratch to have full control over the delivery and logging format and to allow for an effortless integration of third-party plugins. Avoiding commercial software like SurveyMonkey, Qualtrics, or FluidSurvey also allowed me to have full control over data storage and avoid that any participants' data was unintentionally made available outside of the experimental context.

The system was designed to be flexible, give developers and experimenters full control, and allow for high data mobility. The system allowed me to easily integrate dynamic content, such as the Unity player or JavaScript-based applications.

Based on my specifications and experiences with handling the experimental presentation, data collection, and feedback to the user, the platform was later polished and re-implemented to increase usability and reduce grown implementation complexity. The improved system—working title "Bride of the Frankensystem"—was built by the PhD Students Colby Johanson and Jason Bowey. The system implements the model-view controller pattern to display questionnaires, instructions, and dynamic systems, while providing a sophisticated login architecture. The system is built using Python, SQLite, the microframework Flask, SQLalchemy, Jinja2 as a general template engine, a custom built JSON interpreter for easier questionnaire handling, and JQuery for styling forms.

1.5.2 Unity

I chose Unity as the Game Engine for all implemented games and game-features, because it features a large number of predesigned assets, has open-source games available for modification (saving development time), a large and active community, and is relatively easy to learn. Using Unity resulted in reduced development time and more time spent on getting crucial experimental design details right.

1.5.3 Experimental Environment: Amazon Mechanical Turk

Amazon Mechanical Turk (MTurk) is a marketplace for Human Intelligence Tasks (HITs) that connects workers (i.e., people looking for online work) with requesters (i.e., people who request workers for online work). HITs are tasks that require human input to be solved, e.g., image tagging. Amazon Mechanical Turk is the largest crowdsourcing platform; the World Bank estimated that in 2015, there were 500,000 active workers, and the web tracker Alexa.com reports 750,000 unique visitors per day (Hitlin, 2016). Official numbers do not exist.

1.5.3.1 Advantages of Amazon Mechanical Turk

As an experimental environment, I chose Amazon Mechanical Turk, due to several criteria: 1) The large number of workers—estimates suggest more than 500,000 active workers (Hitlin, 2016). 2) The potential for quick iterations, even when running complex experimental setups, such as multi-day studies. 3) Being able to gather different sample sizes dynamically, for example to test data logging or the functionality of other experimental components, such as the Unity or JavaScript-based modules. 4) Access to a sample more diverse than University students in its socio-economic status; see for example (Kees et al., 2017). 5) People perform the study at home, which removes limitations caused by the experimental setup in the laboratory such as observation-induced compliance biases. But even more important, I could study people in a context that is not externally regulated like a laboratory would be. Considering that I'm interested in researching volitional engagement, this last point assures that lab-induced biases are removed from the study.

1.5.3.2 Limitations of Amazon Mechanical Turk

While there are several advantages of crowdsourcing studies, there are also limitations: 1) The most obvious limitation is the lack of control over the participant's context, i.e., the environment they perform the study in, including distractions and other uncontrollable events. 2) Participants can take breaks, which might have negative effects on the study outcome. For example, emotional induction effects usually wear off after a short amount of time (Buchanan et al., 1999). 3) People need to be instructed clearly on how to perform the study. It is important to instruct participants to work quickly and without breaks. 4) Because of the lack of external controls, researchers should evaluate if participants follow instructions by implementing controls and filtering data based on these controls. For example, to sense when participants disengage with a task—be it to take a break or because they got distracted.

1.5.3.3 Measuring Worker Behavior and Outlier Detection

Because of the lack of controls that are given in a lab study, it is beneficial to log user behaviour besides other experimental data such as questionnaire data or performance data (e.g., accuracy in a shooter game). Logging the time spent filling in a questionnaire or the number of times the browser window was not in focus provides valuable information about differences in user behaviour and potentially helps to identify outlier behaviour. The answers to the questionnaires themselves can also be used in a more fine-grained analysis of behaviour—the variance users create when filling in questionnaires or the internal consistency when answering questions in support of the same construct, e.g., Likert-scale responses to items such as "I feel capable doing X" or "I feel competent at X" should produce approximately the same answer.

Another common practice for ensuring quality data in crowdsourced experiments is to include questions like "If you read this, please chose the leftmost response" to assure that participants are attentive (Buhrmester et al., 2011). However, my experience (and reading worker forums) suggested that workers feel overly controlled and mistrusted by these kinds of questions. After carefully controlling multiple datasets using attention questions to identify careless and inattentive work, I omitted attention questions as a mechanism to assure data quality and instead relied on the consistency, variability, and timing of questionnaire responses.

1.5.3.4 Exclusion Criteria

In my work, I logged time-spent on questionnaires, evaluated response behaviour such as the produced variance when responding to an item, and instructed participants clearly about how to fill in questionnaires. I decided not to use attention questions, because of the negative responses from the MTurk community. For all collected datasets, I set exclusion criteria based on the number of violations for three previously-defined criteria. I created a count for violation of my criteria—a flag—and removed participants based on the kind of violation and/or the number of violations; the applied criteria vary between studies based on the importance of a criterion for a study (e.g., how time sensitive filling in the study is) and can be found in the individual manuscripts. Criterion 1: Participants producing zero variance between items in more than two questionnaires were flagged as violating my inclusion criteria and later removed independent of other criteria. Responding with zero variance indicates that questions were not read and that the fastest way possible to fill in a questionnaire was chosen. Criterion 2: I calculated violations of the time to fill in a questionnaire dynamically. I set taking less than -1 standard deviation as the lower boundary, i.e., people who were faster than the majority of all responses were flagged, and larger than +2 standard deviations, i.e., people who responded slower than 97.8% of the sample. Only people who showed multiple violations were removed—this criterion varied between studies depending on the study's sensitivity to response time. Criterion 3: I flagged outliers by looking at participants' responses to individual scales. Participants who scored higher than +3 standard deviations on multiple scales, and participants who scored lower than -3 standard deviation on more than three scales were removed. The number of participants removed by these criteria were usually less than 10%.

1.5.3.5 Worker Treatment

When gathering data from workers on Amazon Mechanical Turk, it is important to keep in mind that they are real people doing the work for many different reasons. In my experience, treating workers fairly, regarding payment, results in workers being motivated to do a good job even when exposed to a complicated task that may even involve extra equipment such as a camera or a microphone (e.g., Miller et al., 2017; Muender et al., 2016).

Workers communicate with each other in forums such as turkernation.com or mturkforum.com and use ranking systems for tasks and requesters, i.e., Turkopticon. Used as a browser plug-in, Turkopticon allows workers to see a requester's ratings immediately, which allows workers to choose who to work with and incentivizes good requester behaviour. Amazon itself does not feature a ranking system for requesters at this point in time. I oriented my payment on the average minimum wages paid in the US, which is about $8 per hour and paid more for tasks that required additional effort; in general, payments averaged $9.725 per hour over the studies described in my dissertation.

1.6 Dissertation Outline

The remainder of my thesis is comprised of four manuscript chapters that each includes an introduction, a problem statement, a solution statement, the actual manuscript, a summary, and a discussion about the relevance of the manuscript in context of my thesis. At the beginning of each chapter, I introduce the study presented in the manuscript, followed by a problem statement that phrases the problem addressed in the manuscript and the motivation to address the problem. With regard to the problem statement, I present a solution statement and the steps I took to implement the solution including the research questions I asked and a detailed design rationale for specific decisions made during the design process. After presenting the actual manuscript, I present a short summary of the insights gained, and the lessons learned. Each chapter is closed with a brief discussion about the relevance of the presented manuscript in the context of my thesis.

After I have presented all four manuscripts, I discuss my findings in context of the overall problem statement of my thesis and present overarching lessons learned and potential future work. I finish my presentation with a conclusion statement. Supplementary material used in my thesis, such as questionnaires, can be found in the appendices.

Chapter 2

Fostering Intrinsic Motivation

Citation: Birk, M. V., Atkins, C., Bowey, J. T., & Mandryk, R. L. (2016, May). Fostering intrinsic motivation through avatar identification in digital games. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 2982-2995). ACM.

Acknowledgment: Under my direction, Jason Bowey and Cheralyn Atkins were responsible for the development of the avatar creator and the implementation of the infinite runner. They also implemented the data logging for the infinite runner. Cheralyn Atkins provided the technical description of the avatar customizer in Manuscript A.

In coordination with my supervisor, Dr. Regan Mandryk, I was responsible for the research direction, experimental design, system design, data gathering, data analysis, and reporting.

My goals in Manuscript A were to stimulate identification with an avatar in a short interactive application to investigate how identification translates into engagement with a casual game. I conducted a study with 126 participants on Amazon Mechanical Turk. Half of all participants watched a video of another person customizing a neutral avatar and the other half were given the opportunity to customize their own avatar. They key finding was that increased identification with an avatar positively affects play experience, reflected in increased self-reported measures of engagement and in engaging with the game for an increased period of time.

2.1 Problem and Motivation

While the effects of identification have been shown in a general game context (Trepte & Reinecke, 2010; Van Looy et al., 2012) (see Section 1.4.2), the question remains if avatar identification fosters motivation in a research context in which internal validity has priority. Games are created artistically coherent—meaning that all elements of a game such as the visual aesthetics, a potential narrative game narrative, and the game mechanics work together to create an experience. In a research context, however, researchers can only partially recreate the experience players have when engaging with a game that is solely created for fun. The way people are invited to an experiment, additional instruments such as questionnaires, and financial incentives to play prevent a coherent play experience. However, researchers and practitioners need well-evaluated design strategies that have been shown to be effective in inducing motivation in controlled, experimental contexts to further understand and leverage design strategies, for example, in an applied context such as in an online training application. From a research perspective, it remains unclear if design strategies such as avatar customization produce inherently-motivating experiences when investigated in a research context.

Depending on the application area, it is important not only to induce self-reported changes in motivation, but to actually show changes in behaviour. When researching persuasive strategies, for example, showing that attitude has slightly shifted as a result of a persuasive intervention is a result that needs to be interpreted with care when the goal is to induce behaviour change, because self-reported attitude measures suffer from social expectation (Ajzen, 1991), leading to a gap between self-reported attitude and behaviour in persuasive research (Cooke & Sheeran, 2004). In my dissertation, I am interested in the effects of avatar customization on motivated behaviour to improve technology-enabled services. Therefore, one goal of Manuscript A is to investigate the effect of avatar customization on motivated behaviour.

The problem addressed in this manuscript is that researchers and designers of technology-enabled services lack well-evaluated inherently-motivating design strategies that have been shown to translate into behaviour.

2.2 Solution and Steps to Solution

To address the lack of research on feasible design strategies, I investigated the effects of avatar identification on play experience and motivated behaviour. I decided to conduct my research in a crowdsourced online study and develop a thoughtful experimental protocol that aimed to address limitations of online experiments by design, prioritizing internal validity while maximizing external validity.

In an online study on Amazon's Mechanical Turk (MTurk), half of the participants were instructed to customize the appearance, personality, and skillset of an avatar; the remaining participants were assigned a random avatar (of their preferred sex) followed by a video of another user customizing the avatar. The avatar was then used by the participants to play four timed rounds of an exertion-in-the-small (Sheinin & Gutwin, 2014) infinite runner game, after which affect, intrinsic motivation, and need satisfaction were measured with the goal of determining how avatar identification relates to self-determined play. Finally, participants played a final untimed round of the game that ended after 30 minutes—a virtually infinite amount of time in a paid online study. Alternatively, participants could click on a quit button to end the game and proceed in the experiment. The untimed version of the game allowed for the measurement of volitional play.

2.2.1 Research Questions

To explore the role of avatar identification, I asked and answered five research questions:

RQ1: Does avatar customization increase identification?

RQ2: Does avatar identification improve in-game needs satisfaction (i.e., competence, autonomy, relatedness, immersion, and intuitive control)?

RQ3: Does avatar identification improve intrinsic motivation (i.e., enjoyment, effort, and reduced tension)?

RQ4: Does avatar identification increase positive affect or reduce negative affect?

RQ5: Does avatar identification translate into motivated behaviour (i.e., do players who identify with their avatar spend longer in the unending infinite runner?)

2.3 Design Rationale

When designing the study to answer my research questions, I was confronted with several challenges requiring carefully thought through design decisions. I needed to decide on the general context of the study. The play environment had several requirements: 1) The game needed to look like a game. 2) The game needed to allow participants to interact with an avatar in a believable context. 3) To measure free-play as a proxy of engagement, the game needed to feature variable play time. 4) The game also needed to provide challenge and allow for player input to feel like a game. 5) To increase internal validity and assure similar behaviour across participants, the game also needed to have a built-in mechanism that would prevent participants to leave the experimental situation for a few minutes and come back, quit mid play, or just wait inactive in front of the screen—all potential options in the privacy of their homes. 6) The game needed to allow me to measure engagement operationalized as self-reported experience and as displayed behaviour in the same context. Measuring behavioral engagement required a task with variable length to avoid ceiling effects. The internal validity of self-reported behaviour, however, benefits from a fixed length to keep the experience comparable between participants. To measure both in the same study, an appropriate game needed to support variable length.

I chose an infinite runner as a game, because it fulfilled all the above listed criteria. Choosing an infinite runner game provided a context that allowed participants to engage in a game with variable length and provided me with the means to measure player behaviour and assure engagement—the game only proceeded if players alternated between pressing N and M on their keyboard; if they stopped pressing these keys, the avatar stopped moving and the game time stopped, preventing participants' study progress. This turned out to be an important decision, because it allowed me to assure that all players stayed engaged over the time of the experiment and to remove those who didn't stay engaged from the analyses.

2.4 Manuscript A

Fostering intrinsic motivation with interactive applications can increase the enjoyment that people experience when using technology, but can also translate into more invested effort. We propose that identifying with an avatar in a game will increase the intrinsic motivation of the player. We analyzed data from 126 participants playing a custom endless runner game and show that similarity identification, embodied identification, and wishful identification increases autonomy, immersion, invested effort, enjoyment, and positive affect. We also show that greater identification translates into motivated behaviour as operationalized by the time that players spent in an unending version of the infinite runner. Important for the design of games for entertainment and serious purposes, we discuss how identification with an avatar can be facilitated to cultivate intrinsic motivation within and beyond games.

2.4.1 Introduction

When people are intrinsically motivated to complete a task—that is, they do so based on the inherent satisfaction derived from the action itself (Ryan et al., 2006)—there are many benefits. Broadly speaking, intrinsically motivated people are willing to invest more effort into a task and derive more enjoyment from it (Deci & Ryan, 2000). In the case of interactive technology, fostering intrinsic motivation with our applications should translate into more effort invested in the task at hand and more enjoyment as a result of using the application (Ryan et al., 2006). This increased engagement has implications for both the designers and consumers of interactive technology. For example, consider an educational application designed to help people learn a language; increased effort invested by the user could translate into improvements in language learning. Or consider a citizen science application in which people contribute to finding new proteins that cure diseases such as HIV/AIDS, Cancer, or Alzheimer's (Cooper et al., 2010); increased enjoyment using the application could translate into more use, and thus a more complete database of proteins. Because of the potential benefits of increased engagement, creators of interactive technology should ask how they can foster intrinsic motivation through design. One place that designers can look for motivation is digital games. Recent estimates suggest that more money is spent purchasing games (92b) than music (18b) and movies (62b) combined (Dring, 2015). Four out of five American households own a device that is used to play video games and 115 million Americans play games (Entertainment Software Association, 2015). Internationally, the global game market is expected to exceed 102 billion by 2017 (Entertainment Software Association, 2014). Although people sometimes assume that it is highly immersive console and computer games that drive the game industry, 35% of those same revenues are expected to be generated through smart phones and tablets, on which people tend to play games that are more casual in nature. With so much time and money being spent (by choice) on digital games, researchers have questioned what it is about games that make them so motivating to play (Ryan et al., 2006) and how we can translate these motivating features into non-game environments—a process known as gamification (Deterding et al., 2011). Serious games—games that leverage this ability to motivate behaviour and retain attention in serious contexts—have been effective at encouraging behaviour change and fostering activities that lead to learning (Rigby & Przybylski, 2009).

There are various theories that explain why games are engaging (Sweetser & Wyeth, 2005; Brockmyer et al., 2009); the most prevalent arises out of Self-Determination Theory (SDT) (Ryan et al., 2006). Being self-determined describes a state in which people have their basic psychological needs for perceived competence (i.e., demonstrating mastery over challenges), autonomy (i.e., doing so under their own volition), and relatedness (i.e., doing so while feeling connected to others) satisfied through the activity. Satisfying these needs leads to people who are intrinsically motivated to perform the activity. Designing with need satisfaction in mind is one way that we can design better games (Trepte & Reinecke, 2010); however, this solution works on the level of the game itself—it helps us build a better interactive application. There is also an argument for seeking ways to foster motivation through methods and approaches that apply across a range of applications.

For example, Trepte and Reinecke (2010) showed that creating an avatar that a player identifies with increases the imagined enjoyment of playing a game with that avatar. And given that so many players invest so much time into creating, equipping, and playing with their virtual representations in massively multiplayer online role-playing games (MMORPGs) (Livingston et al., 2014), it is not unreasonable to assume that extending the use of an avatar that players identify with into multiple new contexts would translate the enjoyment that stems from avatar identification into those new contexts.

However, although it is clear that identifying with an avatar has benefits in the game being played, there are several reasons why it is not feasible that these benefits will extend outside of the specific context of the game. First, there are issues with intellectual property and copyright; second, there could be mismatches between the environment that the avatar was created in (e.g., space fantasy adventure) and the environment in which it would be used (e.g., oil rig training application); third, there is a large proportion of users who do not play the kinds of games that foster avatar identification; and fourth, it is not clear whether there are clear benefits of avatar identification in the kinds of lightweight, casual games that are often used for serious purposes. Thus, it is reasonable to ask whether or not we can leverage the advantages of avatar identification seen in immersive and persistent games and translate them into other games and tasks that are temporary or casual.

In this paper, our goals are to stimulate identification with an avatar in a short interactive application, and then investigate how identification translates in a casual game, as opposed to a game that requires significant and deep engagement. We conducted an online study on Amazon's Mechanical Turk (MTurk), in which we told players that they would be customizing an avatar to use in a series of future tasks. We allowed half of the players to customize the appearance, personality, and skillset of their avatar; the remaining players were assigned a random avatar (of the same sex) and watched a video of another user customizing the avatar. The avatar was then used to play four timed rounds of an exertion-in-the-small (Sheinin & Gutwin, 2014) infinite runner game, after which we collected measures of affect, intrinsic motivation, and needs satisfaction to determine how avatar identification relates to self-determined play. Finally, we included a behavioural measure of motivation in which we had participants play a final un-timed round of the game that had no ending (there was a quit button in the corner that players could press), which allowed us to gather an objective and behavioural measure of their motivation.

We asked five main research questions:

RQ1: Does avatar customization increase identification?

RQ2: Does avatar identification improve in-game needs satisfaction (i.e., competence, autonomy, relatedness, immersion, and intuitive control)?

RQ3: Does avatar identification improve intrinsic motivation (i.e., enjoyment, effort, and reduced tension)?

RQ4: Does avatar identification increase positive affect or reduce negative affect?

RQ5: Does avatar identification translate into motivated behaviour (i.e., do players who identify with their avatar spend longer in the unending infinite runner?)

Our results showed that avatar customization stimulated identification with the avatar. In terms of game experience, we used hierarchical regressions with the three types of identification (similar, embodied, wishful) as individual continuous predictors to show that greater identification increased autonomy, immersion, invested effort, enjoyment, and positive affect. We also show that greater identification translates into motivated behaviour as operationalized by the time spent in the unending infinite runner.

2.4.2 Related Work

We present research from three areas of interest: identification with media characters, identification with avatars in games, and how intrinsic motivation is fostered in games.

2.4.2.1 Identification with Media Characters

Identification is the degree to which individuals like a character, empathize with a character, or perceive a character as being similar to themselves (Cohen, 2001; Ryan et al., 2008). Identifying with a character has been shown to increase media enjoyment (Trepte & Reinecke, 2010), has increased the persuasiveness of messages (Moyer-Gusé & Nabi, 2010; Moyer-Gusé et al., 2011), and has improved health outcomes (Kim & Sundar, 2012). Identification has also been shown to increase aggression (Konijn et al., 2007), addiction (Smahel et al., 2008), and depression (McDonald & Kim, 2001; Bessière et al., 2007). Although we can make assumptions about how identification with a media character from film or television will translate into a game environment, there is one big difference: in games, the player often controls the actions of a character—exhibiting agency in the digital environment (Fullerton, 2014)—which is not possible with traditional film or television characters (Fullerton, 2014). Klimmt et al. present arguments and evidence to describe avatar identification in video games as a shift of self-perception (Klimmt et al., 2010; Christoph et al., 2009) that allows a player to temporarily become one with a character. Murphy (2004) argues that games expand television as a space for identification by allowing people to interact with and through characters. In traditional linear narratives like plays and stories, the action of the protagonist defines their characteristics and personality; however, games have the ability to go beyond mere observation by allowing players to create and dynamically evolve their characters (Neustaedter & Fedorovskaya, 2009). While traditional narratives rely on empathy (Tomlinson, 2005)—the viewer's identification with the emotions of a character—interactive computer games emphasize agency—the direct control of a player over the behavior and development of their character (Fullerton, 2014). Balancing empathy and agency in the design of game characters is of essence to create a compelling experience (Fullerton, 2014), which is positively perceived by the audience (Mallon & Webb, 2005).

2.4.2.2 Identification with Avatars in Games

Players of digital games spend extensive hours playing a game, often with a single avatar (Livingston et al., 2014). Avatars allow players to project their identity into a virtual environment (Ducheneaut et al., 2009), are a means to explore our own identify (Van Reijmersdal et al., 2013), or to play with different forms of identify (e.g., gender swapping; see Hussain & Griffiths, 2008), and are a means to form social relationships in games (Livingston et al., 2014). It is not surprising that players identify strongly with their digital representation; an avatar is a digital artifact that is valued both personally (Livingston et al., 2014) and financially (Castronova, 2004). Prior research has shown that identifying with an avatar has positive outcomes for play experience and enjoyment (Trepte & Reinecke, 2010), shapes our behaviour outside of the game (Yee & Bailenson, 2007), and makes us more susceptible to persuasive messages (Murphy, 2004).

Player-Avatar Convergence Avatars need not just represent us as we are, but also can represent us as we wish to be—otherwise known as wishful identification (Hoffner, 1996; Hoffner & Buchanan, 2005). Convergence between characteristics of an avatar and of the ideal characteristics of a player (their ideal-self; see Przybylski et al., 2012) have been shown to create higher levels of immersion and are predictive for intrinsic motivation and positive affect after game play (Przybylski et al., 2012). However, avatars that represent who we are (actual-self; see (Przybylski et al., 2012)) have value: avatar-player similarity has been shown to be positively related to identification (Trepte & Reinecke, 2010). However, researchers have also showed that players sometimes sacrifice avatar-player convergence (harming identification) for strategic reasons in competitive play; e.g., by choosing avatar attributes that are dissimilar to themselves, but that help their chances of winning (Vasalou & Joinson, 2009).

The idea behind player-avatar convergence can also be extended to the similarity between the avatar and the player's idea of who they are during play, i.e., their game-self (Abuhamdeh et al., 2014). The discrepancy between who we are and who we wish to be is predictive of negative health outcomes, e.g., depression and anxiety (Gonnerman et al., 2000), and these results may also extend into our game-selves—players with lower avatar-self convergence also experience lower levels of well-being.

Facilitating Identification Because of the established benefits of identifying with an avatar, various methods for facilitating this relationship have been suggested. Trepte and Reinecke (2010) showed that customizing an avatar increases identification and leads to higher enjoyment, and also that creating an avatar that has high similarity to the player increases identification, and ultimately game enjoyment. Schneider et al.

(2004) showed that participants who played a first-person shooter game that included a narrative experienced more identification with their avatar than those who played a non-narrative version. Crenshaw and Nardi (2014) showed that when creating an avatar, the given name is important to increase identification.

However, the choices made in avatar creation can be context dependent: whereas identification facilitates enjoyment, players may optimize their characters for a particular game context to maximize their chances of succeeding and to comply with the underlying game mechanism (Trepte & Reinecke, 2010; Livingston et al., 2014). Vasalou and Joinson (2009), for example, show that an avatar's appearance is adapted to the content of a game, e.g., creating an attractive avatar in a dating game. Literature shows that there are multiple ways to facilitate identification; our goal is to use identification to foster intrinsic motivation.

2.4.2.3 Intrinsic Motivation in Games

Self-determination theory (Ryan et al., 2006) is a well-grounded theoretical framework that allows us to explain the intrinsic motivation that people have to play games due to having their basic psychological needs satisfied through game interaction. The traditional model proposes three needs (Deci & Ryan, 2000): **Competence** is the need to experience mastery and control over the outcome of a challenge, e.g., having a clear objective in a game; **Autonomy** is defined as the need to engage in a challenge under one's own volition, e.g., selecting challenges that are in-line with our personal perception of challenge; **Relatedness** is the universal need to feel connected to others, e.g., playing a game with friends. The model has been extended to capture the unique characteristics of digital games with **Presence/Immersion**, the experience of being transported into a virtual environment; and **Intuitive Control**, which describes the naturalness of the game input.

Game designers (VandenBerghe, 2014) have adopted Self-Determination Theory as a model to describe how games facilitate need satisfaction. First, by providing challenging tasks that allow us to experience a sense of mastery; second, by creating autonomy by allowing players to customize their avatars, choose different pathways through a level, or by allowing them to choose the group to play with; and third, by creating social environments that provide a variety of tools to create a sense of relatedness, e.g., in-game messenger, trade systems, and challenges that can only be mastered as a group. The Player Experience of Needs Satisfaction Scale (PENS) (Ryan et al., 2006) was developed to assess the subjective satisfaction of needs in video games and has been broadly applied in research on digital games.

The Intrinsic Motivation Inventory (IMI) defines intrinsic motivation using the following four constructs that comprise intrinsic motivation: interest-enjoyment, pressure-tension, competence, and effort-importance. The need satisfaction constructs and intrinsic motivation constructs are related in the context of gameplay; research has demonstrated that the satisfaction of needs predicts intrinsic motivation (McAuley et al., 1989). These measures are all derived from players reporting on subjective scales; however, there have also been attempts to measure intrinsic motivation to play games objectively by capturing instances of motivated behaviour.

Behavioral Measures of Motivation Subjective instruments give important insights into a player's experience and allow us to make assumptions about how a player's reported state might relate to their behaviour—for example, their game choice, the time spent in game, the recommendation to others, or their financial investment. Ryan et al. (1991) showed that being intrinsically motivated increases the time spent on a puzzle task. In an online game, Yee and Bailenson (2007) demonstrated a relationship between in-game behavior and motivation—that being motivated by achievement resulted in players optimizing their controls. Hars and Shaosong Ou (2001) show that intrinsic motivation drives participation in open source projects, and Childers et al. (2001) show that the enjoyment of online shopping experience translates into shopping behaviour. In the context of games, Birk, Toker, et al. (2015) showed that enjoyment and motivation predicted aspects of in-game behaviour, including more logins and greater social interaction with other players.

In addition to establishing links between subjective reports of motivation and behaviour, researchers have also measured motivated behaviour directly. Abuhamdeh et al. (2014) offered participants a choice of continuing to play a game that they won by either a slim or wide margin, and identified this as a behavioural measure of intrinsic motivation. In developing the PENS instrument, Przybylski et al. (2010) conducted a series of studies and used a behavioural choice for continued play after the study play period (a dichotomous choice) as an indicator of a motivational outcome. Finally, the time spent on boring or repetitive tasks, e.g. knob turning (Sansone et al., 1992), has been used before to measure task motivation.

2.4.3 Experiment Design

We conducted an online study in which we facilitated avatar identification in half of our participants and then had them play a custom infinite runner game. We measured identification, play experience and motivated behaviour.

2.4.3.1 Avatar Creator

To manipulate avatar identification, we built a character creator using C# and the Unity 4.6 game engine (Unity Technologies, 2014), Unity Multipurpose Avatar 2 (UMA Steering Group, 2015), the AL Male Civilian Pack for UMA (AlienLab, 2014), UMA Dresses (Lunatic Fringe Games, 2014), UMA Hair Pack 1 (Will B, 2015), and the UMAzing In-Game Character Creator (Strafejump Studios, 2014). During the creation process, participants were asked to create an avatar and adjust the avatar's appearance, personality, and attributes (characteristics). Participants were required to spend a minimum of four minutes in the character creator, but could take longer if they wished. Once appearance, personality, and character attributes were customized, and the 4 minutes had passed, participants were shown a summary of their character and asked to enter a nickname (Crenshaw & Nardi, 2014) before moving on (see Figure 2.1).

Appearance: Participants could customize the avatar's sex (2 choices), skin tone (5), hairstyle (4), hair colour (4), eye colour (4), upper clothing (3), lower clothing (3), clothing colours, and shoes (2). While hair colour (brown, blond, black, and red), and eye colour (blue, green, brown, and green), were limited to one

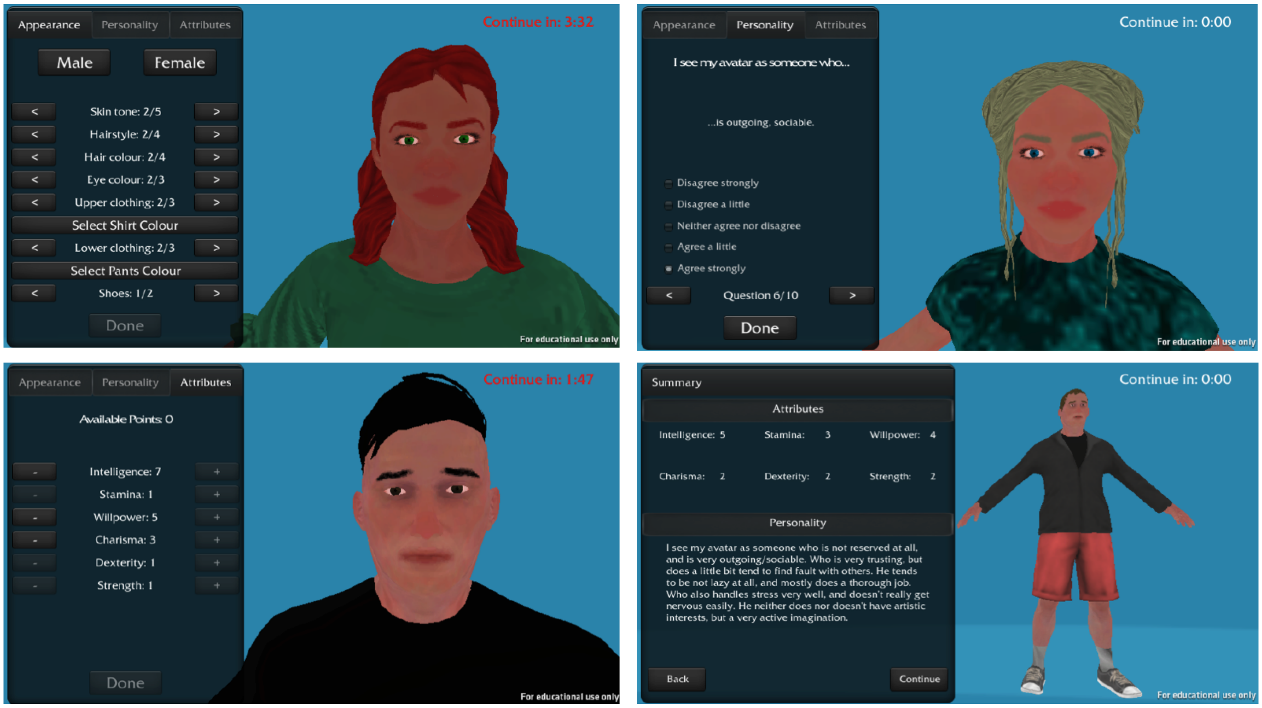

Figure 2.1: Avatar Creator displaying. Top left (a): Appearance; Top right (b): Personality; Bottom left (c): Attributes; Bottom-right (d): Summary. Similarity to authors is purely coincidental.

of naturally-appearing discrete choices, clothing colours were picked using a colour picker supporting 16.7 million colours.

Personality: Participants were asked to answer ten statements describing their character's personality. The personality characteristics matched the items from the BFI-10 personality questionnaire (Rammstedt & John, 2007) (e.g., "is reserved", "tends to find fault with others"); the instructions were changed to "I see my avatar as someone who…" Participants rated their agreement on a 5-pt Likert-scale (disagree strongly to agree strongly).

Attributes: Participants were required to assign eighteen points to their character's intelligence, stamina, willpower, charisma, dexterity, and strength, giving them the opportunity to favour certain attributes. The process required participants to assign at least one point in each category.

Summary The summary displayed the created avatar next to the personality and attribute profile. The personality responses were displayed as prose, using the responses from the BFI-10 questionnaire (Rammstedt & John, 2007). We adjusted the text, matching it to the avatar's sex, e.g., "She tends not to be lazy, and mostly does a thorough job". The purpose of the summary was to reinforce the characteristics of the created avatar, and give the player the sense that their avatar had a profile. After accepting the summary, participants named their avatar.

Figure 2.2: Timed Infinite Runner.

Manipulating Avatar Identification To manipulate identification, we allowed half of our participants to use the avatar customization tool. We told participants (who were MTurk workers) that they would be customizing an avatar to use in a series of MTurk tasks that would be released in the future, using personalized invitations to participate. They were told that in future studies we would make use of the avatar's personality and skills (for example by assigning a task that matched the avatar's abilities), but that the current task was focused on the creation process and thus the avatar would be used in a simple game unrelated to their personality and skills.

The other half of the participants were randomly assigned an avatar of the same sex, from a set of four avatars. We created 2 different personality sets: the first set (*Personality A*) showed low extraversion, high neuroticism, and a tendency for low agreeableness. Openness, and conscientiousness were kept ambiguous. The second set (*Personality B*) was high on extraversion, high on agreeableness, and a tendency to be open. Neuroticism and conscientiousness were kept medium. We also created two sets of attributes for both sexes. The first set (*Attributes A*) emphasized high intelligence (5 points), high charisma (4 points), and elevated dexterity (3 points); the other attributes received 2 points each. The second set of attributes (*Attributes B*) emphasized high strength (6 points), high stamina (5 points) and low intelligence (1 point). The other attributes received 2 points each. The personality and attribute sets were crossed to four different avatars for each sex: *Personality A, Attributes A; Personality A, Attributes B; Personality B, Attributes A; Personality B, Attributes B*. We kept all characteristics of the avatar's appearance very neutral. A mid-range skin tone and brown hair was chosen, and avatars of both sexes were dressed in a grey shirt and black pants.

Figure 2.3: Endless Infinite Runner; the timed version had the same layout except with a timer instead of the quit button.

Following the same procedure used in (Hanus & Fox, 2015) to study the persuasiveness of avatars as virtual salespeople, participants in the randomly-assigned avatar group watched a video of the creation and customization of their avatar. We created four videos for each sex with the four different personality and attribute configuration. Participants could not intervene in the creation process and only passively watched the assigning of appearance, personality, and attributes. In contrast to the customization process, participants who watched the video were not allowed to name their avatar; instead the avatar was represented as "Player 1" throughout the study.

2.4.3.2 Gaming System: Infinite Runner

Infinite Runner is an infinite runner game in which players run down a street, collecting coins, while avoiding obstacles. The game was implemented in C# using Unity 4.6 (Unity Technologies, 2014) and the 3D Infinite Runner Toolkit (Dreamdev Studios, 2014).

Gameplay: Players run down a dark street by controlling the previously-created or assigned avatar. Players have three kinds of actions to avoid obstacles: change lanes on the street, (i.e., left, middle, or right), jump, and roll. The player's score was increased by collecting coins; each coin added one point to the player's score. The game featured 3 types of obstacles: buses, signposts, and cars. Buses could not be jumped over and forced the player to change lanes. Signposts could be avoided by rolling. Cars could be avoided by jumping over them or moving around them. To increase difficulty, hitting an obstacle slowed the player down for one second and deducted 10 points.

Controls: The player had to alternate between pressing the N and M keys in order to keep the character running forward. Forcing players to constantly press buttons served two purposes: 1) it made the task more tiring, and therefore required more effort to stay invested; and 2) the constant input ensured that participants didn't simply wait until the game time expired—the timer stopped progressing if participants stopped running. Players switched lanes by pressing the A key (left), and the D key (right), and pressed the W key to jump over obstacles and the S key to roll.

Versions: Participants played four rounds of a timed version of the infinite runner (see Figure 2.2), and also one round of the endless infinite runner (see Figure 2.3). The only difference between the timed and endless versions was that the timed rounds lasted for 60 seconds, whereas the endless round allowed the participant to play until they pressed a quit button, which appeared in the top left corner (see 2.3). We set a threshold of 20 minutes of time passing (or 10 minutes of active running time) in the endless runner to ensure that participants had time within the MTurk system to complete the post-experiment questionnaires.

2.4.3.3 Measures

We collected both subjective and behavioural measures.

Identification

Identification was measured using the avatar-related sub-scales of similarity identification, embodied identification, and wishful identification from the Player Identification Scale (PIS; Van Looy et al., 2012). Participants were instructed to rate their agreement to identification-related statements, e.g., *similarity*—"My character is like me in many ways."; *embodied*—"I feel like I am inside my character when playing"; *wishful*—"I would like to be more like my character".

Player Experience

Positive and Negative Affect was measured using the Positive Affect / Negative Affect Scale (PANAS; Watson et al., 1988). Participants were instructed to indicate how they felt "right now" using a list of adjectives, e.g., "active" or "upset", and rating their agreement on a 5-pt Likert scale. PANAS has been used to evaluate games before (Greitemeyer & Osswald, 2010).

Intrinsic Motivation was measured using the Intrinsic Motivation Inventory (IMI; McAuley et al., 1989). The IMI measures the constructs *interest-enjoyment*—"I enjoyed this game very much.", *effort-importance*—"I put a lot of effort into this game.", and *tension-pressure*—"I felt tense while playing the game.". Each construct was measured using agreement to statements on a 7-pt Likert-scale. The IMI has been used in games research before (Ryan et al., 2006).

Need Satisfaction , based on Self-Determination Theory (Deci & Ryan, 2000), is a predictor of intrinsic motivation. Need Satisfaction of competence (i.e., demonstrating mastery over challenges), autonomy (i.e., doing so under one's volition), and relatedness (i.e., doing so while feeling connected to others) as experienced during play, was measured using the *Player Experience of Need Satisfaction Scale* (PENS; Ryan et al., 2006). PENS adds two additional constructs: presence and intuitive control that have been identified as being relevant in the context of games (Ryan et al., 2006). Each of the five constructs were measured using agreement to statements on a 7-pt Likert-scale.

Game Performance

Additional in-game metrics were used to measure performance, i.e., the number of obstacles hit and the number of coins that were collected indicated in-game performance.

Motivated Behaviour

Motivated behaviour was operationalized as time spent actively running in the endless round of the infinite runner; investing more effort into a game and showing higher levels of endurance during an engaging but tiring task have previously been used as an objective measure of motivation (Sansone et al., 1992). Objective motivation in the context of Infinite Runner is measured using time spent actively running in the final endless round (i.e., time spent alternating between M and N presses) before pressing the quit button.

2.4.3.4 Participants and Deployment Platform

130 participants (40% female) with an average age of 31.62 ($SD = 8.12$) participated in our study through Amazon Mechanical Turk (MTurk), a platform that acts as a broker between parties offering a range of Human Intelligence Tasks (HITs) (e.g., marketing questionnaires or research studies) and paid workers. Participants received $6 compensation paid through the platform. Although it has been shown that MTurk is a reliable research tool (Mason & Suri, 2012), we measured the time spent per questionnaire to evaluate task performance and ensure that participants were attentive despite the online setting (Hauser & Schwarz, 2016). Participants were excluded from further analysis based on the following: if participants filled in more than 2 surveys with zero variance between items, or showed ratings of ±3SD in more than 2 questionnaires. After these filters were applied, 126 participants (38.9% female) with an average age of 31.61 ($SD = 8.20$) were included in further analysis. Excluding participants based on quality criteria is a standard approach for data collected via crowdsourcing platforms; the applied criteria were suggested by (Mason & Suri, 2012). Ethical approval was obtained from the University of Saskatchewan behavioural research ethics board, and participants were asked to provide informed consent. To comply with ethical guidelines, the HIT was only made available to workers in the USA who were older than 18. Additionally, only workers with an approval rate above 90% were offered the HIT as a means of quality control.

2.4.3.5 Procedure

Participants first received information about compensation and the expectation to answer attentively and quickly, and were then asked to give informed consent. Task duration was also revealed; to avoid time as a source of pressure, the allotted time was 20 minutes longer than the average time needed in a pilot study with 18 participants. At the beginning of the experimental phase, participants filled out a demographics questionnaires asking about, e.g., age, and gender, and PANAS to measure baseline affect. Afterwards, participants were randomly assigned to one of two groups: the customization group or the random avatar group. After creating an avatar (or watching an avatar being created), participants answered the identification scale. Next, each participant was told that they were being connected to a group of online players (to increase the social relevance of the situation and stimulate investment in the experiment), and they then played two blocks with two rounds in each block of Timed Infinite Runner. Each round lasted 60 seconds. After each round, participants were presented with a manipulated leaderboard showing their progress relative to the other simulated players. The first round in each block presented neutral feedback and the second round presented neutral feedback in block one and losing feedback in block two. Leaderboards have been shown to be effective at manipulating play experience and creating controlled variance in the game experience across participants (Bowey et al., 2015).

After each block, participants filled out the PANAS, PENS, and IMI scales. After answering the questionnaires after the second block, participants played Endless Infinite Runner for a maximum of 20 minutes (10 minutes of active running). A red quit button was presented at the top left corner of the screen (see Figure 2.2) to indicate that participants could quit Endless Infinite Runner any time. After 20 minutes, the game ended regardless, and participants were forwarded to the last set of questionnaires, in which we collected the PANAS, and additional demographics questionnaires. Finally, participants stated the purpose of the experiment in their own words, after which the purpose of the experiment was disclosed. We ensured that participants understood the deception of the leaderboard manipulation, and the avatar being used across multiple experiments by asking simple comprehension questions.

Data Analyses Collected data was analyzed using SPSS 23, with the PROCESS-macro for SPSS from (Hayes, 2013). IMI and PENS data after rounds 2 and 4 of the timed running game were averaged to capture the full the experience of *Timed Infinite Runner*. To capture the change in affect starting before the experiment to after the final round of *Endless Infinite Runner*, we calculated the change in affect (positive and negative individually) from baseline measurements to the values gathered after *Endless Infinite Runner*.

To test the effects of identification (as a continuous variable, rather than a group treatment) on the dependent measures, we conducted hierarchical linear regressions. Because sex and age can affect measured need satisfaction and intrinsic motivation during play (Ryan et al., 2006), we entered these two measures into the first block of the regressions. We then report the R^2-change of adding identification as a predictor of: the five needs satisfaction measures, the three intrinsic motivation measures, the two affect measures, and the

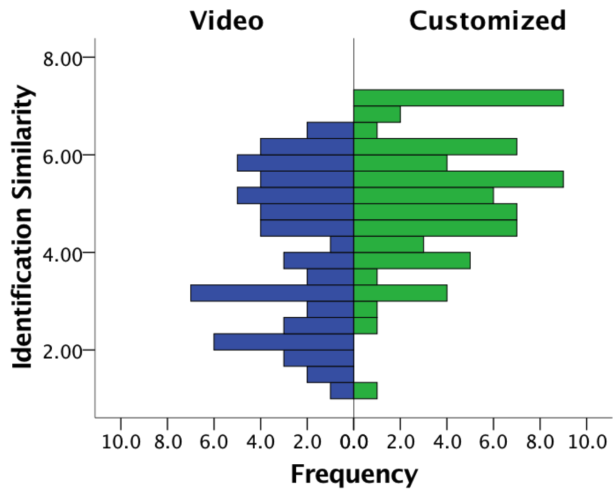

Figure 2.4: Mirrored Histograms for the Video Condition (left) and the Customized Avatar condition (right).

measure of motivated behaviour. Separate regressions were conducted for the three types of identification (similarity, embodied, and wishful). We set α at 0.05.

2.4.4 Results

To understand the effects of avatar identification on player experience, intrinsic motivation, and motivated behaviour, we formulated five research questions based on prior work. We present our results according to these five questions.

2.4.4.1 RQ1: Does avatar customization increase identification?

Our first research question was to check that our manipulation of customization was stimulating a range of identification values. First, we conducted a multivariate ANOVA on the three measures of identification, with customized avatar ($n = 58$) or randomly-assigned avatar ($n = 68$) as the between-groups factor. The slight group size differences are an artifact of random assignment; however, ANOVAs are robust to small group size differences for main effects (Field, 2013). We show that customizing the avatar increases similarity-identification ($F_{1,124} = 19.8, p < .001, \eta^2 = .14$), embodied-identification ($F_{1,124} = 9.41, p = .003, \eta^2 = .07$), and wishful-identification ($F_{1,124} = 7.03, p < .009, \eta^2 = .05$) over randomly assigning an avatar, suggesting that the customization process was effective at stimulating avatar identification, as suggested by previous work (Hanus & Fox, 2015) (see Figure 2.4). Although there are significant group differences, Figure 2.4 shows how the differences are created from overlapping distributions of identification (as is expected of group differences; see Field, 2013). As our interest is not in the results of customization, per se, but in identification (which could be facilitated using a range of methods), we use identification with the avatar as a continuous predictor of experience and motivation measures in the remaining research questions.

2.4.4.2 RQ2: Does avatar identification improve in-game needs satisfaction (i.e., competence, autonomy, relatedness, immersion, and intuitive control)?

The hierarchical regressions (described in the section on data analyses) showed that when controlling for age and gender, identification similarity significantly predicted autonomy and immersion, but not relatedness or competence (see Table 2.1). The same pattern was true for embodied identification and wishful identification. In addition, intuitive control was significantly predicted by similarity and embodied identification, but not wishful identification.

2.4.4.3 RQ3: Does avatar identification improve intrinsic motivation (i.e., enjoyment, effort, and reduced tension)?

The hierarchical regressions showed that when controlling for age and gender, identification similarity significantly predicted enjoyment and effort, but not tension (Table 2.1). The same was true for embodied and wishful identification.

2.4.4.4 RQ4: Does avatar identification increase positive affect or reduce negative affect?

The hierarchical regressions (Table 2.1) showed that when controlling for age and gender, identification similarity, embodied identification, and wishful identification all significantly predicted positive affect, but not negative affect.

2.4.4.5 RQ5: Does avatar identification translate into motivated behaviour (i.e., do players who identify with their avatar spend longer in the unending infinite runner?)

The hierarchical regressions showed that when controlling for age and gender, identification similarity, embodied identification, and wishful identification all predicted the motivated behaviour, which we operationalized as the time spent actively running in the endless infinite runner. Eight participants reached the 10-minute threshold; in these cases the experiment proceeded automatically.

2.4.4.6 Summary of Results

Our results showed that avatar customization stimulated identification with the avatar, which we then used as a predictor in a series of hierarchical regressions (controlling for age and gender). Greater identification predicted increases in autonomy, immersion, invested effort, enjoyment, and positive affect. We also show that greater identification translates into motivated behaviour as operationalized by the time spent in the endless infinite runner. Finally, we show that these relationships are true for the three constructs underlying identification: similarity to the avatar, a feeling of embodied presence within the avatar, and wishing that one could be more like the avatar.

With R^2-values between .07 and .22, the presented effects are considered as small; however, small effect sizes in games user research are not uncommon, because of the complex interaction between individual characteristics of the player and the game (Birk, Mandryk, et al., 2015; Steinemann et al., 2015; Zendle et al., 2015).

2.4.5 Discussion

Although our avatar creation process is a pared-down version of the types of character creators seen in role-playing games, such as World of Warcraft, we still show differences in identification and subsequent experience and motivated behaviour. The identification that can be created in the lab can be considered as a poor copy of the deep identification with a character that happens when players engage with games through digital representations over the long term.

We assume that the observed motivational and emotional responses would increase if we could investigate identification of players with their longtime digital representations. However, our results suggest that a process as simple as avatar creation is a promising step towards a paradigm that allows us to facilitate identification.

Being intrinsically motivated to play a game has benefits for player experience; however, there are also potential economic and societal benefits. If identification in commercial games fostered intrinsic motivation, players may bemore likely to spend more time in the game and recommend the game to others. Taken together, these aspects are fundamental to a game company growing a customer base, which is highly relevant for success. In terms of the societal benefits, intrinsic motivation to play a game has relevance for games that are used in serious contexts.

2.4.5.1 Application in Serious Contexts

Although serious game companies also wish to grow their customer base, their goals for fostering intrinsic motivation may be different. Serious games often focus on motivating behaviour change or promoting learning (Michael & Chen, 2005). For example, the persuasive health game *Escape from Diab* (Archimage, 2006) uses persuasive techniques to improve kids' eating behaviour, whereas *DragonBox* (WeWantToKnow AS, 2012) is designed to help kids learn algebra. These systems rely on repetitive use over longer terms to optimize results. (Ryan et al., 2008) argue that being self-determined—which leads to being intrinsically motivated—can improve the outcomes of health-related interventions, e.g., smoking cessation on or improved dental hygiene. Our results suggest that identification with the digital representation in a game facilitates intrinsic motivation, as evidenced by more enjoyment and invested effort. We saw this relationship in our subjective data, but also observed it in motivated behaviour—players spent more time interacting with the final phase of the game before pressing the quit button. In the context of games for learning or behaviour change, increased effort in play could translate into real change for players—a learner could better understand and retain information; a player of a smoking cessation game (Khaled et al., 2008) could stop smoking sooner;

	Similarity Identifiction					Embodied Identification					Wishful Identification				
	β	R^2	$R^2(c)$	F (c)	p (c)	β	R^2	$R^2(c)$	F (c)	p (c)	β	R^2	$R^2(c)$	F (c)	p (c)
Player Experience of Need Satisfaction (PENS)															
Competence	0.15	0.09	0.02	2.74	0.10	0.12	0.08	0.02	2.04	0.16	0.02	0.07	<0.01	0.07	0.79
Autonomy	0.18	0.07	0.03	4.13	**0.04**	0.29	0.12	0.08	11.50	**<0.01**	0.23	0.09	0.05	7.27	**0.01**
Relatedness	0.05	0.03	<0.01	0.25	0.62	0.10	0.19	0.01	1.17	0.28	0.05	0.03	<0.01	0.37	0.55
Immersion	0.20	0.07	0.04	5.25	**0.02**	0.35	0.15	0.12	17.25	**<0.01**	0.24	0.08	0.06	7.32	**0.01**
Intutive Control	0.22	0.21	0.04	6.83	**0.01**	0.23	0.22	0.05	8.46	**<0.01**	0.08	0.17	0.01	0.98	0.33
Intrinsic Motivation Inventory (IMI)															
Enjoyment	0.21	0.08	0.04	5.81	**0.02**	0.25	0.09	0.06	8.30	**0.01**	0.19	0.07	0.03	4.50	**0.04**
Effort	0.19	0.11	0.03	4.52	**0.04**	0.28	0.15	0.08	11.15	**<0.01**	0.18	0.11	0.03	4.28	**0.04**
Tension	<0.01	0.01	<0.01	<0.01	0.98	-0.03	0.09	<0.01	0.12	0.73	0.01	0.01	<0.01	0.02	0.90
Positive Affect/Negative Affect Scale (PANAS)															
Postive Affect	0.28	0.15	0.07	10.64	**<0.01**	0.30	0.17	0.09	13.32	**<0.01**	0.21	0.12	0.04	5.94	**0.02**
Negative Affect	-0.04	0.02	<0.01	0.18	0.67	0.05	0.14	<0.01	0.31	0.58	0.06	0.02	<0.01	0.47	0.50
Behaviour															
Time Played	0.15	0.15	0.06	8.07	**<0.01**	0.25	0.16	0.06	9.30	**<0.01**	0.25	0.15	0.06	8.73	**<0.01**

Table 2.1: Regression properties including β, R^2, change in R^2 from adding identification, F of change, and p of change for PENS, IMI, PANAS, and behaviour for the three identity constructs: Similarity Identification, Embodied Identification, and Wishful Identification. Variables displaying change statistics from adding identification in the second step of the model are marked with (c); significant results are displayed in bold.

a player of a digital therapy game (Dennis & O'Toole, 2014) could better combat anxiety-related disorders. Serious game designers who wish to inspire invested effort may think about fostering identification with an avatar.

Interestingly, our results suggest facilitating any of the three identification types. Depending on the situation and goal, designers could choose to foster similarity, embodied identification, or wishful identification. For example, in a training game for cashiers who are learning to operate the registers at a large supermarket, fostering similarity might make more sense than nurturing embodied identification; whereas a persuasive game designed to help players reach a goal might use wishful identification. Regardless of how identification is cultivated, applying our results to the context of serious games suggests that increasing identification in a serious game may stimulate real change.

Motivating Crowdsourced Work Facilitating intrinsic motivation through avatar identification is applicable beyond the context of games. Consider for example if we re-interpret our results in the context of us conducting an experiment on a crowdsourcing platform. We showed that workers invested more effort and experienced more enjoyment in their task when they identified with their avatar. The task itself was a game they played, but this was done as an assignment in a work context. Expanding on these results, it may be possible to provide workers with a stable and representative character that they identify with to increase motivation, productivity, and task enjoyment. Currently, the closest point of identification is the worker profile that stores information about previous task performance; MTurk workers respond very promptly to anything that threatens their profile, because a track record of good performance gives access to better tasks.

2.4.5.2 Blurring the Player-Avatar Relationship

Our work shows that similarity with the avatar yields benefits to motivation; however, the relationship between player and character can vary. Players can be very attached to their representation in a game or they can see the character simply as a means to achieve in-game goals (Livingston et al., 2014). The motivation to play likely varies accordingly: a player who feels attached to their character and experiences guilt if the character is neglected has a different motivation compared to a player who uses an avatar simply to get access to game content. The former player may provide their character with a backstory, enjoy inhabiting their character during play, and make choices in appearance and behaviour that are in line with the character's backstory, whereas the latter may simply make choices to beat the system. Our results suggest that identifying with an avatar could result in greater time investment, and also help players enjoy the game more. Even in a game like World of Warcraft—where the overall narrative and character development process facilitates identification—the identification of players with their avatars varies (Livingston et al., 2014). A range of identification in one game is of interest, and so is the variation of identification between games.

Identification with Minimal Representation Compare, for example, *World of Warcraft* with a Multiplayer Online Battle Arena (MOBA) Game like *League of Legends* (LoL, Riot Games, 2009) or *Defense of the*

Ancients 2 (Dota 2, Valve, 2013); several differences in designing for identification emerge. WoW provides a complex system and environment that allows players to have distinct experiences for an individual character. LoL does not allow players to develop one distinct character—the game allows players, called summoners, to select one of over 125 champions to fight in a 5 on 5 team battle. Each champion has different advantages, disadvantages and synergies with other champions. While champions can be customized with items during a match, the customization is temporary and is reset after each game. Thus, unlike in WoW, the LoL summoner account does not have an affiliated avatar with a unique appearance and history. Customization can only be achieved by giving the account a name, through purchases of champions, different skins for champions, and of course through the performance of the summoner. Identity in MOBAs is a result of the player's performance—the ratio of won and lost games, the ratio of kills and deaths in the game, and other statistics define how a player is perceived and distinguish novices from professionals. While in both the cases of WoW and LoL, the key to economic success is identification, how this goal is achieved varies. WoW uses a digital representation as a proxy that allows players to indulge in the fantasy of the game, while MOBAs leverage a player's personal abilities to facilitate engagement with the game. How designers can foster identification with a profile that represents the player's own skill is an interesting problem that can be addressed through future work.

2.4.5.3 Identifying with Ideals or Groups

Our work focuses on identifying with an avatar; however, characters in games can represent more than their individual nature, by representing an ideal (e.g., the nameless hero who saves the world through their courage) or a segment of society (e.g., prostitutes in a car theft game) (Fullerton, 2014). For example, Abe (*Oddworld: Abe's Oddysee*, GT Interactive, 1997), is a slave who discovers that his fellows are slaughtered and processed as meat. He stands up against the authorities and liberates as many of his fellows as possible. The game allows the player to identify with Abe as a hero, but also with Abe as a slave, a person who is exploited for the good of others and who tries to overcome a hopeless situation. Fostering identification with an ideal—rather than with an avatar—may have similar benefits for motivation.

This idea of identification with a segment of society speaks to the potential of using membership within a group to facilitate identification. For example, in *World of Warcraft*, players choose either the side of the horde or the alliance; this choice gives access to different races with different histories, who are perceived differently in the world. Or consider the RTS game *StarCraft II* (Blizzard-Activision, 2010), in which players pick one of three races: Terran, exiled humans, known for their adaptability; the Zerg, a race that has advantages through mere numbers; or the Protoss, the most technologically advanced race. Each race has individual strengths and weaknesses. Player can identify with the race that they play often and are most skilled with.

Other games use a more dynamic approach to create a sense of group membership. *Fallout 3* (Bethesda, 2008), for example, uses a "Karma" system that keeps track of a player's good and wrong doings in the game world. “Good” and “Bad” players have similar advantages in the game, but other aspects of play can also change—for example, players in bad standing may purchase different weapons from a slaver than those in good standing ,who purchase from a normal merchant. Group membership could be less reflective of who a player is and more related to who they wish to be (ideal-self). For example, would a player like to be a splendid example of society, or maybe associate more with seedier elements of society and seek advantages through toeing moral lines? Allowing players to explore aspects of identity and allowing them to find the sweet spot of their moral integrity is an advantage of game worlds.

2.4.5.4 Limitations and Future Work

Our work shows that avatar identification facilitates intrinsic motivation and motivated behaviour. However, there are limitations that can be addressed through future work.

We facilitated avatar identification through a customization process. We offered the players ways of customizing the appearance of an avatar, but they did not have full control over the avatar’s appearance. They could change the hair and clothes, but not, for example, the face or body. Allowing players to represent themselves completely could have fostered greater identification similarity. Future work should investigate how the range of options for avatar creation changes identification and resulting motivation in play.

Our game was played in the context of MTurk workers engaged in an experiment. Although our results showed differences in this context, investigating avatar identification in the context of volitional play would provide ecological validity. Because there is a fundamental mismatch between randomly assigning participants to experimental conditions, and simultaneously having them engage in play under their own volition, we plan to take our results from these controlled experiments and apply them to see whether they extend to games played under the user’s volition.

We investigated identification with an avatar. Several successful games (e.g., Tetris) do not have an avatar with whom to identify. We have started investigating how to reap the motivational benefits of fostering identification in games that do not have traditional avatars.

Fostering intrinsic motivation through avatar identification has implications for the design of games for serious purposes. We suggest that avatar identification can be used to facilitate invested effort and enjoyment—in fact, our results show that not only were players feeling more invested (subjectively) but were also exhibiting more motivated behaviour when they identified more with their avatar. We plan to investigate whether our results extend into serious applications—for example, by helping players invest in the process of learning a language, playing a therapeutic game, or changing behaviour through gameplay.

2.4.6 Conclusion

The goal of our work was to determine whether identification with an avatar could increase intrinsic motivation in a game. We first show that avatar customization stimulates identification with the avatar. In terms of game experience, we used hierarchical regressions with the three types of identification (similar, embodied, wishful) as individual continuous predictors, and showed that greater identification increased experienced autonomy, immersion, invested effort, enjoyment, and positive affect. Interesting for the design of engaging game experiences, we discuss how we can leverage the ways in which commercial games facilitate identification. In addition to affecting their subjective experience, we also show that greater identification translates into motivated behaviour as operationalized by the time spent in an unending infinite runner at the end of the experiment. Our results show that greater investment in a task through an objective measure of behaviour has implications for the design of games for serious purposes, whether that be to persuade, teach, or help players change their behaviour.

Finally, we show that these relationships are true for each of the three types of identification, which gives designers a range of options for facilitating identification—by increasing feelings of similarity, improving the sense of embodiment within the avatar, or cultivating wishful identification with an avatar who represents the player's aspirations.

2.5 Summary of Manuscript A

The results showed that avatar customization stimulated identification with the avatar. In terms of game experience, hierarchical regressions with three types of identification (similar, embodied, wishful) were used. The three identification types were entered as continuous predictors to show that greater identification increased autonomy, immersion, invested effort, enjoyment, and positive affect. The findings show that greater identification translates into motivated behaviour as operationalized by the time spent in the unending infinite runner.

By evaluating the effect of a single design strategy on player engagement, my research investigated a game-based design strategy and showed experiential and behavioural evidence that identification with a representation is associated with increased play experience and behavioural engagement in the short-term. As such, my research suggests that avatar customization might be a suitable motivational design strategy to foster effective implementations of interactive systems that promote sustained engagement.

2.5.1 Lessons Learned

There were several lessons I learned during the execution of this first project. Specifically, the experience of working with MTurk workers on a complex open-ended study, learning about worker compliance, and exploring behaviour in an online study. I also learned about the advantages and drawbacks of avatar customization as an experimental manipulation.

A very positive experience I had during this study was the feedback I got from MTurk workers, and equally important, seeing evidence in my data that workers stayed engaged in open-ended and attentionally-demanding tasks. The positive feedback and showing engagement under their own volition qualified MTurk workers as a group to recruit from for the remaining experiments.

Further, by controlling the response behaviour in questionnaires and overall reliability of my data through the approaches described in Manuscript A, I found that measures of data quality, such as Cronbach's alpha, time spent per questionnaire, and overall variance, suggested that my data was of good quality—a finding that is in-line with previous research on MTurk data quality (Buhrmester et al., 2011; Casler et al., 2013; Kees et al., 2017).

A surprising finding that was not discussed in Manuscript A was that the group manipulation did not show any significant effects—identification was the mechanism that drove change in experience and behaviour, while avatar customization directly did not have any group effects. The lack of a group effect was likely a result of exposure to the customization video, which was relatively long at 4 minutes in duration. The distribution in both groups (customized, generic) suggested that people in both groups identified with their representation (see 2.4). To account for identification effects from mere exposure to the creation process, control groups in later studies were only exposed to images of their avatar and not to the entire creation process.

2.6 Relevance in Context

When evaluating the relevance of Manuscript A in the context of my dissertation, several questions remain. First, it is an open question if avatar customization is a valid design strategy to foster volitional engagement with technology-enabled services for health. Second, I did not gain any insights on the effects of avatar customization in the specific context of technology-enabled services for health in terms of whether avatar customization can promote sustained engagement or improved treatment efficacy. Therefore, a more detailed investigation of the effects of avatar customization on motivation was necessary (see Chapter 3). Third, an investigation of avatar customization in the long-term and in a non-game context (see Chapter 4) is missing. And fourth, whether or not avatar customization has direct effects on task efficacy (see Chapter 5) remains unclear.

The presented manuscript is fundamental for my research and the work presented in Chapter 2 to Chapter 5. The findings support that identification with a representation effectively increases volitional task engagement and therefore justifies further investigations of the motivational and behavioural properties of avatar customization and the benefits for technology-enabled services for health.

Chapter 3

The Motivational Push of Games

Citation: Birk, M. V., Mandryk, R. L., & Atkins, C. (2016, October). The Motivational Push of Games: The Interplay of Intrinsic Motivation and External Rewards in Games for Training. *In Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play* (pp. 291-303). ACM.

Acknowledgment: Under my direction, Jason Bowey and Cheralyn Atkins were responsible for the development of the go-no-go task. Cheralyn Atkins also modifed the avatar creator described in Manuscript B, implemented the data logging for the go-no-go task, and provided a technical description of the avatar creator and the go-no-go task used in Manuscript B.

In coordination with my supervisor, Dr. Regan Mandryk, I was responsible for the research direction, experimental design, system design, data gathering, data analysis, and reporting.

My goal in Manuscript B was to explore the effects of rewards on motivation, specifically how intrinsic motivation—induced through avatar customization—and behaviour are affected. I conducted an 11-day study evolving around a game to train executive functioning. 200 participants were recruited via Amazon Mechanical Turk; half of all participants were randomly assigned to a group that customized their avatar and the other half watched a video of someone customizing an avatar. Similar to the manipulation presented in Manuscript A, I used avatar customization to manipulate identification with a representation and subsequently self-reported variables of motivation and login behaviour.

3.1 Problem and Motivation

Chapter 2 established that identification with an avatar increased self-reported motivation and motivated behaviour, operationalized as time-played in an infinite runner. While the results support that avatar customization has the potential to effectively address waning motivation and subsequently attrition of technology-enabled services for health, several open questions remained.

First, because in Manuscript A I only showed effects on self-reported motivation and motivated behaviour after 10 minutes of play, it remained unclear whether or not identification with an avatar affects motivation in the medium term (i.e., over the duration of days or weeks). Second, to assure that identification affects motivation directly rather than as a latent variable, I needed to test if induced motivational effects are responsive to known manipulations of motivation (i.e., the overjustification effect; Deci et al., 2001; Huillery

& Seban, 2014; Amabile, 1993; Hitt et al., 1992; Seaverson et al., 2009). Third, the avatar customization technique used in Manuscript A did not show group differences—an effect that might have resulted from a lack of options during the avatar creation process—as such, I wanted to test if increased options, and subsequently increased identification would result in group differences of motivation between participants who customized their avatar and those who only watched a video of the creation process. Fourth, in Manuscript A I investigated self-reported motivation and behaviour in the context of a game. While the gained insights were important for showing that avatar customization is an effective game-based motivational design strategy, I lack insights into the motivational effects of avatar customization in the context of a training application.

Training activities, such as those offered by technology-enabled services, require engagement over time to be effective. Motivational effects of avatar customization have been established in the short term (see Manuscript A), but the effects of external rewards (i.e., a financial incentive) on motivation is unknown. Therefore, the problem addressed in this manuscript is that the stability of effects of identification resulting from avatar customization are only understood in the short term and it is unknown whether external rewards thwart or diminish the effect.

3.2 Solution and Steps to Solution

To address the lack of research on psychometric properties of avatar identification in the medium-term, I investigated experience and login behaviour in an 11-day long study before and after providing a financial reward on day 8—external rewards have been shown before to thwart intrinsic motivation (Deci et al., 2001; Huillery & Seban, 2014; Amabile, 1993; Hitt et al., 1992; Seaverson et al., 2009). Similar to the procedure applied in Manuscript A, participants were recruited via Amazon Mechanical Turk and half of the participants were instructed to customize their avatar; the other half were assigned a neutral avatar followed by a staged avatar customization video. In comparison to the work presented in Manuscript A, customizable facial features such as face shape, eye distance, nose size, ear size and ear orientation, were added to the avatar creator. Participants played a daily go/no-go task—a task that requires participants to respond to one type of stimulus but inhibit a response to a visually-similar lure.

3.2.1 Research Questions

To explore the effect of rewards on participants with differing levels of intrinsic motivation, I asked three central research questions in the Manuscript:

RQ1: Does motivation wane over time?

RQ2: Is waning motivation affected by initial engagement?

RQ3: How does the delivery of an anticipated reward affect motivation for differentially motivated participants?

3.3 Design Rationale

To be in a position to answer the above research questions, I made several decisions when designing the experiment that were built on experiences made during the study described in Manuscript A. I designed the study around the overjustification effect, which captures the negative effects of rewards on intrinsically motivated behaviour (Deci et al., 2001; Huillery & Seban, 2014; Amabile, 1993; Hitt et al., 1992; Seaverson et al., 2009)—people enjoying an activity for its own sake will experience lower engagement and show less activity when they are rewarded for their engagement.

I delayed promised payment by 7-days with the intent to gather self-reported experience data and login behavior before and after receiving a reward. I decided to investigate the effects of motivational changes induced through avatar customization over the long-term instead of in single session, because long-term effects are more relevant when addressing effects stipulating the efficacy of technology-enabled services for health, which are usually used over time as well.

In Manuscript A, I already gained insights into behavioural effects of customization in the short-term. Many technology-enabled services require repeated and sustained engagement, which was not reflected in the short-term study presented in Manuscript A. To increase external validity in the present study, I was interested in increasing the ecologically validity of the study by investigating motivation and login behaviour in the context of a believable training intervention over multiple days.

I chose a game-based training, because it allowed me to incrementally adapt the insights I gained in Manuscript A. The customizable or assigned avatar were acting agents in the game, similar to the infinite runner described in Manuscript A. Therefore, I was able to build on insights gained around measuring user input, comparing performance, and investigating the effect of avatar customization on experience and behaviour. I chose a go/no-go task, because the task has shown good training effects in previous research (Dowsett & Livesey, 2000; Thorell et al., 2009), people benefit from multiple training units (Klingberg et al., 2002), the general task specification allowed for a game-based implementation, and the task has been used successfully in the context of game-based training before (Waltz, 2018).

Building on the results presented in Manuscript A, I used identification with an avatar as a proxy for motivation—those who identify less are less motivated and those who identify with their avatar are more motivated. My research design aimed to investigate the effect of participants' identification on waning motivation and the differential effect of an anticipated reward on people differentiated by identification.

3.4 Manuscript B

Games for training aim to keep interest in training activities high by making them more enjoyable, yet interest and motivation often wane over time. Games frequently employ rewards to halt waning motivation; however, research suggests that although this approach may work for less motivated players, it may backfire for players who are already enjoying a game. To explore changing motivation patterns over time, we conducted an 11-day study of a game for training executive functioning with players who were split into two groups that reflected their intrinsic motivation induced through a manipulation of identification with an in-game avatar. Although motivation waned over time, both effort and enjoyment waned more rapidly for players who identified less. After one week, when we delivered a reward (payment), the less-identified group respond posi-tively—increasing their effort and improving performance; however, the more identified group responded negatively—decreasing their effort and declining in task performance.

3.4.1 Introduction

There are many domains in which people need to invest time and effort in a training activity to see future benefit. For example, consider a person who wants to learn a second language—if she puts in the work of memorizing vocabulary, over time she will be able to communicate in another language. Or consider a person who wishes to lose weight—if he makes good eating choices every day, over time he will lose excess weight. Motivation to engage in these types of training activities often starts off quite high, with the person looking forward to the future payoff that will eventually result from their invested efforts. However, over time, motivation often wanes, resulting in people investing less effort in training, or quitting before reaching their goal.

This lost interest in training is partly because the beneficial outcome of training behaviours (e.g., being bilingual, reaching a healthy weight) takes time to achieve and is distinct from the behaviours themselves (e.g., practicing a second language, making healthy eating choices). This form of motivation—engaging in an activity because it leads to a desirable but separable outcome (Ryan & Deci, 2000a)—is called extrinsic motivation. Although extrinsic motivation can encourage participation in a training activity (e.g., Nouchi et al., 2013; Dennis & O'Toole, 2014; Field, 2013), it often wanes when the realization of the outcome is delayed—it takes time to become proficient at another language or lose weight. A remedy for waning extrinsic motivation over time is to make the training activity engaging enough so that people participate because they enjoy the training activity itself, and not just for the beneficial outcome that will result from sustained training. This form of motivation—engaging in an activity because it is inherently interesting—is referred to as intrinsic motivation (Ryan & Deci, 2000a). However, the question is how to make repetitive training activities—such as verb conjugation practice—inherently enjoyable.

It has been proposed that digital games are intrinsically motivating to play (Ryan et al., 2006) because they are inherently enjoyable. As such, the use of game design elements in non-game contexts—called gamification (Deterding et al., 2011)—has been proposed to increase the intrinsic motivation of engaging in a training activity and avoid the waning motivation that plagues extrinsically-motivated training. However, even in games that are played solely for entertainment and are inherently enjoyable, player motivation can wane over time for a variety of reasons (e.g., loss of novelty or changes in a network of friends who play), resulting in declining participation. To retain players, many games inject rewards, tokens, or valuable in-game content. These incentives sometimes aim to increase intrinsic motivation (e.g., new game missions that are inherently enjoyable); however, a more common approach is to target extrinsic motivation (e.g., by providing a free power-up or in-game currency). Although designers intend that these doses of extrinsic rewards act as motivators to encourage players, it is unclear if they are effective at halting waning player motivation and if this approach will work in the context of gamified training activities.

Research in the psychology of motivation suggests that the effectiveness of introducing external rewards into a training activity could depend on the pre-existing motivation of the participants (Vallerand & Losier, 1999)—those motivated by a separable outcome (extrinsic) should respond well, whereas those motivated by the inherent enjoyment of the activity itself (intrinsic) may respond poorly (Deci et al., 1999). This loss of motivation that occurs when external incentives are provided to intrinsically-motivated people has been demonstrated in other domains (Amabile, 1993; Deci et al., 2001; Huillery & Seban, 2014); however, it is unclear whether the use of external rewards will improve motivation, halt waning motivation, or undermine the pre-existing intrinsic motivation of players of a training game. Loss of motivation in training games has major implications as they are used in many serious domains, such as for health (Dennis & O'Toole, 2014) and safety training (Padgett et al., 2006).

As such, we conducted an 11-day study to understand the effects of time and incentives on motivation and behaviour in a game-based training activity with 200 participants. We asked all participants to engage daily in a game-based version of an established task that has been used in various digital interventions (Dowsett & Livesey, 2000; Thorell et al., 2009)—a go/no-go task to train executive function (Nevin, 1969). Although we promised payment, we withheld it for the first week—replicating the context in which the beneficial outcome for an activity is separable from task engagement and is delayed. After seven days, we paid participants, delivering their expected reward. In addition, we facilitated intrinsic motivation in half of our participants using an existing induction approach of avatar identification (Birk, Atkins, et al., 2016) to create two groups who differed in their level of intrinsic motivation. Each day, we measured subjective motivation, affective state, and task behaviour—both before and after delivering the extrinsic motivator of payment.

The results of our study make three important contributions to understanding motivation and behaviour on repeated days of participation in a game-based training. First, we demonstrated that we can use a digital induction method (avatar identification) to foster intrinsic motivation, which lasted over repeated days of participation in the training activity. Second, we showed that motivation waned over time; however, both effort and enjoyment waned more rapidly for players who identified less with their in-game avatar. Third, when we delivered a reward (payment) after one week, the less-motivated group responded positively, increasing their effort and showing improvements in task performance. However, the more motivated group responded negatively to the reward—they showed a decrease in their invested effort and declines in performance on the training task. Explained by theories of human motivation, our findings have implications for games for training, games user researcher, and games for entertainment.

3.4.2 Related Work

We first describe human motivation, followed by motivation in games for training and gamification.

3.4.2.1 Motivation

Human motivation describes a person's inspiration to act—as Ryan and Deci (2000a) note: "to be motivated means to be moved to do something". Nearly everyone who interacts with the world around them experiences motivation, thus it is a topic of interest in a diverse range of fields.

Spectrum of Motivation Motivation varies by level, i.e., how much motivation is experienced, but also by orientation, i.e. what form of motivation is experienced (Ryan & Deci, 2000a). The spectrum of different orientations of motivation is defined by how controlling or volitional an activity is (the regulatory style). Self-Determination Theory (SDT) suggests three regulatory styles of motivation: *internal motivation, external motivation, and amotivation* (Ryan & Deci, 2000a).

Amotivation suggests the complete lack of an intention to act (Ryan & Deci, 2000a), and results from not seeing an activity as valuable (Frederick & Ryan, 1995), feeling inept to do it (Deci, 1975), or not feeling that it will result in a desired outcome (Seligman, 1975).

On the other end of the spectrum, *intrinsic motivation* is defined as engaging in a task for its inherent satisfaction (Ryan & Deci, 2000a)—an activity that is its own means to an end, e.g., playing endless hours of Tetris, not to beat the high score, but because implementing a clever strategy to clear the falling rows is enjoyable in and of itself.

Extrinsic motivation is defined by an activity in which behaviour and outcome are separable, e.g., studying (behaviour) to pass an exam (outcome) (Ryan & Deci, 2000a). Extrinsic motivation varies depending on the congruency of an activity with our goals; the separable outcome may be desired because it leads to a reward (*external regulation*), leads to approval from the self or others (*introjection*), is something that is consciously valued by the individual (*identification*), or is congruent with an individual's self-view (*integration*) (Ryan & Deci, 2000a).

It is important not to confuse extrinsic motivation with external rewards. External rewards, e.g., points, or in-game currency, are representative of external regulation. Rewards are a fundamental concept in games and when combined with appropriate reward schedules, can enforce long-term retention (Deterding et al., 2011). However, once the reward or the expectation to be rewarded is removed, players would likely lose interest and retention would drop (Haw, 2008).

Undermining Intrinsic Motivation There is literature from the psychology of motivation that suggests existing intrinsic motivation can be undermined by the application of an extrinsic reward (Ryan & Deci, 2000a). As such, the efficacy of applying rewards in a training activity could depend on the existing motivation of the participants. Specifically, individuals who are extrinsically motivated to do the task should respond well to a reward, whereas intrinsically-motivated people may respond poorly.

This negative effect of an extrinsic reward on existing intrinsic motivation has been shown in multiple domains (e.g., education, see Deci et al., 2001; philanthropic activities, see (Huillery & Seban, 2014); workplaces, see Amabile, 1993, games, see Hitt et al., 1992; health, see Seaverson et al., 2009), and with a variety of demographics (e.g., kids, see Lin et al., 2015; college students, see Lin et al., 2003). However, there has been controversy surrounding the existence of this so-called "overjustification effect". In a meta-analysis of 128 studies, including four meta-analytical studies, Deci et al. (1999) confirmed that extrinsic rewards—independent of whether they are received or expected—undermine intrinsic motivation. The results showed that receiving expected tangible rewards (e.g., money or marshmallows) consistently undermined intrinsic motivation, as measured by a free choice paradigm (i.e., measuring how long participants spend on a task without any expected reward). Self-reported measures parallel results from the free-choice tasks (with weaker effect sizes), except when the rewards were based on performance (e.g., receiving a reward when a threshold score is reached).

3.4.2.2 Motivating Training Activities using Games

Motivation is a central concept in the discussion of games for both entertaining and serious purposes. With the exception of professional gamers who get financial rewards (external regulation) for playing, games are generally a leisure activity that is engaged in by choice under a player's own volition (Salen & Zimmerman, 2004). Thus players need to be motivated to participate. That motivation may be intrinsic—that is, they participate because they truly enjoy the game play—or extrinsic; they play because they get social value from gaming (introjection), value from gaming as an activity (identification), or self-identify as a gamer (integration) (Ryan & Deci, 2000a).

It is important to distinguish the motivation that people have to play games from the motivational elements that designers employ within games to encourage certain in-game behaviours. Although based in the same theories of what motivates people, design decisions grounded in motivation within games (e.g., rewarding players for in-game purchases) are conceptually distinct from design decisions used to motivate participation with the game itself (e.g., giving an in-game reward every day that the player logs in). In this paper, we focus on the latter—using game elements to motivate participation in the game in general.

Motivation in Games for Training Motivation to participate in games for training is similar to the motivation to participate in games for leisure; however, the training game can be viewed as a means to an end. Consider the example of a person who wishes to learn German and plays a game to help her learn the genders of nouns. She may actually enjoy playing the game (intrinsic motivation), may do so because her company is paying her to learn German (external regulation), because her friends all speak German and she doesn't want to be excluded (introjection), because she sees value in speaking another language (identification), or because bilingualism is congruent with her self view (integration). Although a training activity can be inherently interesting (especially at first), they generally have instrumental value in terms of leading to a separable outcome, as opposed to intrinsic value. The application of games as training activities tries to address this intrinsic value by making them more enjoyable (e.g., Dennis & O'Toole, 2014; Khaled et al., 2007). When playing games for training, players might be motivated for multiple reasons—the motivation to play might be in service of a separable outcome, e.g., learning a language as the main purpose of engaging in the game, or because the game is inherently enjoyable, e.g., because the premise is interesting or the mechanics are enjoyable (Deci et al., 2001). In our example, playing a learning game will likely be a more enjoyable means to learning the genders of German nouns than rehearsing them without the incorporation of game elements. As such, games for training tend not just to affect the spectrum of extrinsic motivation for learning, but address the underlying enjoyment of the activity itself.

The Motivational Pull of Games Games are fun, because they allow us to actively participate in a compelling narrative, provide us with challenging encounters, and give us the opportunity to choose our fate (Rigby & Ryan, 2011). A variety of models (Ryan et al., 2006; Yee, 2006) have tried to capture the essence of player motivation. Self-determination theory (Deci & Ryan, 2000) is a well-grounded theoretical framework that allows us to explain how satisfying basic psychological needs leads to enjoyment in games. The traditional model proposes (Deci & Ryan, 2000)three factors. *Competence* is the need to experience mastery and control over the outcome of a challenge, e.g., mastering the skills of a champion in *League of Legends* (2009, Riot Games), or facing the increasing challenge of *Tetris* (Pajitnov, 1984). *Autonomy* is the need to engage in a challenge under one's own volition, e.g., the diamond shaped pathways in *The Legend of Zelda* (1986, Nintendo), or through race, class, and faction choices in MMORPGs, such as *World of Warcraft* (2004, Blizzard). *Relatedness* is the universal need to feel connected to others, e.g., by playing team-matches in *Counter Strike* (2000, Valve), or by feeling connected to in-game objects or characters, e.g., bonding with the companion cube in *Portal* (2007, Valve). SDT has also been extended to capture the unique characteristics of digital games with Presence/Immersion, the experience of being transported into a virtual environment; and Intuitive Control, which describes the naturalness of the game input. Satisfying these needs has been shown to increase enjoyment (Rigby & Ryan, 2011) and play-time (Birk, Atkins, et al., 2016) in games.

Gamification Because of this motivational pull of games, the use of game design elements in non-game contexts—called gamification (Deterding et al., 2011)—has been proposed as a method to increase the intrinsic motivation of serious games. For example, games have been used to encourage serious behaviours, such as healthy eating (Baranowski et al., 2003), smoking cessation (Khaled et al., 2007), lowered energy consumption (Khaled et al., 2007), and understanding the challenges faced by people with disabilities (Gerling et al., 2014). In terms of promoting training activities, gamification is used to motivate people in the short term in domains from human resources system training (Reeves & Read, 2009) to surgical skills training (Lin et al., 2003). In addition, serious games have also been proposed for use over the long term to motivate sustained and repeated participation, increase the effort invested by the participants, and improve the retention of participants over time, with the goal of ultimately leading to more effectives training (Krause et al., 2015).

There is some controversy surrounding the use of *gamification* in serious contexts and even on the use of the term gamification itself. Intentionally called *pointsification* (Bogost, 2011a) or *exploitationware* (Bogost, 2011a), thought leaders suggest that gamification is often a superficial application of trivial game elements—such as points, achievements, badges, and levels (Nicholson, 2015)—rather than the principled application of the mechanics, dynamics, or aesthetics (Hunicke et al., 2004) that create meaningful, emotional, and engaging play. As Bogost notes, "points and levels and the like are mere gestures that provide structure and measure progress within" the game system (Bogost, 2011a, 2011b). For the purposes of this paper, we leverage the term gamification (and *gamify*) when we talk about the application of game-based elements in non-game contexts (Deterding et al., 2011). However, we clarify that our goal of including game elements is to increase motivation (either intrinsic or extrinsic) to engage in sustained and repeated training activities over time.

Waning Motivation in Games for Training In games designed for long-term training, combating waning motivation is of particular importance. One way that researchers have attempted to maintain motivation in longer-term deployments is to keep enjoyment (intrinsic motivation) high because it is a good predictor for staying engaged in a task over time (Haw, 2008). One approach is to inject novelty in the game over time. (Hernandez et al., 2014) deployed a 10-week trial of a networked multiplayer exergame for children with Cerebral Palsy to exercise and socialize together. To keep the novelty of the game high over the 10-week trial, they included six mini-games, which they introduced progressively every two weeks to maintain interest in the training activity. Logs of game choice and time played suggest that the strategy was effective (Hernandez Alvarado, 2015). Similarly, (Mandryk et al., 2013) deployed a 12-week trial of a neurofeedback training system for children with fetal alcohol spectrum disorder to learn to self-regulate. Rather than creating a neurofeedback game, their system turned any off-the-shelf game into a biofeedback game—this decision was largely motivated by the idea that allowing participants to select a commercial game of their choice to use for training would increase the enjoyment, i.e., intrinsic motivation, of training. In addition, the authors note that they originally deployed five game choices, but participants complained that they were bored of the games and thus two new games were added half-way through to keep interest high.

Gamification of training activities has been proposed as a means of keeping motivation high, and these few studies show how researchers attempt to employ novelty to sustain motivation over the long term. However, to our knowledge, there has been no systematic study of how motivation in game-based training wanes over time, and how methods of fostering intrinsic or extrinsic motivation within the context of gamified training activities affect waning motivation.

3.4.3 Experiment Design

We conducted an online study to understand the effects of time and incentives on motivation and behaviour in a game-based training activity. We withheld payment for the first seven days, but paid participants each day thereafter. In addition, we facilitated intrinsic motivation in half of our participants using an existing induction approach of avatar identification (Birk, Atkins, et al., 2016). Each day, we measured subjective motivation, affective state, and task behaviour.

3.4.3.1 Manipulating Identification: The Avatar Creator

To manipulate avatar identification, we used a character creator that has been shown to facilitate intrinsic motivation (Birk, Atkins, et al., 2016). Participants were asked to create an avatar and adjust its appearance, personality, and attributes (characteristics). A minimum of four minutes in the character creator were required, but participants could take longer if they wished. After customizing their avatar, participants were shown a summary of their character and asked to enter a nickname.

Options for the appearance, personality and attributes are described in (Birk, Atkins, et al., 2016). We additionally added customization of the size of nose, eyes, and ears (small, medium, big), the distance of the eyes (narrow, medium, far), and the shape of the head (oval, round, heart, strong jaw) to better facilitate similarity identification (Birk, Atkins, et al., 2016). At the end of the creation process, the avatar was visually presented along with a summary of its personality and attributes, to give the player the sense that their avatar had a profile.

We manipulated identification by presenting half of our participants with the avatar creator. The other half were randomly assigned an avatar of the same sex, from a set of four. Following the approach used in (Birk, Atkins, et al., 2016), participants in the randomly-assigned avatar group watched a 4-minute video of the creation and customization of their avatar. We created four videos for each sex with four different personality and attribute configurations, similar to (Birk, Atkins, et al., 2016), with the additional face options set at the medium level. Participants who watched the video were not allowed to name their avatar; instead the avatar was represented as "Player 1".

3.4.4 Go/No-Go Task: Zombie Apocalypse

Zombie Apocalypse is a zombie themed go/no-go task in which participants stab Zombies with a sword, and avoid stabbing moles. The game was implemented in C# using Unity 4.6 (Unity Technologies, 2014).

We chose the go/no-go task for a variety of reasons. It is used to train cognitive functions (Dowsett & Livesey, 2000), so it is a representative task for a training activity that must be repeated over multiple days. It is simple to explain and does not require a steep learning curve, making it appropriate for an experiment. Go/no-go is also a common game-mechanic (e.g., whack-a-mole), making it familiar to participants in the context of a game, and also straightforward for us to gamify by adding premise, graphical assets, and a score.

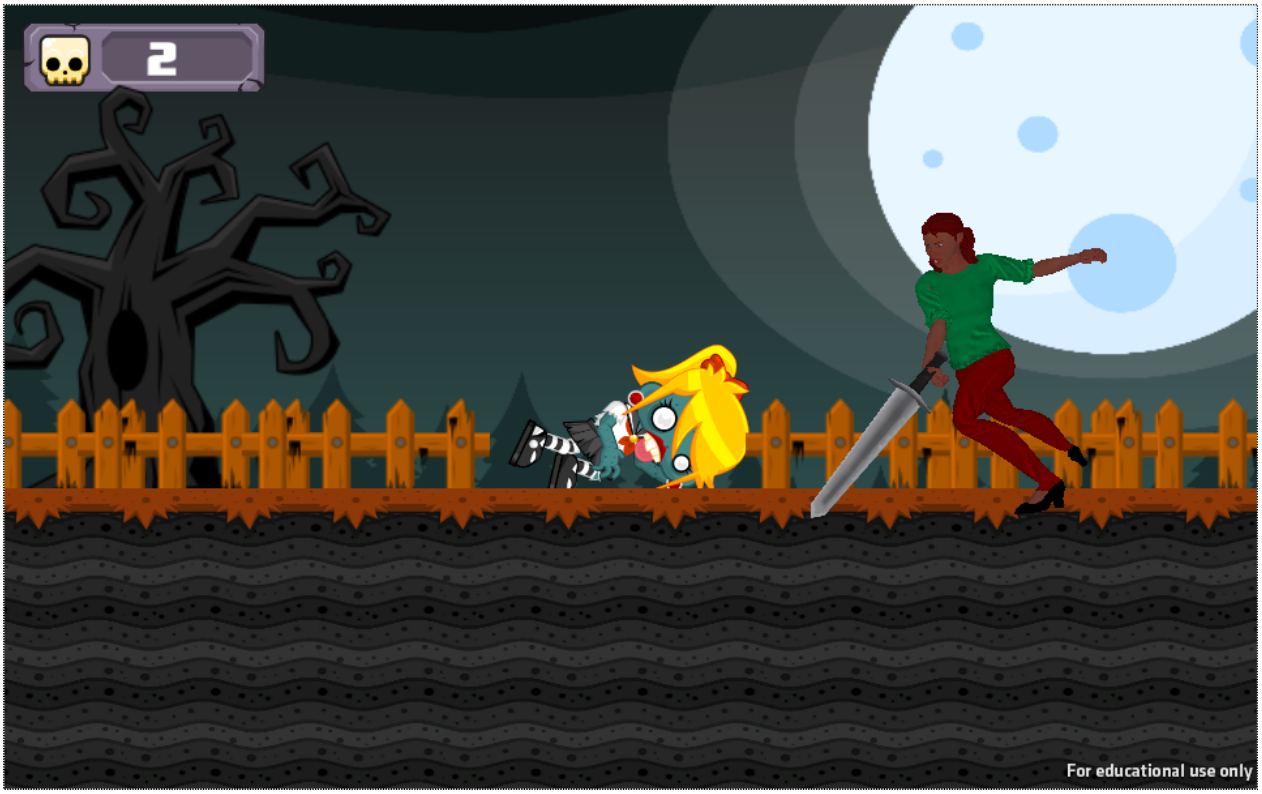

Figure 3.1: Go/no-go task Zombie Apocalypse showing a player hiting a Zombie.

Task: The participants' created or assigned avatar stood in a fixed position holding a sword (see Figure 3.1). Players were instructed to respond to appearing zombies (*targets*), but not to moles (*lures*). Targets and lures were intentionally visually similar to increase task difficulty. A target or a lure appeared every second (popping up from under the ground), giving participants a 500ms window to respond by pressing the spacebar. Correct or false responses to targets or lures results in four response types with different scoring: 1) correctly responding to a target: *hit*, 2) not responding to a target: *miss*, 3) responding to a lure: *false alarm*, 4) not responding to a lure: *correct rejection*. Hits increased the participant's score by 1; misses and false alarms decreased score to a minimum of zero points; correct rejections left the score unchanged. Stabbed zombies (hit) exhibited a death animation, whereas missed zombies (miss) walked away. Stabbed moles (false alarm) turned red before disappearing to give the participant feedback, whereas missed moles (correct rejection) popped back underground safely. Each participant was presented with 80 targets, and 20 lures, presented in a pseudo-random order. Score was displayed and incremented after each target or lure. To ensure that participants understood the task and controls, we presented a training condition with 10 stimuli total (8 targets, 2 lures) before beginning the task each day.

Applications: The go/no-go task requires participants to decide between a target and lure and initiate a response in a very short amount of time, which requires focus and quick decision-making. As a result of the 4:1 relation between targets and lures, participants get accustomed to responding to targets and are thus required to inhibit the most common response, press the spacebar, to correctly reject a lure. Assessment and training of executive functioning is used in both clinical (Casey et al., 1997) and non-clinical (Dowsett & Livesey, 2000) contexts.

Measures We collected both subjective and behavioural measures.

Identification was measured using the avatar-related sub-scales of similarity identification, embodied identification, and wishful identification from the Player Identification Scale (PIS, Van Looy et al., 2012). Participants rated their agreement to identification-related statements measuring similarity—"My character is like me in many ways"—on a 7-pt Likert-scale.

Motivation was measured using the Intrinsic Motivation Inventory (IMI, McAuley et al., 1989). The IMI measures the constructs *interest-enjoyment*—"I enjoyed this game very much.", *effort-importance*—"I put a lot of effort into this game.", *perceived competence*—"I think I am pretty good at this game.", and *tension-pressure*—"I felt tense while playing the game.". Each construct was measured with multiple items using agreement with a 7-pt Likert-scale. The results were then aggregated into the four constructs of the IMI.

3.4.4.1 Task Performance

Performance in a go/no-go task is defined by *hits* and *correct rejections*. Negative performance is described as the number of *misses* and *false alarms*. Following (Kane et al., 2007; Snodgrass & Corwin, 1988), hits and false alarms are used to compute *sensitivity (dL)* and *bias (CL)*. dL separates targets and lures and represents an index of difficulty for discriminating the two types of stimuli—a higher dL indicates that targets and lures are more discriminable. Negative CL scores indicate a liberal bias to respond to target or lure, while positive scores indicate a conservative bias to respond.

3.4.4.2 Participants and Deployment Platform

We conducted our experiment online. 200 participants (43.9% female) with an average age of 31.68 ($SD =$ 8.94) were recruited through Amazon's Mechanical Turk (MTurk), which connects workers willing to do Human Intelligence Tasks (HITs) with requesters of the work, and has been shown to be a valid environment for conducting user studies (Kittur et al., 2008). Participants were re-invited daily to participate in our series of HITs for 11 consecutive days. Participants received $4 for the first day (40 min), and $1 for each consecutive day (10 min). Ethical approval was obtained, and participants provided informed consent. The HIT was only made available to workers in the USA older than 18 with an approval rate above 90%. Participants were excluded from further analysis if: they showed zero variability in more than 2 questionnaires, suggesting that they indicated the same response on all items independent of scale or reversed items; the number of hits in the go/no-go task exceeded 80, indicating that they reloaded the task; or they did not participate on day 4, 5, or 6, which were the days on which we measured identification (30 participants excluded).

3.4.4.3 Procedure

We recruited the initial pool of participants on the first day. Daily HITs were independent, so that people who skipped a daily HIT were not excluded from further participation.

Day 1 Participants were first informed about the amount of compensation they would receive ($4 for approximately 40 minutes); additionally, we informed them that the first payment would be delayed for 7 days, and that they were expected to answer attentively and quickly. It is important to note that by promising a reward at day 7, we introduce an anticipated extrinsic reward from day one. Therefore, participants might experience two different sources of motivation, i.e., the financial reward as an extrinsic source, and identification as a source of intrinsic motivation.

After being informed of the reward, participants were randomly assigned to either the customized or random avatar group. After creating an avatar (or watching an avatar being created), participants performed the go/no-go task, and then completed the IMI. Finally, participants provided basic demographic information, e.g., frequency of play, age, and stated the purpose of the study in their own words.

Day 2–11 Each day, participants were informed about the compensation they would later receive ($1 for about 10 minutes), and were asked to provide informed consent. Participants entered their MTurk identification code to load their saved avatar, and then performed the go/no-go-task, before completing the IMI. On day 4, 5, and 6, we additionally asked them to complete the avatar identification scale.

3.4.4.4 Delivery of Extrinsic Motivator

On Day 8, we paid participants for their initial HIT and continued to pay them each day thereafter. This represents us delivering on a promised (but delayed) outcome (i.e., an expected tangible reward) that is separable from the training activity. Paying a delayed reward at day 8 fulfills three purposes: 1) the separable outcome becomes salient; 2) experienced extrinsic motivation is temporarily increased; 3) paying participants reinforces trust that their efforts would be compensated. We were not interested in the specific reward, but in a proxy for receiving an expected, but delayed beneficial outcome separated from the training task.

3.4.4.5 Data Analyses

All data were logged to a database on a server at the University of Saskatchewan and were analyzed using SPSS 23. IMI data were computed after each completion of the go/no-go task. Identification was computed as average identification of Day 4 to Day 6. To create two groups that varied in their initial intrinsic motivation based on our avatar customization manipulation, we performed a median split on the average similarity identification.

To analyze changes in motivation and performance over time, we performed Hierarchical Linear Models (HLM) with *day* on level-1, and the dependent measures on level-2; *identification* was entered as the covariate.

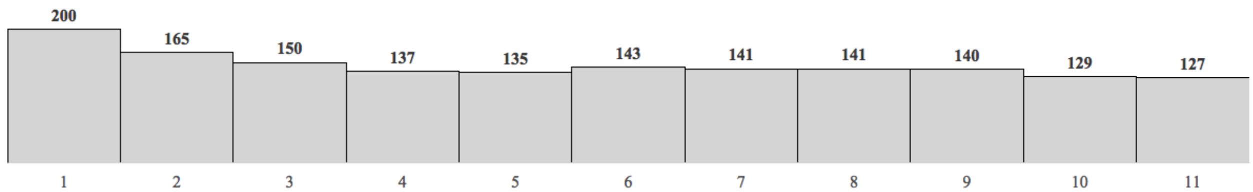

Figure 3.2: Participation by Day (1-11).

Estimates were computed using Restricted Maximum Likelihood with a maximum of 100 iterations, and a maximum step size of 10. To avoid biasing the analysis, covariance was kept unstructured for random intercepts. The α-niveau was set to .05.

To compare improvement or change over time (days) for the two identification groups separately, we computed the slopes for the daily averages within an identification group. Because we used a between-participants design, the absolute performance differences on the go/no-go task between the more- and less-identified groups is not meaningful (due to the individual differences in executive function between players in these two groups). As such, our analyses of the task performance measures calculate the difference between the day of interest and performance on Day 1. Comparisons are made using this difference, which reflects absolute improvement from Day 1 values.

3.4.5 Results

We first present the general participation rates, followed by the results on the initial phase, and then following the reward. In general, participation dropped over time (Figure 3.2). Day 1 was a Tuesday (following a holiday Monday). The average drop-out was 30.6%. Chi-squared tests for each day showed no significant differences in participation between the two identification groups (all $p > .05$).

3.4.5.1 Initial Phase of Waning Motivation

Does motivation wane over time? The HLM (described in the data analyses section) showed a main effect of day on invested effort ($F_{6,820.32} = 6.96, p < .001$). Figure 3.4 shows that effort decreases in general over days. There was no main effect of day on enjoyment ($p = .31$). A general waning of effort confirms our assumptions, reflects prior research (Lin et al., 2015), and is a prerequisite for investigating the effect of a delivered reward on motivation.

Does training improve task performance? All performance data were corrected for baseline (Day 1) performance. HLMs showed that hits ($F_{6,831.33} = 5.50, p < .001$) increase over time, whereas false alarms decrease ($F_{6,823.14} = 4.11, p < .001$), indicating that players improved over time (Figure 3.5). Participants also showed an increase in dL over time ($F_{6,828.76} = 11.86, p < .001$), suggesting that discriminating targets and lures becomes easier over time. In addition, participants were aware of their improving performance.

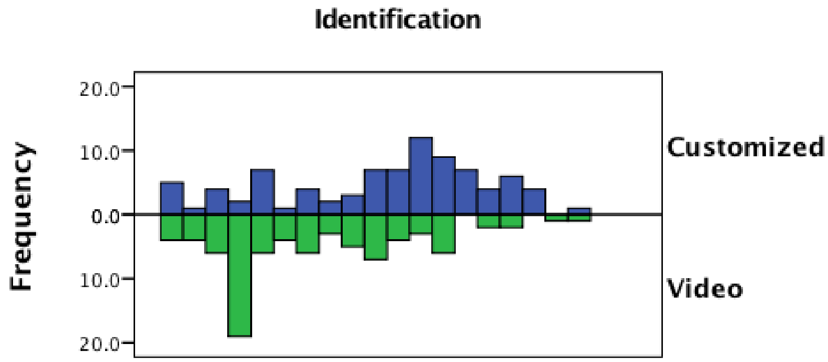

Figure 3.3: Split-Histogram of Identification by Condition.

Perceived competence increases over time ($F_{6,825.35} = 4.84, p < .001$). There was no difference in bias ($F_{6,829.50} = 1.33, p = .243$), indicating that tendency to respond to target or lures didn't change over time.

Can intrinsic motivation be fostered in a gamified training task using avatar identification? Identification can be facilitated through a variety of methods; we chose to use customization to create variability in similarity identification. While the group differences between conditions were significant ($t_{167} = 4.534, p < .001, d = .70$), Figure 3.3 shows how the differences are derived from overlapping distributions of identification. We used a median split of identification in the remaining research questions to capture more-identified and less-identified participants regardless of their group assignment.

The HLM showed that identification is linked to motivation: players with high identification experienced more enjoyment ($F_{1,169.15} = 23.19, p < .001$) and invested more effort ($F_{1,168.65} = 10.96, p < .01$) compared to those who experienced low identification (Figure 3.4). The results confirmed that intrinsic motivation can be fostered through identification.

Is waning motivation affected by identification? To determine whether motivation waned at different rates for the two identification groups, we regressed a line on the daily mean of participant data from Day 1 to Day 7, separately for each group. Comparison of the slopes (explained in the data analyses section) showed significant differences for experienced enjoyment ($\beta = .074, p < .016$) and marginally different slopes for effort ($\beta = .058, p = .065$). The slope values (Figure 3.4) show a decline of effort and enjoyment for participants who experienced lower identification, whereas motivation was more stable for participants with higher identification. See Figure 3.4.

Do group differences in effort translate into performance? We applied slope-analysis (UCLA, 2016) to investigate differences in the improvement over days for the two identification groups. Although identified participants reported higher effort, they did not improve at a faster rate than less-identified participants (Figure 3.4). As Figure 3.5 shows, performance was initially very high on all measures, showing that there

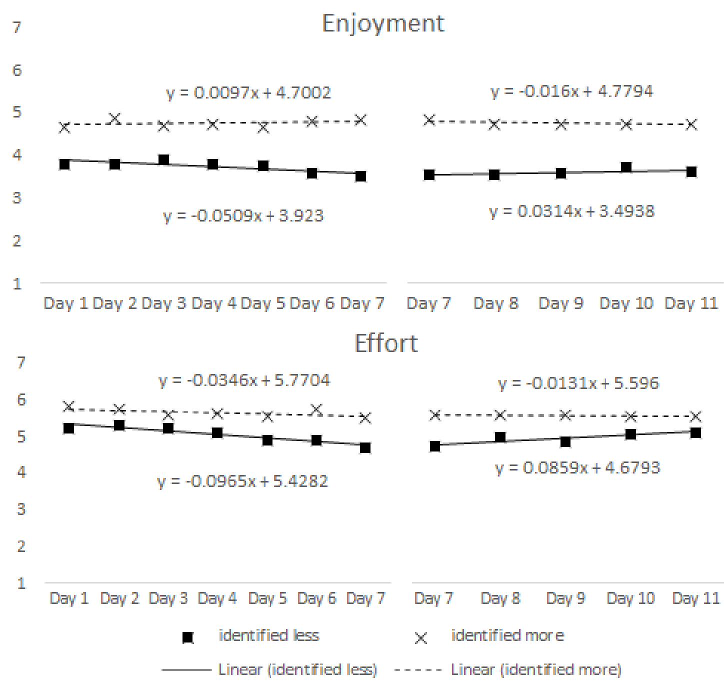

Figure 3.4: Means (marks), Regressions (lines), and Equations for Enjoyment and Effort from Day 1 to Day 7 (left) and Day 7 to Day 11 (right) for less and more identified participants.

was not a lot of gains to be made from practice alone. Differences in performance would be due to training executive function. We address this further in the discussion of the results.

3.4.5.2 Response to the Delivery of an Anticipated Reward

The previous results were focused on Day 1 to Day 7 of our study to establish a differential pattern of waning motivation and determine its effects on performance. In this section, we focus on Day 7 to 11, to show the differential response of participants to receiving the external reward.

How does the delivery of an anticipated reward affect motivation for differentially motivated participants? To investigate the impact of a reward on motivation, we compared the slopes of effort and enjoyment data on days 7-11. Figure 3.4 shows that the lower-identified participants report an increase in their invested effort, whereas effort reported by higher-identified participants was more stable ($\beta = -.086, p = .028$). There was no significant difference in the slopes for enjoyment ($p = .75$).

Do group differences in invested effort following the delivery of an anticipated reward translate into task performance? We compared the slopes of task performance following the reward for the two groups. As Figure 3.5 shows, less-identified participants (who reported increased effort after the payment), showed improvements in performance, reflected in an increasing number of hits. However, more-identified participants showed a reversal in the improvements seen over days 1 to 7: there was a decrease in the number

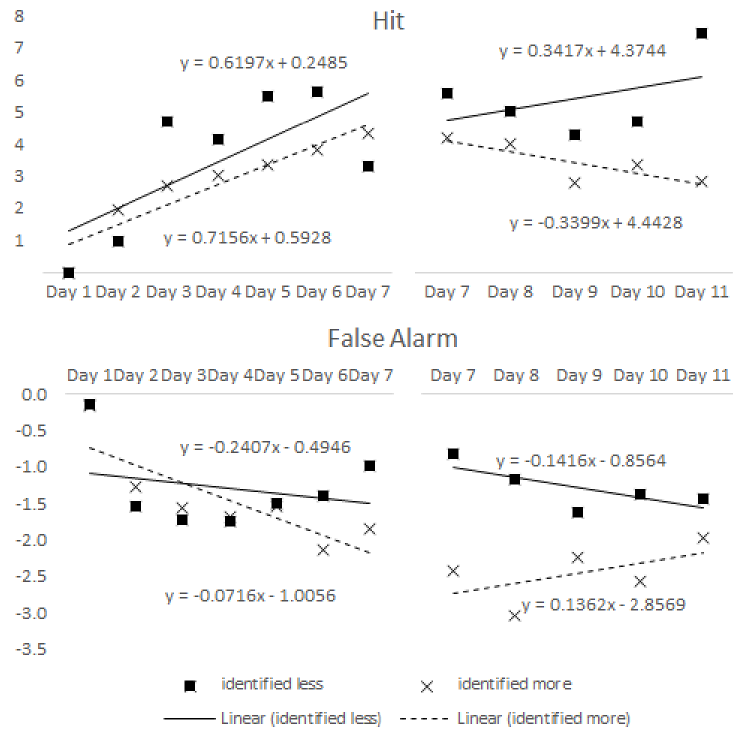

Figure 3.5: Means (marks), Regressions (lines), and Equations for baseline corrected Hits and False Alarms from Day 1 to Day 7 (left) and Day 7 to Day 11 (right) for less and more identified participants.

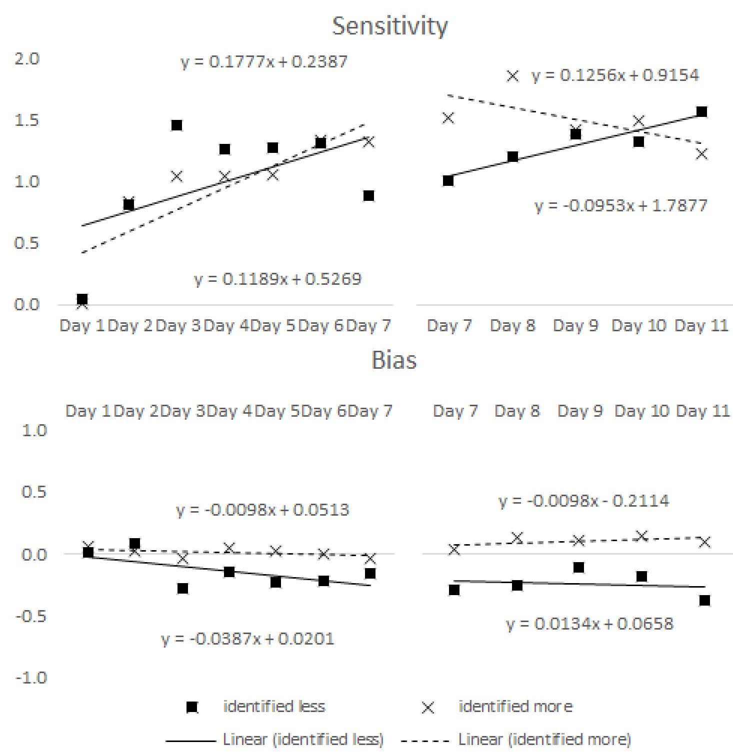

Figure 3.6: Means (marks), Regressions (lines), and Equations for baseline corrected Sensitivity and Bias from Day 1 to Day 7 (left) and Day 7 to Day 11 (right) for less and more identified participants.

of hits. This difference in slopes was significant ($\beta = -1.227, p = .012$). There was a marginally-significant difference in the slopes for sensitivity ($\beta = -.240, p = .052$), showing the same trend. There was no difference in the slopes of false alarms ($p = .270$) or bias ($p = .247$).

Together, these results show that although motivation wanes for each group, identified (i.e., more motivated) participants show less of a decline in the initial 7 days in motivation and greater improvement in some measures of task performance. Following the delivery of the extrinsic motivator, the less-identified group saw improvements in motivation and performance, whereas the more-identified group actually showed declines in motivation and performance, potentially exhibiting the "overjustification effect".

3.4.6 Discussion

We first summarize our results, and then give design recommendations for approaches to foster intrinsic and extrinsic motivation in games for training. Our discussion focuses on the application of our findings to games for training, games user research, and games for entertainment.

3.4.6.1 Summary of Results

Prior to the delivery of the external reward (i.e., on days 1-7), we found the following results.

- We first establish that motivation to invest in our game-based-training activity does wane over time. Although enjoyment of the task does not change with repeated exposure, effort—the primary measure of subjective willingness to invest in the training—decreases over time.

- Second, we show that there is improvement in the training task over time on three of the four baseline-corrected measures of behaviour (i.e., increases in hit, decreases in false alarms, increases in sensitivity). Additionally, participants were aware of their improved performance—their experienced competence also increased over time.

- Third, we demonstrate that we can foster intrinsic motivation in a gamified training task using a manipulation of avatar identification. Although prior work demonstrated this manipulation in the context of a casual game for entertainment [22], we show that the results extend to games for training. Specifically, we show benefits of identification on experienced invested effort and task enjoyment.

- Fourth, we show that the benefits of avatar identification on subjective motivation remain over repeated exposure. Although we expected differences in motivation and effort on Day 1 based on (Ryan & Deci, 2000a), we show in this study that the motivational benefits of this type of manipulation persist over time. In particular, participants with lower avatar identification experienced declining task enjoyment and invested effort over time, whereas participants with higher avatar identification show stability in both enjoyment and effort over the initial week of training.

- Fifth, we show that the group differences in invested effort do not translate into task behaviour—although the measures trend in this direction, the identified group did not improve their performance on the training task at a significantly greater rate than the non-identified group. Following the delivery of the extrinsic reward—the delayed payment—we find the following additional results:

- Sixth, the delivery of the extrinsic reward resulted in an increase in invested effort for the non-identified participants, but not for the identified participants (i.e., those who were more intrinsically motivated).

- Seventh, the training data show that this increased effort reported by non-identified participants translated into task performance in terms of improvements in the number of target hits and marginally in greater sensitivity (i.e., a greater ability to discriminate between targets and lures).

- Finally, in a behavioural exhibition of the effect of an extrinsic reward on intrinsically-motivated people, we found that the participants who identified with their avatar showed a decrease in performance after we delivered the extrinsic reward, as reflected in a decrease in the number of target hits and marginally decreasing sensitivity.

3.4.6.2 Halting Loss of Interest through Motivational Strategies

We motivated our work with the notion that people are initially relatively enthused to begin training programs in various domains, such as to learn a new language or begin an exercise program. However, the initial enthusiasm often begins to wane after the novelty wears off, skill improvement slows down, or the person gets tired of waiting for the beneficial outcome of their efforts. Our work demonstrates that people with differing motivational engagement in a training task will respond in contrasting ways to an intervention intended to halt waning interest. Players who are less intrinsically motivated should respond well to an extrinsic reward, whereas more intrinsically-motivated players likely will not. A good strategy of piquing the interest of less-motivated players is to provide them with a reward, whereas players who are still demonstrating interest in the task might respond better to keeping their interest high. In this section, we present different in-game strategies that could be used to engender interest in a training game.

Extrinsic Motivators Designers wishing to use extrinsic motivators to engage players in a task can reward them explicitly, but can also choose from other approaches. Here, we provide guidance based on the spectrum of extrinsic motivation (Ryan & Deci, 2000a).

External Regulation: Rewards are the classic form of external regulation as they are disconnected from the player's goal but are salient to their interests. In-game examples of external regulation are currencies, tokens, or power-ups.

Introjection: Designing for introjection means focusing on the player receiving approval from themselves or others. Providing a player with a status item (e.g., pets in World of Warcraft), a rank that has social value (e.g., Platinum status in League of Legends), or an unlocked achievement that has personal value (e.g., Win 5000 rounds in Counter Strike:GO) target introjection.

Identification: Providing motivation for players who consciously value the activity they are undertaking should reinforce their self-endorsement of their goals. Asking a player to invest effort in service of a greater goal (e.g., grinding for a quest in World of Warcraft), or learning low-level details to achieve a desirable advantage (e.g., Champion statistics in League of Legends) are examples of identification with the personal importance of a behavior and the acceptance of the regulatory aspect.

Integration: Targeting the congruence that integrated players experience should emphasize the synthesis of the player's goals with the self-view. Challenging oneself because of internal beliefs (e.g., No-Kill run in Fall-Out 3), by accepting the demanding role to lead a team (e.g., Raid leader in World of Warcraft), or by training hard to compete in tournaments (e.g., the League of Legends World Championship) are examples of integrated motivation in games.

Intrinsic Motivation Design choices that target intrinsic motivation should be about maintaining the interest of the players.

Enjoyment: To maintain increase of enjoyment, designers should consider adding novelty to the game (Koster, 2013). This could be achieved through new game levels to play, characters to inhabit, places to explore, weapons to use, or missions to complete. Although novelty is a central concept in maintaining the interest of players, there are other ways to engender enjoyment. Self-determination theory suggests that motivation is formed through the satisfaction of the psychological needs of competence, autonomy, and relatedness. We address design options for each of these in turn.

Competence: Players should feel that they are satisfying their need for competence -that they experience mastery over challenges. Game systems use feedback in the form of scores, stars, or achievements to reinforce the experience of competence. However, a game for training (and skill acquisition in general) suffers from a fundamental problem in that improvements in skill follow a power law—that is, we see smaller improvements with more practice (Newell & Rosenbloom, 1981). If we consider learning any new skill (e.g., instrument, language, game), the massive improvements seen for each unit of effort invested at the beginning begin to level off. Players often quit when the satisfaction of competence wanes (i.e., when it takes a lot of effort to see minor gains in improvement). Designers should consider how to introduce new skills or accelerate skill acquisition to avoid the leveling off of competence satisfaction that occurs over time.

Autonomy: Players should feel like they are making choices in the game and that they are acting under their own volition to feel that their experience of autonomy is being satisfied. We manipulated feelings of autonomy using avatar customization, which previous work has shown to translate into increased effort and motivated behaviour (Birk, Atkins, et al., 2016). Other examples of autonomy manipulations are the branching narratives seen in The Legend of Zelda, the character behaviour choices seen in Mass Effect 3 (Bioware, 2012), and the intensive customization of Avatars seen in Second Life (Linden Labs, 2003). Focused mainly on the choice aspect of autonomy, designers could also highlight the feeling that players are acting under their own volition in choosing to play—a potentially difficult task in a training game.

Relatedness: The feeling of making and maintaining social connections is key to the satisfaction of relatedness. Multiplayer games already have several ways in which the experience of relatedness is satisfied within gameplay. Support for clan play (e.g., League of Legends), games based on a player's social network (e.g., Farmville, PotFarm, or Clash of Clans), and matchmaking algorithms that balance the skills of players in forming a team (e.g, League of Legends, Dota 2, Fifa) all enhance the experience of relatedness in multiplayer games. In addition, relatedness can be fostered in single-player games (Ryan et al., 2006) by helping the player to feel close to the characters within the game (e.g., the companion cube in Portal, or Ellie in the Last of Us).

We have presented several ways in which both extrinsic and intrinsic motivation can be fostered within games. In the next section, we discuss how to apply these motivational designs in several design contexts.

3.4.7 Implications for Design

Our results have several implications for the design of game-based training systems, for games user research, and for games for entertainment and leisure.

Personalizing Training Our results demonstrate how motivational patterns change over time and how responses to a reward differ for players who are more or less motivated to begin with. Characterizing the *motivational level* of players can inform design decisions on when interactive training systems should intervene to halt waning motivation. However, our work also suggests that characterizing the *motivational orientation* of users of these training systems is just as important to inform the appropriateness and timing of interventions. Understanding that people engage in training games for differing reasons, with differing expectations, and with differing levels of interest is essential.

Previous work has suggested that the efficacy of persuasive games can be improved if they are personalized to the individual player type, because different types of players are motivated by different persuasive strategies (Orji et al., 2014). Our work extends on this idea of tailoring games for training by suggesting that knowing the motivational level and orientation of players at any given time can—and furthermore should—inform the choice of the motivational strategy employed, for example by introducing novel content (Mandryk et al., 2013; Hernandez et al., 2014) or by providing game-based rewards (Mandryk & Stanley, 2004). As opposed to (or perhaps in addition to) the trait-based tailoring proposed in (Orji et al., 2014), our work suggests a state-based personalization that takes into account player patterns in motivation over time.

Application to Games User Research Our findings have implications beyond games for training, including in games user research (GUR), which is interested in understanding player experience and applying this knowledge to improve design. The methods deployed have many parallels to our context of games for training. It is common in GUR to pay people to participate in studies—sometimes over multiple days. Similar to players of games for training, gameplay testers may both enjoy the experience and be externally regulated through the payment. As such, a reward is likely to affect players' invested effort, enjoyment, and performance, depending on their level of intrinsic motivation. As such, games user researchers should be careful about how and when rewards are given to participants. In addition, researchers have to be cautious about the conclusions they draw from their tests—actual players of their games are likely to be higher in intrinsic motivation than their testers and thus will likely respond differently to the application of extrinsic in-game rewards.

Applications to Games for Entertainment Although our research was designed to inform the area of games for training, our findings have implications in games that are designed solely for entertainment. Although the intentions of players choosing to play for leisure may differ from those engaged in games for training, there are parallels. In both cases, the goal is to retain players—the purpose in retaining players in games for training is so that they achieve a separate beneficial outcome; in games for entertainment, the purpose of retaining players is driven by the financial outcomes for the company. Next to units sold, the average revenue per paying user (ARPU) is a key metric of success for commercial games and is directly related to the number of daily active users (Casual Games Association, 2012). Acquiring new players is also vital for success. A common approach to acquire new players is to leverage the existing user base through recruitment reward programs in which players receive premium currency, or vanity items (such as mounts or skins) as rewards in exchange for successfully recruiting players from their social network. Our findings suggest that the intertwined reward structure of social outreach has complex implications: the reward might be desirable for a player, but also may be perceived as externally-controlled and may negatively affect their underlying motivation and play experience if they were already enjoying the game.

3.4.7.1 Limitations and Future Work

Although our results provide several important implications for waning motivation in games for training, there are limitations in our study that can be addressed in future work. First, the results that we present are not causal: we facilitate intrinsic motivation using an established induction paradigm (Birk, Atkins, et al., 2016). The paradigm creates variability in avatar identification, which in turn fosters differing motivation in players; however, we cannot interpret our results as causal. Future work can address this limitation by working with existing players who are differentially motivated to participate in a game for training a specific activity; however, randomly assigning intrinsic motivation to a subset of participants in an experiment is not possible by definition. Second, our experience with the avatar identification manipulation raised ideas on how to improve the facilitation of motivation through identification. Third, while MTurk has advantages for research in general, and longitudinal research in particular, the platform is also limited in terms of conducting research on motivation. Workers have the explicit goal to work in exchange for a financial reward. When researching motivation, it would be ideal to have full control over what drives participants to engage in a task. We plan to investigate the differences in reward types such as token rewards (e.g., money), access (e.g., premium features), or social rewards (e.g., attention from others), to create a taxonomy of rewards and their effects on players. Finally, the go/no-go task has the advantage of being simple, well described, and constrained, while also training executive functioning. Most gamified training tasks are more complex and involve layers of gamification, e.g., points, leaderboards, or social reinforcement. Therefore, our results need to be shown in the context of a more complex task, ideally "in the wild".

3.4.8 Conclusion

There are many domains in which people need to invest effort in a training activity over time to see benefit at some future point. Motivation to engage in these types of training activities often starts off quite high, but there is a loss of interest over time, and people often quit before reaching their goal. Games for training use enjoyment to keep interest in the training activity high, even when motivation to achieve the outcome starts to wane. To explore changing motivation patterns over time and to determine how rewards affect players with differing interest levels, we conducted an 11-day study of a game for training executive functioning with players who were split into two groups that reflected their intrinsic motivation induced through a manipulation of identification with an in-game avatar. We show that motivation wanes over time; however, both effort and enjoyment wane more for players who identify less. After one week, when we delivered a reward (payment), the less-motivated group respond positively—increasing their effort and showing improvements in task performance; however, the more motivated group responded negatively in terms of their invested effort and declines in performance on the training task. Explained by theories of human motivation, our findings have implications for games for training, games user researcher, and games for entertainment.

3.5 Summary of Manuscript B

The results showed that avatar identification fosters intrinsic motivation, but peoples' motivation wanes differently over time depending on their level of identification with their in-game avatar; for participants who did not identify with their avatar, extrinsic rewards increased motivation and task performance, while participants who did identify with their avatar showed stable self-reported motivation but showed a decrease in task performance after receiving a reward. The findings are in-line with the overjustification effect and show that motivation induced through avatar customization is responsive to reward delivery. Furthermore, the study demonstrated that motivation is affected by avatar customization in the medium-term.

3.5.1 Lessons Learned

Besides the presented research findings, I have learned several lessons when conducting the presented study. I primarily learned about payment in an online context, but I also gained further insights into avatar customization, motivation, and task choice in online paradigms.

3.5.1.1 Effects of Payment

While previous research has used Amazon Mechanical Turk for multi-day studies before (Daly & Nataraajan, 2015), the presented study was our first experience with crowdsourcing a multi-day study. Participants' commitment and willingness to engage on a daily basis was surprisingly high with an average of 63% engaging every day even if there were not paid daily. The personal feedback Turker's send through Amazon Mechanical Turk's system indicated that people were waiting for payment.

Participants' payment concerns and our results—which show that participants respond differently depending on their motivation—suggest that payment needs to be carefully considered in online studies. A finding that is in line with previous laboratory findings (Staw et al., 1980) and theoretical considerations around the effects of incentives in motivational research (Locke, 1968).

Specifically, my research reveals parallel effects of experience and behaviour, showing that subjective motivational measures have a degree of predictability for motivated behaviour. In the context of my research, it is relevant to see that enjoyment and effort show steeper declines for those who identify less with their avatar over time, indicating that avatar customization might positively affect engagement over time. The overjustification effect is more pronounced on a behavioural level, which is also in line with previous research suggesting that people are not necessarily consciously aware of their decisions, especially not if the behavioural change is minimal. These findings suggest that it is difficult to identify motivational changes from self-reported data. Therefore, spatial measures of motivation that can distinguish extrinsic and intrinsic motivation are required. Furthermore, because the effects on motivated behaviour are difficult to measure, it is important to design online studies with potential effects of financial incentives in mind, e.g., by developing payment plans that prevent potential effects on engagement.

3.5.1.2 Avatar Customization and Motivation

In Chapter 2, I showed that avatar customization affects motivated behaviour in a game context. One important outcome of the presented study was that avatar customization affects experience and behaviour in the context of a simple task instead of a game-context. In the context of my thesis, to show that avatar customization applies outside of the game context is an important interim step to investigate the potential of game-based motivational design strategies applied to training tasks for health.

However, avatar customization does not produce group differences. Similar to the experiment discussed in Manuscript A, identification is the mediating factor. When considering identification as a proxy for motivation, there are several issues: 1) Median splits of variables decrease the finding's internal validity, because the results do not include the entire range of the variable. 2) Because there are no group differences, it is not possible to claim causality, i.e., avatar customization cannot be discussed as the reason for the presented findings. 3) There are open questions around the levels of motivation and how being differently motivated might explain my findings in more detail.

3.5.1.3 Levels of Motivation

In the presented study, I did not directly measure levels of motivation but argued that the effect of avatar customization on motivation has been shown before in Manuscript A. Therefore, I claimed that avatar customization affects intrinsic motivation measured by identification as a proxy. In future studies, it would be beneficial to measure different levels of motivation to gain more fine-grained insights into participants' motivation to engage with a training task. Another added benefit of more fine-grained measures would a better understanding of the impact of financial incentives in multiday studies—are participants only engaging because of a separable outcome (i.e., money) or because they enjoy the activity?

3.5.1.4 Advantages and Drawbacks of the Go/No-Go Task

The go/no-go task allowed me to implement a very simple task that could easily incorporate a game-theme and allowed for believable user interactions with the task. While the task could be perceived as game-like, I intended it to train peoples' ability to learn better control over their responses—a skill that has been shown to have positive effects on mental health (Dowsett & Livesey, 2000; Thorell et al., 2009). Unfortunately, the task showed ceiling effects, i.e., most participants learned to perform the task flawlessly. To measure performance differences, a task with high enough difficulty would be needed. Another alternative would be to avoid performance tasks altogether and rely on other outcome measures, such as measures of mood change or self-reported responses to constrained situations such as emotional inductions.

3.5.2 Relevance in Context

The presented study addressed differential effects of reward delivery on intrinsic motivation facilitated through avatar customization. Within my dissertation, Manuscript B contributes in the following ways: 1) I present my first investigation of medium-term effects of avatar customization; 2) I investigated the responsiveness of avatar customization-induced motivation to known effects on motivation, i.e., the overjustification effect (Deci et al., 2001; Huillery & Seban, 2014; Amabile, 1993; Hitt et al., 1992; Seaverson et al., 2009); 3) In the manuscript, I introduced a more nuanced perspective on motivation following research on Self-Determination Theory (Ryan & Deci, 2000a); and 4) my research touches on the tension between monetary incentives in the context of research on motivation—highlighting that it is important to find ways of incentivizing participants without subtly affecting results, especially not when investigating intrinsic motivation.

In the context of my dissertation, the presented study contributes to my goal of building a scientific basis to understand and utilize avatar customization as a motivational design strategy by replicating previous effects presented in Chapter 2, along with a more nuanced look at motivation as a result of avatar customization. While the study confirms that avatar identification drives effects of self-reported motivation and behaviour, the study goes beyond the finding itself by showing that induced motivation is responsive to known manipulations of motivation, highlights the impact of incentives on study designs, and provides insight into MTurk workers' motivation and behaviour over time.

Chapter 4

Combating Attrition

Citation: Birk, M. V., & Mandryk, R. L. (2018, April). Combating Attrition in Digital Self-Improvement Programs using Avatar Customization. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (p. 660). ACM.

Acknowledgment: Under my direction, Cherylin Atkins implemented the breathing exercise and modified the avatar customizer described in Manuscript B.
In coordination with my supervisor, Dr. Regan Mandryk, I was responsible for the research direction, experimental design, system design, data gathering, data analysis, and reporting.

My goals in Manuscript C were to study the effects of avatar customization on sustained engagement, operationalized as login behaviour. In this study, I adjusted the study protocol in response to the lessons learned in Manuscript A and B, and replaced the video control group with a control group in which an avatar was assigned instead of showing a video of the avatar creation process. The study design was strongly influenced by reflecting on previous lessons learned, for example those described in Chapter 3, in which I showed that incentives affect experience and behaviour—consequently, I adapted our payment procedure. Our findings are explained in the context of my previous work—understanding user behaviour as a function of identification, and in relation to levels of motivation as described in Manuscript B.

In the study presented in Manuscript C, 400 participants were invited to participate in a three-week long study using Amazon Mechanical Turk. Central to the study was a 19-day long breathing exercise divided into a pre-notification phase (Day 2-7), a notification phase (8-15), and a post-notification phase (16-19) to compare the effects of customization to a better-known (and more explicit) manipulation of engagement (i.e., notifications). Participants created their avatar on Day 1 or were introduced to an assigned avatar. Pre-and post the multi-day breathing exercise (i.e., Day 0 and Day 20, respectively), we measured levels of motivation. Additionally, we measured demographic information on Day 1 and conducted an exit survey on Day 20.

Participants were paid \$2 on Day 0 for filling in demographic information, which took about 10 minutes. On Day 0, participants were re-invited to partake in the multi-day study, starting the next day. 62.5% (250 participants) returned to the multi-day breathing exercise on Day 1. Building on the insights about reward delivery gained in Chapter 3, participants were paid \$10 upfront for Day 1-19, as participants who fully engaged voluntarily in the study would spend approximately 60 minutes over three-weeks working on the study. On Day 20, participants were paid an additional $2.50 for filling in a final survey. In the context of crowdsourced online studies, the payment plan was designed to minimize the effects of delayed payments on participants' motivation and to be considerate about the effects of incentives on motivation in general.

4.1 Problem and Motivation

In Chapter 3, I showed in a multi-day game-based training that motivational effects associated with avatar identification persist over time and respond differentially depending on the level of motivation to reward delivery—the results are in line with previous research (Deci et al., 2001; Huillery & Seban, 2014; Staw et al., 1980). The results are relevant to address the problem formulated in the problem statement of my dissertation—Manuscript A provided insights into the general effects of identification on experience and behaviour. And Manuscript B suggested that identification can facilitate sustainable engagement. However, both studies deployed video games (i.e., an infinite runner in Manuscript A) or game-based intervention (i.e., a game-based training in Manuscript B), limiting the generalizability and applicability of the results. Therefore, the problem I addressed in this Manuscript is that avatar customization shows promising effects when avatars are displayed either embedded into a game or into a game-based training application, but it remains unclear if avatar customization can facilitate sustained engagement with a non-game-based training task over time.

4.2 Solution and Steps to Solution

To address the question if avatar customization has the expected effects on sustained engagement with a non-game-based training task, I conducted a 21-day long online study in which participants engaged for 19 days in a daily breathing exercise, either with a customized avatar or an assigned avatar. Based on my previous research (See Chapter 2 and 3), it is fair to assume that avatar customization can increase engagement and reduce login attrition in the medium-term. Reduced attrition in a non-game-based training task would indicate that avatar customization is indeed a motivational design strategy that fosters inherent motivation to engage over time outside of a game-context and facilitates sustained task engagement. While previous effects are encouraging, the studies presented in Chapter A and Chapter B prioritized internal validity, while maximizing external validity—isolating small effects was our main priority in both studies. The manuscript presented in this Chapter prioritizes external validity by focusing on a research context that allows for extrapolating potential user behaviour in-the-wild as a result of customization. Design strategies that facilitate motivation are especially important in contexts with low external regulation or non-existing regulation as common, for example, in internet-based training. Medical treatment faces similar issues—medication adherence, for example, is a known issue that thwarts the efficacy of many medical products and has spawned digital products to combat low adherence (Huckvale et al., 2015; Griffiths et al., 2010; Heber et al., 2016; Lorig et al., 2010). External regulation, however, is expensive and systems such as notifications

(Cutrell et al., 2001), nudging (Thaler & Sunstein, 2008), or human support (Schueller et al., 2017) can easily be ignored or are unavailable or too expensive. In the study presented in Manuscript C, I investigated avatar customization pre-notifications, during notifications, and post-notifications to contrast the effects of intrinsic motivation with a common interface design approach for increasing adherence. From the research presented in Chapter 2, I learned that rewards can affect experience and behaviour in online tasks. Therefore, I introduced a sophisticated payment plan to minimize potential effects of incentives on my participants. The payment plan was deployed with the aim to allow participants to engage with the daily exercise without feeling that they need to engage with the task because of a separable outcome, but because they want to engage with the task under their own volition. This payment plan allowed us to investigate intrinsic motivation and engagement without regard to payment (see Section 4.3.3). To measure and control for other forms of motivation besides intrinsic motivation, I used the Situational Intrinsic and Extrinsic Motivation Inventory (SIMS) to measure different levels of motivation (see Section 4.3.4).

4.2.1 Research Questions

To address the efficacy of avatar customization on a non-gamed-based self-improvement program over time, we investigated the following research questions:

RQ1. Does customizing an avatar combat attrition?

RQ2. Does customization affect avatar identification?

RQ3. Does customizing an avatar affect situational intrinsic or extrinsic motivation over time?

RQ4. Are the differences in logins explained by increased identification (controlling for pre-existing motivation)?

4.3 Design Rationale

The study design was geared to investigate the effect of avatar customization on attrition in a semi-naturalistic setting, allowing me to infer the efficacy of avatar customization in-the-wild. To create a study that maximized external validity, while assuring internal validity, I did the following: 1) I moved away from a game-based training task as used in Manuscript B. 2) I used Amazon Mechanical Turk, because our previous experiences suggested that MTurk workers show commitment to a study, but vary enough in their behaviour (i.e., drop-out) to fulfill our requirements. Most importantly, however, workers are not constrained by lab times or build any personal connection with the experiment that increases their commitment—very similar to the experience of using a technology-enabled service for health without supervision. 3) I carefully considered the impact of incentives aiming to avoid that incentives affect participants' login behaviour over time. 4) I measured and compared subjective experiences of levels of motivation as described in Chapter 3 with the goal of gaining a better understanding of why participants engaged with the breathing exercise over time. 5)

In addition to evaluated scales of motivation, I asked participants at the end of the study about the reasons why they DID and DID NOT participate using "check all that apply" questions.

4.3.1 Non-game-based Training: Breathing Exercise

In my previous studies, participants customized an avatar or watched a video of someone else customizing an avatar before engaging with the same avatar in a game-based context, e.g., an infinite runner (Manuscript A) or a Zombie themed go-no-go task (Manuscript B). To extend the applicability and usefulness of my work beyond games, I decided to move to a simple breathing exercise, which is a) a non-game-based training, b) allows for daily practice, c) has believable health implications (i.e., the participants would believe that it was the intervention under investigation), and d) allows for the integration of an avatar.

4.3.2 Long-term Worker Behaviour and Study Commitment

Having worked with more than 5000 Amazon Mechanical Turk workers across multiple studies gave me valuable insights into the behaviour of workers. Previous research has already shown that results produced online are comparable to lab results (Buhrmester et al., 2011; Kees et al., 2017; Crump et al., 2013; Kittur et al., 2008) and that the majority of workers earns extra, rather than primary, income on Amazon Mechanical Turk (Ross et al., 2010) (only a minority actually makes a living on MTurk, see Ross et al., 2010). These findings combined with my experiences with workers led me to have confidence that workers are a sample that might be slightly skewed—there are more white, well-educated workers than would be representative in the US population (Huff & Tingley, 2015; Berinsky et al., 2012)—but that their general behaviour is similar to what would be expected from participants in the lab, i.e., they hold commitments for the most part, work as agreed on, and even provide unsolicited feedback to potential protocol errors or technical issues. Because workers operate remotely from their homes, there is no control over the environment and it is important to build in processes that allow for the assurance of data quality (e.g., the button pressing mechanic used in Manuscript A), but the trade-off of dealing with a few participants who show deviant behaviour—which can also be a problem in lab studies—is well worth the advantage of studying effects of interface design (i.e., avatar customization) on user behaviour in a semi-natural setting. Because of the natural behaviour of workers and the advantage of a semi-natural study environment, we decided that when optimizing for external validity, MTurk would provide advantages that a lab study could not provide without a substantial amount of resources and would even then be inferior, e.g., most people would not commute to partake in a one-minute digital intervention.

4.3.3 Incentives

In a Letter to the Editors of World Psychiatry, John Torous and Joseph Firth comment on the dichotomy of actual versus aspirational digital health (Torous & Firth, 2018) and highlight that novel study formats such as crowdsourced user studies have huge potential to add to our understanding of mental health, but that we need to consider the effects of incentives and how they might affect the performance of real-world (i.e., not in a study context) applications. For the study presented in Manuscript C I prioritized external validity—i.e., in study length, application choice, sample source—which is reflected in the experimental design, the study material I chose, and guided our decisions around our payment protocol. I decided to pay participants upfront to assure that payment did not affect their decision to engage with the task on a daily basis. With $10 for 3 minutes per day over 19 days (57 minutes in total), I decided to pay fairly but not to overpay with the intention of avoiding that participants' engagement could be a function of payment. $10 for 400 participants upfront is a serious investment of resources. Therefore, I decided to build an initial pool of 400 participants (Day 0) who were re-invited to partake in the multi-day part of the study (starting Day 1)—250 people returned on Day 1 and were paid $10. My reasoning here was that this way I would only have participants in my pool who have shown that they return after an invite and are therefore more likely to show further commitment. I assured that participants were swiftly paid, accepted the task and completed the first run through on Day 1. By paying participants with only a three-hour delay, I minimized the risk of potential effects on participants' motivation. I separated the payment of the exit survey from the payment of the first study, because I was interested in getting as many responses as possible to the exit survey—both from people who regularly completed the breathing task and from those who did not.

4.3.4 Subjective Experiences of Situational Intrinsic Motivation

In Manuscript B, I inferred levels of motivation based on my findings from Manuscript A—I inferred that avatar customization increased intrinsic motivation, reflected in self-reported intrinsic motivation and behaviour (i.e., reward response), for those who identified with their avatar. However, the assessment of levels of motivation as described in Chapter 3 could be improved by using validated scales that allow for the assessment of levels of motivation beyond intrinsic motivation. The Situational Intrinsic and Extrinsic Motivation Scale (SIMS) (Guay et al., 2000) is composed of four subscales and designed to assess intrinsic motivation, identified regulation, external regulation, and amotivation—the scale has been shown to be responsive to experimental inductions (Ibid). Adding an explicit measure of different levels of motivation allowed me to measure participants' motivation over time and investigate if avatar customization affects self-reported motivation—distinguishing for example if participants engaged because they inherently enjoyed the task (i.e., intrinsic motivation) or felt they needed to because of an external reason such as a reward (i.e., external regulation).

4.3.5 "Check All That Apply"-Reasons for Participation

During the exit survey of the experiment, I asked participants to reply to an "all-that-apply"- question with the intent to gauge why participants DID and why they DID NOT participate. My reasoning for proving preselected answers was to gain data-driven insights into questions that potentially would lead to a better understanding of participants' behaviour and potentially even spawn additional research questions. For example, "I liked seeing my avatar" (assigned 9%, customized 16%) confirms our assumptions and shows that the effects of customization even effect explicit statements of liking. The question "It became a habit for me" (assigned 18%, customized 28%) is interesting from a research perspective, because it infers that when using customized interfaces, I might be more likely to form habits, which would have implications for every application that relies on habit forming, e.g., social network games, medical adherence programs, or dating apps.

4.4 Manuscript C

Digital self-improvement programs (e.g., interventions, training programs, self-help apps) are widely accessible, but can not employ the same degree of external regulation as programs delivered in controlled environments. As a result, they suffer from high attrition—even the best programs won't work if people don't use them. We propose that volitional engagement—facilitated through avatar customization—can help combat attrition. We asked 250 participants to engage daily for 3 weeks in a one-minute breathing exercise for anxiety reduction, using either a generic avatar or one that they customized. Customizing an avatar resulted in significantly less attrition and more sustained engagement as measured through login counts. The problem of attrition affects self-improvement programs across a range of domains; we provide a subtle, versatile, and broadly-applicable solution.

4.4.1 Introduction

People often find it difficult to activate themselves to undertake behaviours that are good for them, such as engaging in regular exercise, learning a language, or taking care of mental wellness by practicing mindfulness. Even when people are motivated to begin a program of self-improvement, they may find it difficult to adhere long enough to see the resulting benefits. To help activate these behaviours, people seek out—or societies institutionalize—controlled environments that provide the scaffolding to help obtain the desired outcomes, such as being fit, healthy, or educated. For example, people who wish to undertake an exercise program may hire a personal trainer even if they already know which exercises to do; societies establish formal school environments for children to support learning; people with serious addictions may cloister themselves in formal rehab programs; and people may seek out counselling services to improve their mental health.

These controlled environments rely on *external regulation*—taking action to satisfy an external demand or obtain an externally-imposed reward (Ryan & Deci, 2000a)—to encourage continued participation and avoid attrition, even when motivation wanes due to competing demands on our time, attention, or resources or simply because we lose interest over time. The teacher, trainer, and counselor all help reinforce the commitment to self-improvement through the use of verbal and written contracts that have penalties for breaking them.

External regulation can be an effective method of promoting participation in self-improvement programs; however, access to these programs is often limited. The cost of a personal trainer is much greater than that of an exercise app; attending weekly Spanish classes requires instruction and facilities that are completely negated by a language-learning program such as Duolingo[1]; and access to counselling services can be limited by geography (Canadian Mental Health Association, 2016) or health-care systems that do not have enough therapists to meet the demand (World Health Organization, 2014). To improve the accessibility of programs of self-improvement, researchers and designers have turned to digital approaches (e.g., apps, internet-based training, forum groups) so that program delivery can be delivered *at scale*—i.e., at the size required to address the problem. However, digital self-improvement programs delivered at scale often cannot employ the same degree of external regulation as their non-digital counterparts.

External regulation has been introduced into digital self-improvement programs delivered at scale, for example through notifications (e.g., https://www.myfitnesspal.com/), login streaks e.g., http://www.udacity.com), locking out access to other programs (e.g., http://www.saent.com), and contingent rewards (Jeni, 2017; Dale et al., 2018); however, these are typically weaker forms of external regulation that can be easily circumvented or ignored by people. As such, in digital self-improvement programs delivered at scale, we must compensate for the loss of external regulation, so that program efficacy is not compromised in service of increased access.

We propose that increasing the volitional engagement of participants can help to counteract the loss of external regulation in digital programs delivered at scale. Volitional engagement implies that people are engaging as a choice (Gagne, 2003), and is associated with being more intrinsically motivated—i.e., doing something because it is inherently interesting (Ryan & Deci, 2000a). A common approach to making digital self-improvement programs inherently interesting is to leverage mechanics and features from digital games, which are known to be intrinsically motivating to play (Ryan et al., 2006). Often referred to as gamification, i.e., the use of game design elements in non-game contexts (Deterding et al., 2011), these approaches have been shown to increase participation in self-improvement programs in the short term and over repeated exposure (Richter et al., 2015).

Previous work has shown that one such technique—that of avatar identification through customization—increased enjoyment of a casual game and translated into increased motivated behavior as measured by greater time spent in free play (Birk, Atkins, et al., 2016). Further work showed that the same technique sustained the increase in enjoyment and invested effort in a simple training task (a gamified go-no-go task) over repeated exposure. However, in this prior work, behavioural differences were only shown in the immediate-term, i.e., directly following play; there was no evidence of increased retention of participants over time. Over the medium-term, there was evidence of increased enjoyment for those who identified with their customized avatar, but this effect was negatively influenced by experiment payment in an example of the overjustification effect (Deci et al., 1999)—in fact, the use of daily payment meant that participation rates could not be interpreted, as they could not be attributed to increased volitional engagement as a result of the technique

[1] http://www.duolingo.com

(Birk, Mandryk, & Atkins, 2016).

In this paper, we investigate the use of avatar customization as a method of decreasing attrition in a medium-term digital program of self-improvement delivered at scale. We either assigned participants a generic avatar or asked them to customize their own avatar who would lead them through a one-minute daily breathing exercise. We chose breathing as the program due to its simplicity and believability as an intervention for improved mental wellness (Ahtinen et al., 2013)—the experimental cover story. To avoid the external regulation associated with payments, we paid participants in advance and asked them to complete the daily breathing exercise at their leisure—no reminders were given for the first week, a daily notification was provided for the second, and no notifications were again provided for the third and final week. We logged their participation and their state anxiety before and after breathing. We also used validated scales to assess traits that might explain their behaviour and their degree of amotivation, external regulation, identified regulation, and intrinsic motivation over time.

Our results show that avatar customization increases participation over a 3-week program, that this difference cannot be attributed to accompanying increases in identified regulation, and that identification with the avatar does not fully explain the benefits of avatar customization on logins. Even the best-designed self-improvement programs don't work if they aren't used; our work provides a step in using interaction design to decrease attrition and promote engagement.

4.4.2 Related Work

We present work on understanding our need for self-improvement, theories why we deviate from positive and intended behaviour, and the cost of attrition. We provide background literature on interaction techniques that combat attrition. We then explain the theoretical background of volitional engagement and current approaches for fostering volitional engagement in digital self-improvement.

4.4.2.1 Challenges in Self-Improvement

Be it acquiring knowledge in educational settings, improving physical or mental health, or learning new skills, as people, we enjoy and seek out opportunities to improve ourselves and grow. Lifestyle decisions (i.e., how we individually address personal growth) and self-improvement (i.e., taking action to become better) have long been addressed in psychological research, for example by Self-Determination Theory (Ryan & Deci, 2000b)—as the axiom of personal growth; in Self-Efficacy Theory (Maddux, 1995)—as the inherent belief that one can succeed in facing new challenges; and in Positive Psychology (Seligman & Csikszentmihalyi, 2014)—captured in the attempt to answer the question "What is a good life?".

Most self-improvement actives, such as learning a new skill or exercising, require persistence, but motivation wanes over time (Bean & Eaton, 2000; Dishman, 1991; Eysenbach, 2005; McAuley et al., 1990) and life circumstances change; starting a new job or working to meet a deadline can change our focus and compromise our ability or willingness to engage in behaviour that leads to self-improvement.

How to practically enable personal growth and avoid the pit falls of waning motivation has been addressed by, for example, applying Self-Efficacy theory to exercise programs (Marcus et al., 1992) or Self-Determination Theory in the context of learning (Rigby et al., 1992). Broadly speaking, theories of user engagement can be separated into social-based theories (Festinger, 1954; Schilling & Hayashi, 2001)—those that investigate the role of others in our behaviour; need-based theories (Atkinson & Litwin, 1960; Maddux, 1995; Locke & Latham, 2002; Ryan & Deci, 2000b)—ones that assume that our actions are determined by internal needs, e.g., the need to experience mastery; and reward-based theories (Fishbein & Ajzen, 1975; Skinner, 1972)—theories that attribute actions to expected or learned outcomes.

The Impact of Attrition Although people are often motivated to begin programs of self-improvement, waning motivation results in reduced participation over time. Attrition—the loss of participants over time—has been shown to negatively affect the outcomes of different forms of training (Bean & Eaton, 2000; Dishman, 1991; Eysenbach, 2005) as even the best forms of training do not work if people do not use them. Attrition is, for example, studied in the context of eHealth interventions (Christensen, Griffiths, & Jorm, 2004), because randomized clinical trials (RCTs) need to capture drop-out to evaluate the quality of an intervention. Dropping out of a digitally-delivered intervention like MoodGYM (Christensen et al., 2002) means that the positive outcome is lost—if training is not performed, it can't show the desired outcome. Eysenbach defines the law of attrition as "... the observation that in any eHealth trials, a substantial proportion of users drop out before completion or stop using the application" (Eysenbach, 2005). Whereas Eysenbach focuses specifically on eHealth applications, the issue—not necessarily the health implications—of drop-out are the same across domains. Systems like MoodGYM (Christensen, Griffiths, & Jorm, 2004), (Christensen, Griffiths, & Korten, 2004) show higher completion rates under external regulation compared to completion rates "in-the-wild"; 22.5% in a clinical trial ($N = 182$) compared to only 0.5% for the completion of all 5 modules in-the-wild ($N = 19607$).

In the context of mental health interventions and applications, Killikelly et al. (2017) suggest four approaches to study the flip side of attrition, called adherence: (1) the analysis of drop-out rates, (2) relationships between varying factors, e.g., intervention time, (3) post-study questionnaires of experience, and (4) experimental factors to manipulate adherence. Previous studies have also highlighted the importance of support (Mohr et al., 2011), to cater to the target audience's behaviour (Doherty et al., 2012; Mohr, Tomasino, et al., 2017; Mohr et al., 2011), and community (Schlosser et al., 2016). Quantifying the negative impact of attrition is challenging (Christensen et al., 2009; Eysenbach, 2005), because once people have dropped-out of a study or a training program, they are difficult to access (Eysenbach, 2005). As an alternative, researchers have developed models to better understand drop-out in the context of health (Christensen et al., 2009; Eysenbach, 2005), learning (Bean & Eaton, 2000; Berge & Huang, 2004), and exercise (Dishman, 1991; McAuley et al., 1990; Mullen et al., 2013) programs, as well as identifying potential predictors of attrition (Killikelly et al., 2017).

4.4.2.2 Combating Attrition

Traditional methods to combat attrition rely on a variety of methods, including: peer-support (Dennis, 2003)—e.g., Alcoholics Anonymous (AA) defines a support group, but also fosters the relationship with an experienced alcoholic (i.e., a sponsor) to help participants abstain (Kast et al., 2012); oral or written contracts (Rousseau, 1995)—e.g., a recovery contract defines goals, a timeline, and consequences (personal or legal implications), to support recovery; and financial penalties (Pearson & Lieber, 2009)—e.g., personal trainers or counselors will usually charge for an agreed-on-appointment, regardless of attendance.

Technology-Supported Methods to Combat Attrition Technology offers novel opportunities to combat attrition. For example, push-notifications remind users to engage in an activity, showing positive effects on attrition (Kelders et al., 2012). Research on notifications has investigated general notification preferences and behaviour (Iqbal & Horvitz, 2010), notifications that are sensitive to context (Iqbal & Bailey, 2008) and attention (Horvitz et al., 1999), personalized notifications (Etter et al., 2006), and subtle changes in user interfaces (Kerber et al., 2016). A main driver for research on notifications is the finding that notifications are disruptive (Bailey & Konstan, 2006; Czerwinski et al., 2004), and negatively affect memory (Cutrell et al., 2001) and task performance (Bailey & Konstan, 2006).

To combat attrition, platforms incorporate strategies to create more engaging experiences. For example, the design of SilverCloud—an online platform designed for delivery of Cognitive Behavioural Therapy (Doherty et al., 2012)—incorporates four strategies to engage users: *Interactivity*—supporting active user engagement, i.e., allowing for meaningful choices; *Personalization*—content that is geared towards user needs and allows for customization; *Support*—enabling contact to a health professional or therapist; and *Social*—enabling community building and peer-interactions.

Online peer-groups have been formed around many topics of interest, e.g., diet, exercise. But online services that cater to training-at-scale, e.g., Udacity[2] or Lynda[3] as learning platforms, go beyond mere information exchange and offer specialist services that promise personal advancement everywhere at any time. Being able to communicate with others, exchanging what has been learned, and discussing concepts and theories with others, are essential components of online learning-at-scale and have been shown to keep users engaged (Mohr, Tomasino, et al., 2017).

4.4.2.3 The Theory of Volitional Engagement

Approaches for increasing adherence and reducing attrition in programs of self-improvement delivered at scale generally work to increase the voluntary participation of users—referred to as volitional engagement (Gagne, 2003). Volitional engagement draws on theories of human motivation to explain why people choose to engage in a task (Ryan & Deci, 2000b). For example, in a learning program, motivation can be described as different

[2] http://www.udacity.com
[3] http://www.lynda.com

regulatory styles: *intrinsic motivation*—i.e., learning for the inherent satisfaction of gaining knowledge; *identified regulation*, i.e., learning because one values being knowledgeable; *external regulation*—i.e., learning because the outcome (gained knowledge) is perceived as means to get a good job; and *amotivation*, i.e., the lack of any intention to learn. It is generally understood that more intrinsic forms of motivation are beneficial for volitional engagement.

Multiple strategies have been applied to foster volitional engagement in a digital context. For example, gamification—the application of game elements in a non-game context (Deterding et al., 2011)—has successfully been applied to increase volitional engagement in a variety of contexts (Richter et al., 2015; Seaborn & Fels, 2015). The at-scale mobile language learning Duolingo (*http://www.duolingo.com*, n.d.), for example, successfully incorporates multiple mechanics to increase user engagement and combat attrition, such as streak-based incentives, completion-based rewards, leaderboards and avatar customization. In a similar vein, persuasive technology uses strategies to bring about change by shaping or reinforcing behaviors or attitudes (Fogg, 2002). Two strategies commonly employed in persuasive technologies are personalization (Kaptein et al., 2015; Nouchi et al., 2013)—which is *system-initiated* tailoring that offers content or services personalized to an individual; and customization (Orji et al., 2014; Sundar & Marathe, 2010)—which is a system that supports *user-initiated* tailoring of content or services. Sundar and Marathe (2010) argue that although personalization will increase the relevance of content for individuals using an interactive system, customization yields systems and content that are not only relevant, but also boost the agency and self-determination of the individual because it is they themselves who perform the tailoring. Customization fosters autonomy, a sense of control, and a sense of identity, making the person feel relevant in the context of their interactions (Ibid).

4.4.2.4 Application of Volitional Engagement

Customization has been used in internet-based training, for example, by allowing users to create individualized training plans (Burns et al., 2011; Meyer et al., 2015), by providing users choice about training approaches they want to take, e.g., a CBT approach versus a mindfulness approach (Meyer et al., 2015), and by customizing when an application sends a reminder (Burns et al., 2011). Game-based approaches often increase a user's options to customize: The exergame Zombies, Run![4], for example, features selectable missions with different training intervals and a currency that users can spend to choose rewards. Pokemon Go[5], was a huge success in activating people to walk (Althoff et al., 2016); the game, which focuses on collecting monsters, gives players the option to customize their avatar and incentivizes movement—players can, for example, hatch rare companions from eggs after walking a certain distance. The idea that digital representations (i.e., avatars) can be customized to foster volitional engagement has been explored before (Birk, Atkins, et al., 2016; Trepte & Reinecke, 2010).

[4]https://www.zombiesrungame.com/
[5]http://www.pokemongo.com

Avatar Customization In the context of online games, avatar customization has been shown to increase avatar identification and subsequent enjoyment of a game; furthermore, avatar identification translated into motivated behavior as measured by greater time spent in a free play task (Birk, Atkins, et al., 2016). In the context of internet-based training, we suggest that customization may reduce attrition and increase training efficacy.

4.4.3 Study Methods

We conducted an online experiment in which participants were paid upfront to partake in an online breathing exercise. Participants were either assigned an avatar or were asked to customize their own. After 7 days of self-actuated participation, we sent out daily reminders for 8 days, and stopped sending the reminders for the remaining 4 days. Including the pre- and post-surveys, it was a 3-week experiment. Considering that attrition occurs rather quickly (Eysenbach, 2005), we chose the experiment duration based on previous research on habits, which suggests that small habits, like performing a breathing exercise, are formed in a time frame of approximately 21 days (Lally et al., 2010).

4.4.3.1 Avatar Creator

To manipulate avatar identification, we built a character creator using the Unity 5.4 game engine (Unity Technologies, 2016), Unity Multipurpose Avatar 2 (UMA Steering Group, 2015), built for WebGL. Participants were asked to create an avatar by adjusting the avatar's appearance, personality, and attributes (see Figure 4.1). Participants were required to spend a minimum of four minutes in the character creator, but could take longer if they wished; afterwards a summary of their created avatar was shown.

Appearance: Participants could customize the avatar's sex (2 choices), skin tone (5), hairstyle (4), hair colour (4), eye colour (4), upper clothing (3), lower clothing (3), clothing colours, and shoes (2). Participants could also adjust the avatar's facial features, choosing between 4 different head shapes, i.e., heart shaped, oval, round, squared jaw, and 3 different options for eye-, ear-, lip-, nose size, (i.e., small, medium, large), eye distance (i.e., close, medium, far), and ear orientation (i.e., flat, medium, sticking out). While hair colour (brown, blond, black, and red), and eye colour (blue, green, brown, and green), were limited to one of naturally-appearing discrete choices, clothing colours were picked using a colour picker supporting 16.7 million colours.

Personality: Applying the BFI-10 personality questionnaire (Rammstedt & John, 2007), participants were instructed to state their agreement on a 5-pt Likert-scale (disagree strongly to agree strongly) to ten statements about their character's personality: "I see my avatar as someone who. . . ", e.g., "is reserved", "tends to find fault with others".

Figure 4.1: Avatar Creator.

Attributes: Drawing from the character creation process of role-playing games (e.g., Dungeons & Dragons), participants were required to distribute eighteen points to six different character attributes, i.e., intelligence, stamina, willpower, charisma, dexterity, and strength. Participants were required to assign at least one point in each category.

4.4.3.2 Conditions of Volitional Engagement

To manipulate volitional engagement through avatar identification, half of the participants were asked to customize their avatar using the described avatar creator, whereas the other half were assigned an avatar with generic physical and facial features, plain clothes, and medium hair and skin colour (see Figure 4.2). To avoid unwanted gender mismatch between participant and avatar in the assigned condition, we asked participants "If you were to create an avatar to represent yourself in digital environments, that digital representation would be:" and limited possible responses to male and female. We assigned the generic avatar accordingly. The avatar creator in the customization condition provided the options male and female.

4.4.3.3 Recurring Task: Breathing Exercise

Our goal was to ask participants to engage in a program of self-improvement, but the specific program itself did not matter for the purposes of demonstrating decreased attrition as a result of avatar customization;

we simply needed a believable program of self-improvement in which an avatar could be present. To reduce variability associated with topic interest (e.g., a music training program, such as ableton.com), individual ability (e.g., a language learning app, such as Duolingo), or uncontrollability over the environment (e.g., a step counting program, such as Pedometer that might be subject to inclement weather), we wanted a very simple activity that could be done sitting at a computer and did not require any skill. As such, we asked participants to engage in a daily 1-minute breathing exercise, which has shown positive effects on stress management and relaxation (Wongsuphasawat et al., 2012; Zhu et al., 2017). Participants could freely choose when to engage in the breathing exercise, but were limited to one exercise unit per day to avoid effects resulting from differential usage amounts. The breathing task allowed us to integrate an avatar, was a believable program of self-improvement, did not require skill, and was not subject to uncontrollable factors or individual differences in ability or usage.

The 1-minute long exercise was timed to 6 breath cycles per minute—5 seconds inhaling, 5 seconds exhaling per cycle because this rate has been shown to help with relaxation (Wongsuphasawat et al., 2012; Zhu et al., 2017). To time the exercise and to indicate diaphragmatic expansion for the participant, a circle of variable size was displayed inside a circle of fixed size on the thorax of either the customized or assigned avatar, as has been used in similar breathing interventions (Ibid); see Figure 4.2. Participants were instructed to inhale during expansion of the circle, and to exhale during contraction. Although we could have used the avatar to show animations to pace breathing (e.g., through arm movements or a breathing animation), we did not want to conflate the motivational effects of the customization process with further interaction with the avatar. After an avatar was created or assigned, participants were only exposed to their avatar during the breathing exercise.

4.4.3.4 Measures

We measured situational motivation, state-trait anxiety, Big Five personality, and identification with the avatar using standardized scales.

Situational Motivation was measured using the Situational Motivation Scale (SIMS, Guay et al., 2000). SIMS assesses intrinsic motivation, identified regulation, external regulation, and amotivation, following early studies on Self-Determination Theory (Ryan & Deci, 2002). Using a 7-point scale from 1 "corresponds not at all" to 7 "corresponds exactly", participants rated responses (e.g., "Because this activity is fun") to the question "Why are you currently engaged in this HIT?".

State-Anxiety and Trait-Anxiety were measured using the State-Trait-Anxiety Inventory (STAI, Spielberger et al., 1970). We included STAI immediately before and after the breathing task to enhance the believability of the breathing activity as a relaxation intervention.

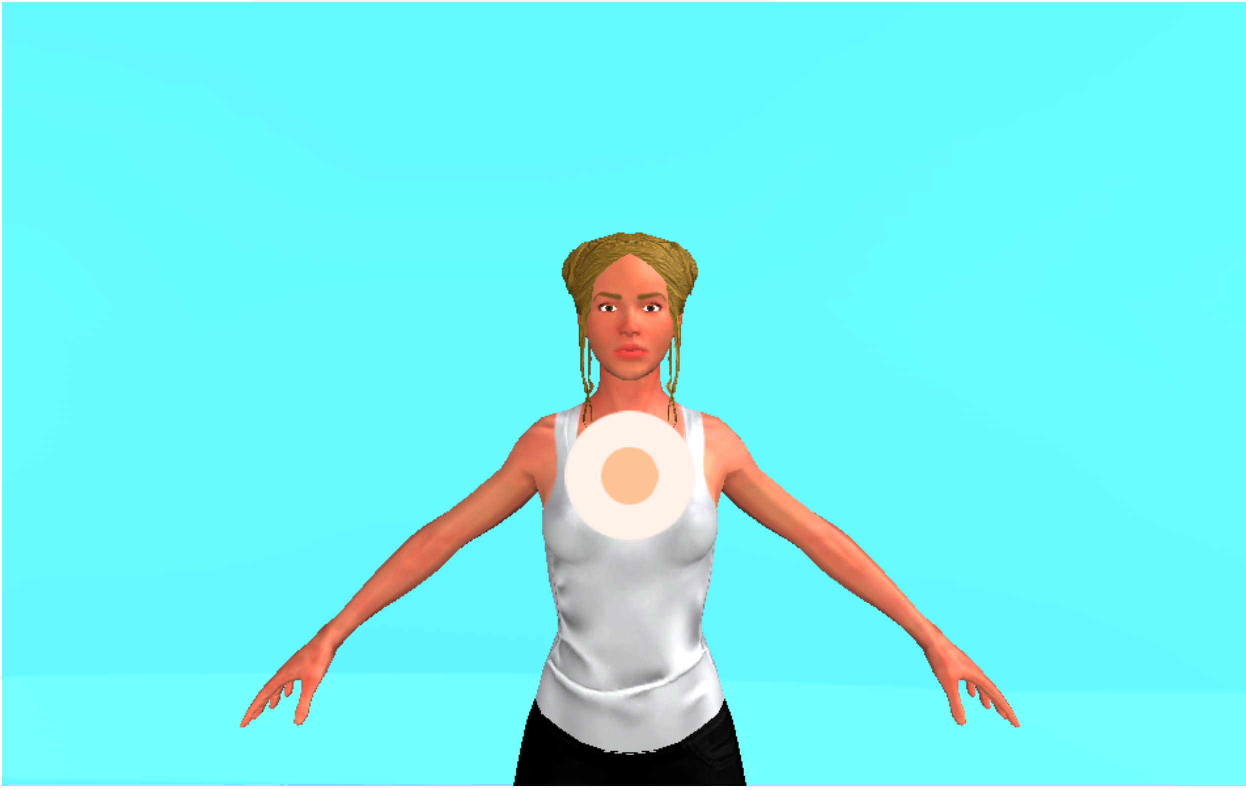

Figure 4.2: Generic Avatar with Breathing Indicator.

Personality Traits were assessed using the Ten-Item-Personality-Inventory (TIPI, Gosling et al., 2003), which assesses the big five personality factors through 10 items that are rated on a 7-point Likert-scale, ranging from 1-"Strongly disagree" to 7-"Strongly agree".

Identification was measured using the avatar-related sub-scales of similarity identification, embodied identification, and wishful identification from the Player Identification Scale (PIS, Van Looy et al., 2012). Participants rated their agreement to identification-related statements—"My character is like me in many ways"—on a 7-pt Likert scale from 1-"strongly disagree" to 7-"strongly agree". Identification has been shown to be important in research on the effects of avatar customization on motivation (Birk, Atkins, et al., 2016; Birk, Mandryk, & Atkins, 2016).

4.4.3.5 Procedure

We recruited participants through Amazon's Mechanical Turk (MTurk). MTurk acts as a broker between parties offering a range of Human Intelligence Tasks (HITs) and paid workers. MTurk has been shown to be reliable as a recruitment tool for research (Kittur et al., 2008; Mason & Suri, 2012). Ethical approval was obtained from behavioural research ethics board at University of Saskatchewan, and participants were asked to give ongoing informed consent at the beginning of each day that they participated. To comply with ethical guidelines, the task was only available to workers from the USA who were older than 18. Only workers with an approval rate above 90% were offered the task as a means of quality control.

The 3-week study was conducted in six phases.

First Phase: Pre-screening (Day 0) In the first phase (Day 0), we recruited 400 participants through a single HIT, advertised as a "3-week daily breathing exercise". Participants were informed that the completion of this first HIT would result in an invite to a second HIT. Participants were asked for demographic information and to complete state scales (i.e., STAI-state, SIMS) and trait scales (i.e., STAI-trait, TIPI). The first HIT neither exposed participants to their avatar—customized or assigned—nor to the breathing task. Within 3 hours, participants received $2 for participating in the 10-minute-long first phase of the study.

Second Phase: Multi-Day Breathing (Day 1) We invited all 400 participants from the first phase back to complete the multi-day breathing exercise, expecting about 1/3 dropout based on our previous experiences with MTurk; 250 (62.5%) returned. Upon providing ongoing consent, we randomly assigned participants to their condition: customized or assigned avatar. All participants started by completing SIMS to measure motivation, followed by either customizing their avatar or viewing their assigned avatar. Participants then completed the identification questionnaire, followed by the breathing exercise, including pre- and post measures of state anxiety. Finally, they were encouraged to "come back tomorrow" and given the URL of the task.

Within 3 hours, we paid participants $10. We chose to pay participants immediately and in advance, regardless of whether or not they participated in the breathing activity. Advance payment was provided to avoid that participant engagement could be attributed to a separable outcome, i.e., daily payments. Furthermore, payment frequency and timing has shown to affect the behaviour of workers (Kaufmann et al., 2011). Participants who completed the daily breathing exercise would spend approximately 60 minutes in total over the 3 weeks; to assure that participants were ethically paid (Silberman, 2016), we paid all participants $10, regardless of daily participation. Random condition assignment was conducted during the first phase, resulting in a study sample of 250 participants. There were 119 people assigned to the customized avatar condition and 131 people assigned to the assigned avatar condition. The differences in group size are simply an artifact of randomly assigning participants to the conditions.

Third Phase: Multi-Day Breathing (Day 2–7) In the third phase (Day 2–7), participants engaged in self-actuated daily breathing exercises. There was no communication with the participants or notifications that they should complete the exercise, as we intended to observe the participants' volitional engagement without any external regulation. The daily exercise only involved the breathing exercise, including pre/post measures of STAI-state. They were not paid explicitly for this phase, as they were paid in advance (see Second Phase). Furthermore, participants were only able to login to the breathing task once per day (starting at midnight central standard time). If they attempted to login again in the same day, they were given instructions to return again tomorrow.

Fourth Phase: Multi-Day Breathing (Day 8–15) In the fourth phase (Day 8–15), we sent out daily notifications via email, reminding participants to engage in the breathing exercise: "Please visit <url> today and complete your daily breathing exercise with your avatar.". We were interested in the differential effect of external regulation (through reminder) on participants who created an avatar and participants who were assigned an avatar. Notifications are common to increase retention in Internet-based training, for example Duolingo[6] notifies users daily to practice. All other aspects were identical to the third phase.

Fifth Phase: Multi-Day Breathing (Day 16–19) In the fifth phase (Day 16–19), notifications were again removed, reverting the task to the version without notifications, as described in phase three. This was done to observe differences in volitional engagement, once external regulation (or at least prompting) was again removed.

Sixth Phase: Exit Survey (Day 20) In the sixth and final phase (Day 20), we sent out an exit survey as a separate HIT to gather information about their motivations and reasons for participating and not participating over the 3 weeks. All participants were invited; 202/250 (80.8%) completed the exit survey. They were asked to complete the SIMS, complete the breathing exercise and the identification questionnaires, plus answer select-all-that-apply questions about why they did and did not do the task (see Table 4.1 for the questions). Participants were paid $2.50 for the task, which took approximately 5–10 minutes.

4.4.3.6 Data Analyses

All survey data and system logins were logged to a database on a server at the University of Saskatchewan and were extracted and analyzed using SPSS 24. Our primary measure of interest was attrition and thus we report primarily on daily participant login counts (Boolean) using non-parametric tests. Motivation and Identification data were analyzed using ANOVA. We used the SPSS Process Macro 2.16 by Hayes (Hayes, 2013) to perform mediated regressions. Mediation analyses tests the influence of a mediator on the effect of a predictor variable (X) on a criterion variable (Y).

Pre-existing Group Differences Because we used random assignment to create our groups, we did not expect them to differ in any systematic way; however, we used one-way ANOVA on pre-existing motivation (measured on Day 2 before the avatar was customized or assigned), conscientiousness, and age, and chi-squared tests on gender to check for group differences.

As Table 4.2 shows, our groups were balanced in gender, age, conscientiousness, amotivation, and external regulation; however, we did find significant pre-existing group differences in intrinsic motivation and identified regulation. We address these pre-existing differences by including intrinsic motivation and identified regulation separately as a covariate in the ANCOVAs (see results).

[6]http://www.duolingo.com

Check all that apply	**Assigned**	**Customized**
. . . check the reasons why you DID participate:	**(True/103)%**	**(True/99)%**
I liked seeing my avatar	9%	16%
I was hoping for an additional bonus	20%	29%
I was bored and looking for something to do	4%	6%
I wanted to meaningfully contribute to the study	**66%**	**62%**
It previously relaxed me and I wanted that feeling again	**56%**	**49%**
I enjoyed the breathing exercise	**57%**	**65%**
It became a habit for me	18%	28%
It was a quick task and did not require much effort	**46%**	**43%**
I'm interested in managing stress and and this could be helpful	**53%**	**61%**
I had been paid in advance	**64%**	**64%**
I felt a responsibility to do it as part of this HIT	**73%**	**72%**
Does not apply	4%	5%
. . . check the reasons why you DID NOT participate:		
I was too busy	**53%**	**48%**
I did not feel like doing it	14%	8%
I forgot about it	**60%**	**60%**
I lost interest	12%	6%
I was not at a computer or other device that supported it	28%	38%
I did not get paid for daily participation	13%	3%
I did not enjoy the breathing exercise	7%	3%
I felt silly doing it	2%	2%
I remembered, but not at the moment I was able to do it	32%	27%
Does not apply	10%	12%

Table 4.1: Percentages of indicated reasons for participating/not participating. Items indicated by more than 40% of participants are displayed bolded and in red.

	Assigned	**Customized**			
	M (SD)	M (SD)	F 1,248	p	η_p^2
Intrinsic Motivation	4.6 (1.62)	5.05 (1.3)	5.7	0.018	0.02
Identified Regulation	4.98 (1.38)	5.49 (1.1)	10.294	0.002	0.04
External Regulation	1.98 (1.09)	2.2 (1.41)	1.802	0.181	0.01
Amotivation	1.96 (1.04)	1.98 (1.09)	2.563	0.111	0.01
Conscientiousness	5.37 (1.5)	5.46 (1.34)	0.527	0.61	0.00
Age	34.06 (10.69)	5.46 (1.34)	1.741	0.188	0.01
	N(Female/Male/Other)	N(Female/Male/Other)	2	p	
Gender	64/65/2	60/58/1	0.285	0.867	

Table 4.2: Test statistics for pre-existing Situational Motivation, conscientiousness, age, and gender.

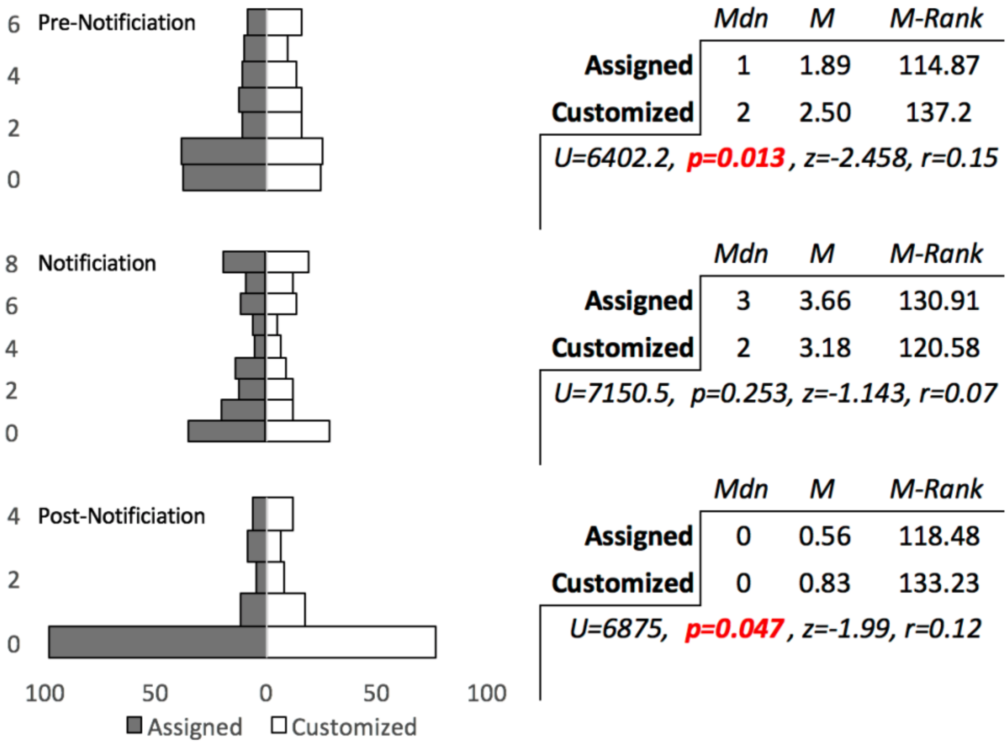

Figure 4.3: Left: Mirrored histograms of login count for the three notification phases (pre-notification, notification, and post-notification) by avatar condition (assigned, customized). The X-axis shows the frequency of login amount bins; the Y-axis the number of logins as a category. Right: Mann-Whitney U test statistics for each notification phase by avatar condition. Significant results are bolded and in red.

4.4.3.7 Research Questions

We ask the following research questions:

RQ1: Does customizing an avatar combat attrition?

RQ2: Does customization affect avatar identification?

RQ3: Does customizing an avatar affect situational intrinsic or extrinsic motivation over time?

RQ4: Are the differences in logins explained by increased identification (controlling for pre-existing motivation)?

4.4.4 Results

We present our results by research question.

4.4.4.1 RQ1. Does customizing an avatar combat attrition?

Our first question is whether or not customizing an avatar can help to combat attrition. Figure 4.3 shows mirrored histograms of login count for the three notification phases (i.e., pre-notification, notification, and

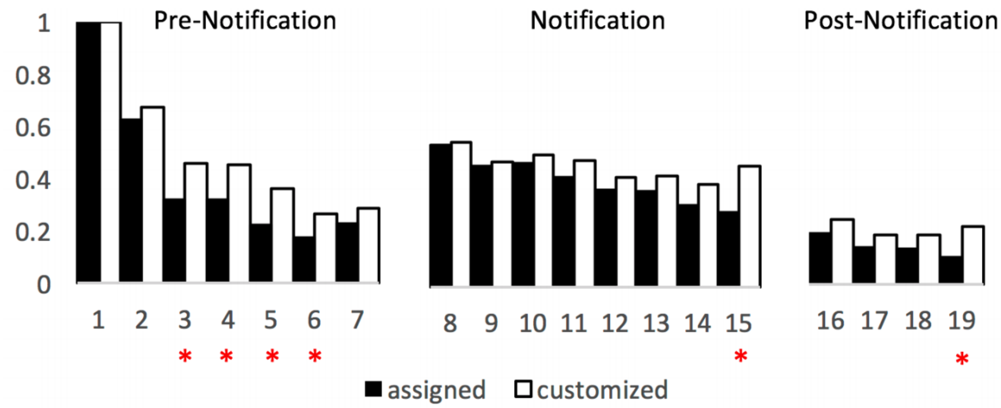

Figure 4.4: Logins by Day, normalized by starting group size, and partitioned by phase: Pre-Notification, Notification, and Post-Notification. *= significant differences by day ($p < .05$).

post-notification) by avatar condition (i.e., customized or assigned). Figure 4.4 shows the counts of logins for each day as a percent of the group size (i.e., customized or assigned). A Mann-Whitney U test of login count by participant confirms our assumption that participants who customized ($Mdn = 7, M = 8$) their avatar showed more logins over all Days (1–19) compared to participants with assigned avatars ($Mdn = 5, M = 6.63$), $U = 6752, p = 0.035, z = -1.834$.

To investigate the effect of notification, we compared the number of logins during each phase (pre-notification: Day 1–7, notification: Day 8–16, post-notification: Day 17–19) between participants who customized an avatar and those who participated with an assigned avatar. The results suggest that in the pre-notification and post-notification phases, participants who customized an avatar had a higher overall login count compared to participants who were assigned an avatar. See Figure 4.4 for results.

These results suggest that participants who customized an avatar showed significantly less attrition overall and that this difference was driven by the phases in which there were no notification reminders. We were further interested in not just the consistency of logging in, but also how login patterns changed over time. Figure 4.4 shows the actual login counts per day per condition (normalized by the size of the starting groups as seen in the equal bars for Day 1). First, it is clear that there is attrition over time in both conditions. Second, it is clear that there is more participation each day by people in the customized avatar group. Comparing daily login rates by avatar group using a Chi-squared test reveals that during pre-notification, the customization group has higher login rates on Day 3 ($p = .011$), Day 4 ($p = .015$), Day 5 ($p = .007$), and Day 6 ($p = .037$) compared to the assigned group. Once notified, the groups initially equalize, but descriptively diverge again over time, and by Day 15, the difference is again significant ($p < .001$). Once notifications were removed, the customized group still shows higher login rates and the difference is significant on Day 19 ($p < .001$). Taken together, our results show that people with a customized avatar showed less attrition than people with an assigned avatar, evidenced through greater overall login rates—especially when no notifications were given—and through greater numbers of daily logins.

Ruling out Pre-existing Group Differences in Motivation as an Explanation of Login Behaviour
We observed significant pre-existing group differences in intrinsic motivation and identified regulation, but not external regulation or amotivation; see Table 4.2. Because the customized avatar group had higher incoming intrinsic motivation and identified regulation, it is possible that their pre-existing motivation explains or accounts for their significantly higher login count. To test whether incoming motivation explains group difference in logins, we conducted univariate ANOVAs on total logins for each phase (and overall), followed by separate ANCOVAs with incoming intrinsic motivation and incoming identified regulation separately. We did not use ANCOVAs to look for group differences originally because the underlying data is comprised of login counts over time; however, comparing the F-statistic for the ANOVA to that of the ANCOVAs will allow us to determine whether the variance in the group differences can be explained by pre-existing motivation. As shown in Table 4.3, the F-statistic does not decrease when either intrinsic motivation or identified regulation are included in the models. In fact, the F-statistic increases in all cases, showing that controlling for the variance of incoming motivation makes the login differences between conditions even more clear.

4.4.4.2 RQ2. Does customization affect avatar identification?

Previous work showed that avatar customization fosters intrinsic motivation in part by facilitating identification with the avatar (Birk, Atkins, et al., 2016; Trepte & Reinecke, 2010). As such, our second question asks whether identification with an avatar is more prevalent in the customized condition and whether customization affects changes in identification over time.

We investigated the effect of avatar condition on identification by comparing identification on Day 1 to identification on Day 20 (i.e., exit survey). The data were analyzed using a repeated measures ANCOVA, controlling for an individual participant's total number of logins to account for the fact that people had different exposure to their assigned or customized avatar over time based on their participation.

Our results reveal a main effect for avatar condition, showing that avatar customization increases similarity identification ($F_{1,198} = 81.1, p < .001, \eta_p^2 = .29$), embodied identification ($F_{1,198} = 36.8, p < .001, \eta_p^2 = .15$), and wishful identification ($F_{1,198} = 37.8, p < .001, \eta_p^2 = .16$). A main effect of Day on similarity ($F_{1,198} = 13.1, p < .001, \eta_p^2 = .06$) and embodied identification ($F_{1,198} = 5.5, p = .020, \eta_p^2 = .03$), shows that identification wanes over time; both are lower on Day 20 than Day 1.

The interaction between survey day (Day 1, Day 20) and avatar condition (assigned, customized) shows a significant interaction on similarity identification ($F_{1,198} = 11.8, p = .001, \eta_p^2 = .06$), and wishful identification ($F_{1,198} = 4.5, p = .034, \eta_p^2 = .02$). For both types of identification, pairwise comparisons show that identification was higher on Day 1 than Day 20 for the customized group, but that identification was stable over days for the assigned group.

Taken together, our results show that customizing an avatar facilitates identification of all types, but that the initial similarity and wishful identification measured right after customization drops after 20 days of a breathing activity; however, identification on Day 20 was still higher for those with a customized avatar than

	Pre-Notification		Notification		Post-notification	
Covariate Included	F	p	F	p	F	p
None	5.915	0.016	1.596	0.208	3.063	0.081
Intrinsic Motivation	6.578	0.011	1.806	0.180	4.276	0.040
Identified Regulation	6.405	0.012	1.872	0.172	3.973	0.047

Table 4.3: F-statistic and p-value for the three notification phases with "None", Intrinsic Motivation, and Identified Regulation, seperately included as covariates.

	Assigned	Customized	Condition		Control	
	M (SD)	M (SD)	F	p	F	p
Intrinsic Motivation	4.15 (1.76)	4.65 (1.54)	0.90	0.34	166.85	0.00
Identified Regulation	4.74 (1.55)	4.95 (1.39)	2.27	0.13	182.37	0.00
External Regulation	2.34 (1.29)	2.71 (1.48)	1.29	0.26	63.51	0.00
Amotivation	2.13 (1.31)	2.01 (1.27)	0.07	0.79	38.98	0.00

Table 4.4: Test statistics for Condition on SIMS subscales on Day 20, controlled for values on Day 1.

	Similarity Identification					
	X → Y	**X → M**	**M → Y**	**Total**	**Direct**	**Indirect**
	p	p	p	p	p	p
Pre-Notification	0.011	<.001	0.626	0.013	0.011	0.352
Notification	0.436	<.001	0.554	0.162	0.436	0.555
Post-Notification	<.001	<.001	0.001	0.043	<.001	0.002

X: Condition, Y: Login Count, M: Similarity Identification

Control: Identified Regulation

Table 4.5: Mediation results for Avatar Condition on Login Count by Day mediated by Similarity Identification and controlled for identified regulation at Day 1.

an assigned one, even when accounting for the additional exposure through login count.

4.4.4.3 RQ3. Does customizing an avatar affect situational intrinsic or extrinsic motivation over time?

Previous work showed that customizing an avatar helps facilitate intrinsic motivation (Birk, Atkins, et al., 2016) and keeps subjective motivation from waning as much as when the avatar is assigned (Birk, Mandryk, & Atkins, 2016). As such, our third research question was whether customizing an avatar affects situational intrinsic or extrinsic motivation, and whether this changes over time. We conducted separate ANCOVAs on the four SIMS subscales (i.e., intrinsic motivation, identified regulation, external regulation, and amotivation) by avatar condition (i.e., assigned, customized), with Day 1 values as a covariate for each subscale. See Table 4.4 for results.

The results suggest that customizing an avatar has no effect on how motivating the task is perceived to be in terms of actually enjoying it (i.e., intrinsic motivation), feeling like it is something that a participant feels would be good for them to do (i.e., identified regulation), feeling more responsible to engage in the task daily from external regulatory pressure (i.e., external regulation), or feeling not motivated at all (i.e., amotivation). Taken together, these results show that customizing an avatar does not affect self-reported situational intrinsic or extrinsic motivation after 20 days.

4.4.4.4 RQ4. Are the logins differences explained by increased identification (controlling for pre-existing motivation)?

Previous work has shown that self-reported motivation and time spent on a task can be explained by identification with an avatar (Birk, Atkins, et al., 2016). As such, our third research question connects the findings from RQ1–RQ3 by considering the mediating effect of identification on how customization affects participation during the different notification phases.

We conducted mediation regression (Hayes, 2013) analysis (model=4) of avatar condition (i.e., assigned, customized) on login count ($X \rightarrow Y$) mediated by similarity identification ($X \rightarrow M, M \rightarrow Y$), separately for each notification phase, and controlled by identified regulation. We then repeated this analysis, also controlling for intrinsic motivation. We only present results for identified regulation, because the outcome is the same pattern when controlling for intrinsic motivation. See Table 4.5 for results.

Our results show that, as expected based on RQ1, customization affects login count in the pre-notification and post-notification phases, but not in the notification phase. Customization predicts similarity identification in all three phases (RQ2). However, similarity identification only predicts login count in the post-notification phase and not in the pre-notification or notification phases. In these two initial phases, including similarity identification in the model does not explain additional variance and the effect of avatar customization on logins is not explained by similarity on Day 1. However, in the post-notification phase, there is a significant indirect effect, indicating a partial mediation of similarity identification on the relationship between condition and login count. The direct effect stays.

We repeated this process for embodied identification and wishful identification; however, those mediations were all non-significant, showing that neither embodied nor wishful identification on Day 1 explains the variance of how customizing an avatar increases login count.

4.4.5 Discussion

We summarize our findings and discuss why customizing an avatar increases login behaviour in a daily breathing exercise, how to interpret these effects in the light of payment, and how employing existing and novel techniques in combination can help to combat attrition. Finally, we discuss how our results have practical significance in the context of self-improvement programs.

4.4.5.1 Summary of Findings

Our results show that participants who customized an avatar showed less attrition over time. Additionally, we investigated the effect of notification on attrition and found that login patterns changed during the notification phase: we show that during the pre- and post-notification phase, participants who customized their avatar logged in more; while during the notification phase, the customization and the assigned group showed no differences. Further unpacking the notification phase, we show that the customization group logs in more, these differences grow in magnitude over time, and they become significant by Day 15. Our findings show that people with a customized avatar showed less attrition overall—especially when no notifications were given; furthermore, these findings hold when controlling for individual differences in incoming motivation.

Considering the role of identification—an important factor for understanding avatar customization-induced increases in motivation and behaviour (Birk, Atkins, et al., 2016; Birk, Mandryk, & Atkins, 2016)—we replicate previous findings showing that customization increases identification with an avatar (Birk, Atkins, et al., 2016; Trepte & Reinecke, 2010). However, we additionally show in this study that identification wanes over time in the context of our task. Further, we performed mediation analysis and confirm that although customization predicts increased login behaviour, this relationship is not mediated by the role of similarity identification. Finally, we investigated how customization affects situational motivation on the final day, and show no significant differences between intrinsic or extrinsic components of motivation, showing that while participants exhibit an increase in motivated behaviour, they subjectively don't express higher levels of intrinsic motivation or identified regulation.

4.4.5.2 Customization Increases Adherence

Creating an avatar—as opposed to having an assigned generic avatar—shows a group effect on login behaviour and can be interpreted as causal. The change in behaviour is a direct effect of customization and can neither be explained by increases in identification, nor in situational motivation, as our mediation results show (RQ4).

Having choice, i.e., being allowed to customize, has been linked to increased motivation, persistence, performance, and satisfaction before (Patall et al., 2008). Furthermore, previous work has shown that avatar customization increases invested effort and behaviour in the short-term (Birk, Atkins, et al., 2016).

While the satisfaction of needs is an important predictor of intrinsic motivation, in the context of avatar creation, a sense of ownership over the avatar provides an additional explanation for increases in observed behaviour. Previous work has shown that being the creator of a product, e.g., designing a t-shirt, leads to a sense of psychological ownership (Pierce et al., 2003), becomes a part of the consumer's extended self (Mittal, 2006), and affects behaviour, as displayed in an increased willingness to pay (Franke et al., 2009). In our studies, participants who customized their avatar also expressed greater agreement with "I liked seeing my avatar" as motivation for completing the breathing task (assigned: 9%, customized: 16%). These findings from marketing psychology highlight how customization is not a trivial process, but deeply affects how we feel about an object—physical or digital (Trepte & Reinecke, 2010)—and how the sense of ownership over a digital artifact can affect our behaviour on a fundamental level. We are willing to invest more: more time, more effort, more money.

The willingness to invest is crucial when we consider that self-improvement programs rely on the user's volitional decision to engage with them; while there are external cues, e.g., notifications, they can only nudge the user to engage or lower the bar for engagement, for example, by providing short units of training that are accessible in free moments.

Role of Payment A common challenge in motivation research is how to account for the effects of incentives on behaviour—how does the compensation to participate in a study affect a participant's motivation and behaviour? In-the-wild studies can avoid this issue, but in experimental research, participants are often incentivized with cash or course credit due to ethical considerations. Previous research has demonstrated the effects of incentives on motivated behaviour (Deci et al., 1999; Kaufmann et al., 2011), including effects as subtle as how a promised payment being delivered can affect motivation (Birk, Mandryk, & Atkins, 2016). We intentionally paid participants in advance for the entire 3 weeks ($10 for asking them to participate for 2 minutes per day plus the 10-minute first day), to avoid effects of payment on motivation (Birk, Atkins, et al., 2016) and to rule out effects of external regulation from payment for their participation (Birk, Mandryk, & Atkins, 2016). When considering participants' reasons for not completing the breathing exercise (see Table 4.1), we find that participants in the assigned group provided "I did not get paid for daily participation" more often as a reason for not doing the task (assigned: 13%, customized: 3%), whereas more participants in the customized group expressed that they completed the breathing exercise because they were "...hoping for an additional bonus" (assigned: 20%, customized 29%). However, paying participants in advance at ethical but reasonable rates appears to have eliminated payment as a mitigating factor of participation; overall, financial reasons for participating are low, and most participants acknowledge that they got paid in advance (both: 64%).

The most prevalent responses for not completing the breathing exercise in both conditions were "I was too busy" and "I forgot about it"—the latter clearly explains why notifications after Day 7 were successful at increasing logins.

Role of Notifications Notifying participants on Day 8 almost doubled participation compared to Day 7 (from about 30% to almost 60%), showing the effectiveness of an established technique in combating attrition. We further saw that notifications initially equalized the groups in terms of participation, but that as the efficacy of the notification wore off over days, attrition was greater for the assigned avatar group—by the end of the notification phase, the participation rates between groups were once again significant.

Timed notifications are effective for tasks that are easily forgotten (e.g., doing a single brief exercise per day), but are not well suited to address other reasons that underlie attrition, such as that the task itself is not enjoyable, or that a user can't attend to a program at the moment of reminder. Tailored approaches such as personalized or context-aware reminders (Etter et al., 2006; Horvitz et al., 1999) can help to increase the efficacy of reminders. While notifications are a generalizable technique to combat attrition, other known techniques such as social support (Kelders et al., 2012) or in-app rewards (Richter et al., 2015) may also benefit from being employed in combination with customization.

Combining Multiple Approaches We can study techniques for combating attrition in isolation or in combination. We chose to both investigate avatar customization in isolation to characterize its efficacy in combating attrition, and also how it works in combination with a prompting reminder—effectively layering different approaches for combating attrition into a single intervention. In addition to existing approaches that could be combined with avatar customization (e.g., notifications, rewards), the use of an avatar opens opportunities to leverage experienced attachment with the avatar as motivation to engage.

For example, if we consider the importance of identification (Birk, Atkins, et al., 2016; Van Looy et al., 2012; Trepte & Reinecke, 2010), focusing on creating memorable experiences shared between the avatar and the user could enhance engagement. For example, characters in World of Warcraft have been shown to be valued by players for being a collection of memorable experiences, e.g., defeating a difficult boss (Livingston et al., 2014), and leveraging this attachment has the potential to further increase the effectiveness of avatar customization on adherence. Training programs for self-improvement could foster interaction between the user and the avatar and also interaction with an application through the avatar.

Considering different layered or isolated techniques of combating attrition, developers of self-improvement programs need to make an informed decision about, which features to integrate and, which features to ignore.

4.4.5.3 Practical Significance

Our results show small but significant differences in adherence as a result of avatar customization. However, small effects have big practical significance in our context of combating attrition in digital programs of self-improvement delivered at scale. Even a small increase in a large sample affects many individual people. The example of MoodGYM demonstrates how an intervention tested in-the-wild saw only 97 (0.5%) of 19607 people completing the exercises. An increase of only 1% would increase the number of participants who might be positively affected by the intervention by 196 participants—in our study, we see closer to 10% improvement in logins as a result of customization. Increasing adherence for thousands of people using a self-improvement program has real potential to change lives.

4.4.6 Conclusion

Digital programs delivered at scale have the potential to increase access to support for people interested in self-improvement in contexts as diverse as behaviour change, lifelong learning, and mental wellness. However, even the best-designed programs won't work if people don't use them. Research in a variety of contexts has shown that digital programs that were effective in an RCT are subject to high attrition when delivered in-the-wild, i.e., at scale. We present a general-purpose interaction technique for combating attrition—avatar customization—and show that it significantly improves retention over three weeks of a daily stress-relief intervention (a breathing exercise) over being assigned a generic avatar. Given the broad accessibility and wide reach of programs delivered at scale, increasing adherence through interaction design has the potential to benefit a great number of people working at self-improvement.

4.5 Summary of Manuscript C

The results show that customizing an avatar decreases attrition in a non-game-based training exercise, especially pre- and post-notification. The notification phase shows a strong effect on login behaviour; however, even during the notification phase, a tendency emerges over time, suggesting that participants who customized their avatar show less attrition under notifications compared to those who participated with an assigned avatar. Considering my previous findings (see Chapter 2, and Chapter 3), I investigated the effects of customization on identification (RQ2) and motivation (RQ3). The results suggest that identification declines over time but is in general higher for those who customized their avatar. Asking if identification explains login differences, my results show that identification doesn't explain login differences. Further, there were no self-reported differences of motivation between avatar groups. The results primarily show the effects of avatar customization on login behaviour and investigate constructs that have shown to be of relevance in previous work (see Chapter 2, and Chapter 3). While the research results are in themselves of importance to contribute to our understanding of how game-based interface manipulations can be used to combat attrition in technology-enabled services for health, there are important lessons I learned from the research process and the results presented in Chapter 2, and Chapter 3 that I applied in the study presented in this Chapter. Considering previous lessons enabled me to refine my research protocols and improve the research quality by, for example, being considerate about payment procedures.

4.5.1 Lessons Learned

The lessons learned in this study are a product of adjusting my research based on the lessons learned in previous studies (see Chapter 2, and Chapter 3). Applying previous experiences, I was able to 1) adapt avatar customization, resulting in group differences between the customization and the assignment group. Group differences are crucial, because understanding causality is important to build a reliable research foundation of the effects of avatar customization on player experience. For example, understanding causality is relevant when leveraging avatar customization as a means to increase participation over time in an application used in a clinical context.

2) I was able to minimize the potential effects of payment delay and payment expectations on the results by pre-paying participants. This practice greatly improved the external validity of the results and allows me to perceive the results in the light of an actual versus an aspirational improvement to digital health applications (Torous & Firth, 2018) and problems around attrition (Eysenbach, 2005; Christensen & Mackinnon, 2006).

3) I already had experiences with workers in medium-term studies (see Chapter 3), but during this study, I gained further insights into worker behaviour: Turkers are self-organized and engage in tasks even without notifications, which points to a great level of commitment in general, but also suggests that there is need to explore long-time Turker behaviour in more detail, e.g., are there other forms of instructions to assure higher rates of worker engagement or did a subset use calendar notifications outside of the study protocol?

4) The results show that participants engaged differently, but SIMS could not detect any differences between groups. The lack of a group difference suggests a dissonance between behaviour and conscious experience. The dissonance might be a function of the subtlety of the manipulation and participants' perception of this simple daily task in general. Some of the "check all that apply" questions shed some light on more subtle group differences that accumulated and might have driven my findings, e.g., participants who customized liked their avatar better, were quicker to form habits, provided fewer reasons why they didn't participate, and stated less often that they lost interest. SIMS might also simply not be sensitive enough to detect the subtle changes in subjective motivation that avatar customization induced.

5) The "check all that apply" questions were a good tool to investigate additional explanation in a quick manner. We do not know how reliable group differences indicated by our "check all that apply" questions are, but the responses are useful to support other points when constructing an argument, have potential to inform research impulses, and reveal patterns that would justify further investigation—it is for example remarkable that 10% more participants with an assigned avatar state that they did not get paid for daily participation, which—having ruled out differences in the delivery of the study protocol—suggests that participants are less engaged when reading the instructions or as a result of decreased participation they are more likely to justify their behaviour.

In summary, applying previously-learned lessons around constructing control groups, participant payment, worker behaviour, measuring motivation, and understanding participant behaviour in context, we further deepened our understanding of how we can utilize game-based motivational design strategies to improve the implementation of technology-enabled services for health.

4.6 Relevance in Context

While Chapter 2 provided a foundation to understand the effects of identification with a representation on experience and behaviour in a game-based context, and Chapter 3 showed the feasibility of investigating the effects of identification in a game-based context in the medium term and highlighted how important it is to consider incentives, the study presented in this Chapter demonstrated that game-based design elements can be applied in a non-game-based context and affect login behaviour over a three-week long study. Considering the aims of my thesis, the results show that avatar customization fosters intrinsic motivation with a non-game-based health training over time and facilitates sustained engagement.

Chapter 5

Improving Cognitive Training Efficacy

Citation: Birk, M. V., & Mandryk, R. L. (in press). How Customization Improves the Efficacy of an Online Digital Intervention for Mental Health. *Journal of Medical Internet Research.*

Acknowledgment: Under my direction, Benj Hingston implemented the ABMT task and modified the avatar customizer described in Manuscript B.
In coordination with my supervisor, Dr. Regan Mandryk, I was responsible for the research direction, experimental design, system design, data gathering, data analysis, and reporting.

Whereas Chapters 2 to 4 produced evidence that game-based motivational design strategies facilitate sustainable engagement by investigating the effects of avatar customization on motivated behaviour in the short-term and medium-term, Chapter 5 applies the learned to modify a computerized cognitive therapy. In this Chapter, I investigate the effect of avatar customization on immediate task efficacy, in particular on increasing the efficacy of an established attentional retraining task—i.e., the Attention Bias Modification Training (ABMT) (Hakamata et al., 2010).

I conducted a study with 317 participants on Amazon Mechanical Turk. Following a 2x2 between-subject design plan, participants were randomly assigned an avatar, or they customized their own avatar before they were assigned either to an attentional bias retraining condition or a control condition. I measured state-trait anxiety before the training condition and measured state-trait anxiety again after participants were exposed to a Negative Mood Induction using images with strong negative valance—ABMT has been shown before to increase resilience to negative mood induction (Lang et al., 2008; Hakamata et al., 2010). The results suggested that avatar customization increases engagement in the moment and positively affected task resilience for those who were exposed to the attentional retraining. Additionally, the results showed that participants who customized their avatar and did not receive any training experienced increased state anxiety after a negative mood induction when controlling for baseline state-anxiety, further supporting that the customization facilitates in-the-moment engagement.

5.1 Problem and Motivation

The primary focus of the previous Chapters has been on designing for training efficacy by increasing sustained engagement over time. In Chapter 2, avatar customization was investigated as an interface design technique to increase engagement in the short-term and in Chapter 4, avatar customization was applied to increase participation in a 3-week training exercise. However, neither of these studies investigated the immediate effect of avatar customization on task-efficacy, i.e., the positive effect of increased engagement in the moment in terms of training outcome.

Video games have shown to increase in-the-moment engagement, expressed, for example, as increased attention (Green & Bavelier, 2006), being mentally deeply involved (Jennett et al., 2008), or having a distorted temporal experience (Sweetser & Wyeth, 2005). Research suggests that video games provide environments that facilitate deeply engaging experiences (Rigby & Ryan, 2011), but researchers don't know if avatar customization applied in a computerized cognitive training (CCT) can positively affect the efficacy of training.

Therefore, the problem I addressed in this Manuscript is that avatar customization shows promising effects on engagement in the short-term and long-term, but the effects of avatar customization on immediate training efficacy remain unclear.

5.2 Solution and Steps to Solution

To investigate the effects of avatar customization on immediate training efficacy, I conducted an online experiment on Amazon Mechanical Turk; 317 participants were split randomly split in two avatar groups (i.e., participants either customized their avatar, or were assigned a neutral male or female avatar), each group was then assigned to a training group (i.e., participants were either exposed to a cognitive retraining or a control condition) before they were exposed to a negative mood induction that allowed me to measure post-induction state-anxiety.

While previous projects provided valuable insights about the effects of avatar customization on motivation, effects on behaviour, and Amazon Mechanical Turk workers as a participant pool, I had little experience with increasing task engagement in the moment of use and, as a result, task efficacy.

Previous research on video games suggests that games facilitate in-the-moment engagement (Green & Bavelier, 2006; Jennett et al., 2008; Sweetser & Wyeth, 2005). Research on skill acquisition shows the crucial importance of attention when learning a new skill (Grossberg, 1999). Considering both bodies of literature and my previous research, I was confident that integrating a game-based design strategy, such as avatar customization, into a computerized cognitive training would improve training efficacy.

Considering that the evaluation of game-based design strategies to increase the efficacy of computerized tasks is a novel approach and at the time of the research I had limited experiences with studying efficacy, I decided to conduct a controlled experiment prioritizing internal validity by using a well-known computerized cognitive training paradigm (i.e., ABMT) and a rigid experimental plan that implemented design decisions from previous research (Hakamata et al., 2010), while maximizing external validity by conducting the experiment online. A laboratory study would have had several limitations when extending my findings to computerized online training, e.g., external control in lab study, a a selective sample, and in a constrained situation that doesn't occur naturally.

ABMT is a simple cognitive retraining, which in the unmodified version relies on having participants look simultaneously at a picture of a person with a neutral facial expression and a picture of the same person with an angry facial expression. Participants are exposed to 720 pictures sets in total. After one image set has been presented for 500 milliseconds, an arrow pointing either to the left or to the right is presented at the same position as one of the images—whether the arrow is presented only behind the neutral image or both images equally defines the training and control condition. In both cases, participants need to press the corresponding arrow key on their keyboard to proceed. In the training condition, participants' attention is retrained by learning that the arrow key is behind the neutral images and therefore that they need to preferentially focus on the neutral image for best performance. In the control condition, the arrow is presented equally behind the neutral and the angry images. For a detailed description of the task, see Section 5.4.4.2.

5.2.1 Research Questions

To investigate the effects of avatar customization on training efficacy, I investigated four research questions:

RQ1. Does customization improve the efficacy of attentional retraining?

RQ2. Does customization increase identification for trained and non-trained participants?

RQ3. Does avatar identification increase the efficacy of attentional training?

RQ4. Does training efficacy vary depending on basic needs satisfaction?

5.3 Design Rationale

Moving away from the effects of avatar customization on engagement over time and towards in-the-moment engagement spawned several challenges. Many computerized cognitive training tasks are developed for research purposes and research usually is focused on investigating the efficacy of the training; as a side-effect, the design is often very clean to avoid effects of uncontrolled stimuli such as coloured interface elements or progress indicators. As a result, most CCTs are not optimized for engagement. Hence, there are several open design questions. In the manuscript presented in this chapter, I aim to address questions around the effects of game-based design on task engagement by manipulating a well-known training task, i.e., ABMT. However, there are still open questions that I would be a in better position to address in retrospect than I was when designing this particular project. For example, questions around how to operationalize in-the-moment engagement, or considerations around online induction methods in general and specifically around necessary considerations when inducing negative mood online.

5.3.1 Task Choice

When looking for a task that would be suitable to evaluate the effects of avatar customization on task efficacy, there were several requirements: 1) My experiences with the effects of avatar customization on motivation and engagement up to this point stemmed from studies with embedded avatars, i.e., avatars that were presented as part of a game (see Manuscript A and B) or within a task (see Manuscript C). Hence, the task needed to be laid out and structured in such a manner that it would be possible to embed an avatar into the task. 2) That task needed to be well-established to prevent introducing design flaws and for an interpretation of the results in context. 3) To further build on my previous experiences, and to address the overarching goal of this dissertation, the task needed to be adaptable to an online context. 4) Another prerequisite was to use a clearly-formulated task protocol to assure that implementation differences were not a source of potential effects. 5) Previous research on the task should have shown results for the general population and not just for a special population, e.g., ADHD, to assure broader applicability of my findings. 6) The ideal task would have also been evaluated using adapted interface components with the aim to maximize external validity. Knowing that a cognitive training task has already been applied in a research context that is more ecologically valid would suggest that the training effects of a task are revealed independent of an interface manipulation.

I chose ABMT, because it fulfilled these criteria: 1) ABMT allows for the integration of an avatar, 2) is well-established (Hakamata et al., 2010), and 3) has been used online before (Carlbring et al., 2012). 4) The protocol is clearly formulated and allows for an exact implementation, 5) previous research was not primarily conducted with special populations, 6) and the ABMT interface has been successfully adapted before—see Dennis and O'Toole (2014), who presented the ABMT using a touch interface with a gamified version of the task.

In addition to fulfilling my requirements, ABMT has been used as a training approach to reduce anxiety. One line of my research is focused on game-based design for mental health. I have, for example, investigated player habits and preferences in an online context and showed that people with depressive symptoms play games (Mandryk & Birk, 2017), which suggests that game-based assessment and treatment might be an interesting and feasible research direction. Investigating a task that has been used in the context of mental health was a deliberate choice, because it produced insights and helped me to gain experiences that could provide advantages when moving forward on applying game-based design elements in the context of mental health—a central component of my future research agenda.

One limitation of ABMT was the repetitiveness of the task—participants have to click the left and right arrows after the presentation of negative or neutral pictures 720 times in total. Most cognitive training tasks rely on repetition, which in the lab is acceptable, but I was concerned about participants' behaviour online, i.e., if they get bored after a while and took breaks. However, I deliberately gave participants short breaks in between sets as described in the manuscript and felt that it would be worth taking the risk, especially because engagement could be measured through participants' response times in the task to determine whether or not the took frequent or long breaks.

5.3.2 Avatar Customization

Similar to the approach described in Chapter C, participants either customized their avatar or were assigned a neutral avatar before engaging with the task. While the avatar assignment approach was comparable to the assignment in previous tasks, the presentation was slightly different. In previous tasks, the avatar was dynamically embedded in the task and provided participants either with agency in the game (Manuscript A and B) or participants followed the avatar's guidance, i.e., in Manuscript C the breathing exercise is paced by the avatar. In the study presented in this Manuscript, the avatar is presented statically, without any agency, and there are no animations, except for within the customization process. The avatar only varies in terms of its facial expression: neutral or angry. Compared to all other presented avatar studies, the manipulation in this particular study is the most subtle, which has implications for potential results, i.e., small effects derived from such a subtle manipulation would support the reliability of the induction and further support the relevance of avatar customization as an interface manipulation technique relevant for the improvement of technology-enabled services.

5.3.3 Validated Measures

By including the Player Identification Scale (De Grove et al., 2017), the Basic Psychological Need Satisfaction Scale (Kasser et al., 1992), and the State-Trait-Anxiety Inventory (Spielberger et al., 1970), I chose several well evaluated and validated measures.

5.3.3.1 Player Identification Scale

The measure for identification by De Grove et al. (2017) has been used in all previous manuscripts and was included to allow for the investigation of identification as a mediator of increased task efficacy. My prior research experience has shown the potential relevance of identification and because I had little experience with applying avatar customization as a game-design element to increase training efficacy, I decided to include identification as potentially explanatory measure.

5.3.3.2 Basic Psychological Need Satisfaction

Following SDT as described in Chapter A, I included a measure of need satisfaction. Previous measures either investigated experienced need satisfaction as a function of playing a video game. In this study, however, I was interested in the general need satisfaction people experience in their life. My reasoning was two-fold: 1) experiencing low need satisfaction in life is a predictor for increased vulnerability (Vansteenkiste & Ryan, 2013); 2) people can be characterized by their need satisfaction: a) is their need for autonomy satisfied in life, e.g., do they experience agency over their choices. b) do they feel connected to others, an important predictor of mental health (Baumeister & Leary, 1995; Ryan & Deci, 2002); or c) are they feeling competent and therefore mastery over the activities they regularly engage in, e.g., their job. Low need satisfaction might be an indicator of high vulnerability and I decided to include a measure of need satisfaction to potentially further explain individual differences of training efficacy. Measuring and understanding individual differences and responses to training are important contributions, because a precise fit between an individual's needs and the design of a task has huge potential to increase task efficacy further and addresses drivers of individual motivation by design.

5.3.3.3 State-Trait-Anxiety Inventory

The State-Trait-Anxiety Inventory is commonly used to measure a person's general tendency to experience anxiety (trait) and short-term anxiety responses to a situation or a set of stimuli (state). Because state anxiety is a situational response measure, it allows for the measurement of short-term changes in anxiety, e.g., after a negative mood induction, allowing researchers to infer the effect of interventions geared to increase resilience against an anxiety-inducing situation. As such, the state-scale of STAI was well suited to evaluate potential effects of avatar customization on training. The scale has also been used before in ABMT research.

5.3.3.4 Negative Mood Induction

While all studies that I have been a co-author of had ethical approval from the ethics board of the University of Saskatchewan, the induction of negative moods online requires specific considerations, which I believe are worth highlighting in the context of my dissertation:

First, measuring changes in resilience requires a difficult or challenging situation. Because ABMT is supposed to increase resilience to negative emotional situations, a negatively-valanced induction is required. Second, I used gruesome imagery from the International Affective Picture System (IAPS) (Lang et al., 2008), which are pre-rated in terms of experienced valance and arousal, assuring that the pictures would elicit the expected effect. Third, a main difference to inducing negative mood online versus in the lab is that there is little opportunity for aftercare. For example, if the image triggered a negative response or resulted in reliving a challenging moment in life, there is no one in immediate proximity to care for the affected person—a situation that might cause potential harm and needs to be addressed.

To accommodate potential discomfort arising from being exposed to gruesome imagery, I took several steps. First, I warned participants early on about the content of the study. Tasks presented on Amazon Mechanical Turk are announced on a webpage that lists all available tasks. Workers can click on a task and are informed about the content of the task and then can decide to accept the task or find a different task. I presented disclaimers that the task would include gruesome imagery within the job advertisements and on the actual task website. Second, I stated clearly in the consent form that withdrawing from the study at any time would be without consequences. And third, at the end of the study, I explained clearly my intentions, provided contact information to a help line (i.e., https://www.crisistextline.org/), and links to images of baby animals to shift participants' cognition away from the gruesome imagery and towards positively-valenced images that comfort and help with mood repair.

I assured that participants' well-being was placed first, while also being aware of potential effects on the study. For example, the multiple disclaimers and trigger warnings might have affected my sample; people who perceive themselves as most susceptible to negative imagery likely did not choose to engage in the study, potentially resulting in lower average state-anxiety as a sample compared to the population as a whole. Lower variance in response to the induction would decrease the likelihood of finding potential group differences in the sample, assuming that avatar customization indeed affects training outcome.

5.3.4 Experimental Design Implications

In summary, the study was designed to identify small effects, if they exist. Showing group differences is challenging—i.e., the manipulation is subtle, the strongest responders are likely not in the sample. Whereas the design allowed me to show the investigated effects, i.e., the sample size is adequate, training time is in-line with previous research and long enough, and the conditions are distinct enough to potentially show effects. As a result of the challenging but fair study design, small group differences would be considered as meaningful.

5.4 Manuscript D

Background: The success of Internet-based mental health interventions in practice—i.e., in-the-wild—depends on the uptake and retention of the application and the user's focused attention in the moment of use. Incorporating game-based motivational design into digital interventions delivered in-the-wild has been shown to increase uptake and retention in Internet-based training; however, there are outstanding questions about the potential of game-based motivational strategies to increase engagement with a task in the moment of use, and the effect on intervention efficacy.

Objective: Designers of Internet-based interventions need to know whether game-based motivational design strategies can increase in-the-moment engagement and thus improve digital interventions. We investigate the effects of one motivational design strategy (avatar customization) in an example mental health intervention (computerized cognitive training for attention bias modification).

Methods: We assigned 317 participants to either a customized avatar or an assigned avatar condition. After measuring state-anxiety (STAI), we randomly assigned half of the participants in each condition to either an attentional retraining condition (ABMT) or a control condition. After training, participants were exposed to a Negative Mood Induction using images with strong negative valance (IAPS), after which we measured state-anxiety again.

Results: Avatar customization decreased post-training state-anxiety when controlling for baseline state-anxiety, for those who trained attention bias modification; however, those who did not train experienced decreased resilience to the negative mood induction ($F_{1,252} = 6.86, p = .009, \eta_p^2 = .027$). Together these results suggest that customization increased task engagement with the intervention in the moment of use. Avatar

customization also increased avatar identification ($F_{5,252} = 12.46, p < .001, R^2 = .23$), regardless of whether or not participants trained attentional bias ($F_{1,252} = .79, p = .378$). Avatar identification reduced anxiety after the negative mood induction for participants who underwent training, but increased post-stimulus anxiety for participants who did not undergo training, further suggesting that customization increases engagement in the task ($F_{1,252} = 6.19, p = .014$). The beneficial effect of avatar customization on training was driven by participants who were low in their basic satisfaction of relatedness ($F_{10,248} = 18.5, p < .001, R^2 = .43$), which is important because these are the participants who are most likely in need of digital interventions for mental health (Cacioppo et al., 2015).

Conclusions: Our results suggest that applying motivational design—specifically avatar customization—is a viable strategy to increase engagement and subsequently training efficacy in a computerized cognitive task.

5.4.1 Introduction

Internet-based mental health interventions are necessary to address a growing gap between an increased demand for treatment of mental health issues and the capacity of traditional therapeutic approaches to meet this growing demand (Kazdin & Blase, 2011). It has been argued that Internet-based mental health interventions have some benefits over traditional approaches, especially related to the increased accessibility of treatment (e.g., due to geographical constraints of living in remote areas with access to clinical treatment), the ability to scale (e.g., to seamlessly address the growing demand of people who could benefit from treatment), the ease of access (e.g., through deployment of smartphones or websites), and the broad appeal (e.g., to treat subclinical populations who may not qualify for treatment in health care systems that are already stretched in meeting the needs of clinical populations) (Kazdin & Blase, 2011; Mandryk & Birk, 2017; Mohr, Tomasino, et al., 2017).

To be successful, an Internet-based mental health intervention requires good intervention design as evidenced by clinical efficacy. There are several examples of Internet-based mental health applications that have demonstrated treatment effects (Curran; et al., 2012; Mohr, Lyon, et al., 2017) in Randomized Control Trials (RCTs) (Meyer et al., 2015; Thiart et al., 2015; Christensen, Griffiths, & Jorm, 2004). However, because these mental health interventions are delivered digitally and often outside of a laboratory or clinical context, success is not only defined by efficacy in research, but also defined by efficacy in practice, which depends on validated implementation models and a focus on external validity—that is, demonstrated success in the context of intended use (Curran; et al., 2012; Gilbody et al., 2015; Mohr, Lyon, et al., 2017). Success in practice requires uptake (i.e., people have to use the applications), retention (i.e., people have to use the applications over a long enough term to experience benefits), and demonstrated treatment effects under conditions of practice (i.e., the clinical efficacy must translate into less controlled environments with multiple competing demands for a patient's attention and time). These elements of demonstrated success in practice have been sources of failure for Internet-based mental health interventions that have been effective in RCTs, but are compromised by low uptake (i.e., recruitment challenges (Mohr et al., 2012), attention (i.e., failure to effectively engage participants in the moment of intervention use; see Ryan et al., 2008), and retention (i.e., failure in adherence over the medium-term; see Donkin et al., 2011) when delivered "in-the-wild". Researchers have recently promoted a need for demonstrated success of treatment efficacy in both research and practice; see Gilbody et al., 2015; Mohr, Lyon, et al., 2017 and the failure to close the research-practice-gap has prompted suggestions that researchers still have a lot to learn about how to implement these types of interventions in-the-wild (Mohr, Lyon, et al., 2017).

Researchers have argued that one source for improving the in-the-wild success of Internet-based mental health interventions is improving the user experience design (Mohr, Lyon, et al., 2017). Internet-based applications built for the purposes of leisure or enjoyment also depend on user engagement—i.e., good uptake, attention, and retention—because the micro-transaction model of monetization for the companies that build these applications requires that end-users chose the application (uptake), engage with it (attention), and continue to use it (retention). Various interaction design strategies have been employed by leisure application designers to improve user engagement, including gamification—i.e., the use of game-based elements in non-game contexts (Deterding et al., 2011), the inclusion of extrinsic rewards (Deci et al., 1999), or leveraging social pressure to engage (Armitage & Conner, 2001). Based on increasing a user's motivation by increasing their enjoyment of and invested effort in engaging with an application, these types of interaction design strategies are part of a growing field of research on motivational design for interactive technologies that have also been recently applied to the context of digital intervention design.

For example, applying motivational design principles has been shown to foster engagement with a 12-week neurofeedback treatment for children with fetal alcohol spectrum disorder (Mandryk et al., 2013), to increase motivation to use a stroke rehabilitation program (Smeddinck et al., 2015), and to improve enjoyment of a social physical activity intervention for children with Cerebral Palsy (Hernandez et al., 2014). Further, a series of studies employing avatar customization (a motivational design strategy built on increased autonomy and identification) was shown to increase enjoyment of and effort invested in a game for training (Birk, Mandryk, & Atkins, 2016), to increase the time spent (free-choice) in a game (Birk, Atkins, et al., 2016), and to combat attrition in an Internet based daily breathing exercise deployed over three weeks (Birk & Mandryk, 2018).

Previous work has suggested that motivational design principles could be employed to help close the research-practice gap for the design of Internet-based mental health interventions by improving user engagement with the application (Mandryk & Birk, 2017); however, prior work that led to these ideas has measured success through two approaches: first, by focusing on subjective measures of motivation, such as increased enjoyment of the intervention or perception of effort invested in the task (e.g., Birk, Mandryk, & Atkins, 2016). And second, through metrics that operationalize intervention usage statistics, such as the time spent in treatment or number of returning sessions (e.g., Birk & Mandryk, 2018). However, there is a third approach to characterizing success of motivational design as applied to digital interventions that has been underserved: that is, metrics that characterize increased engagement with the task in the moment of use. As researchers, we must differentiate between motivational design that results in greater exposure to treatment (i.e., more time spent in training, more adherence)—which should result in improved efficacy through mere exposure, and motivational design that fosters task engagement in the moment (e.g., greater attention and focus, and reduced response to distraction)—which should improve treatment efficacy without an accompanying increase in exposure.

In this paper, we employ one motivational design strategy (avatar customization) in an example mental health intervention (computerized cognitive training) to demonstrate that motivational design principles can not only improve exposure to treatment through greater uptake and retention, but that they can improve focused engagement in the moment of intervention use.

5.4.2 Computerized Cognitive Training as a Mental Health Intervention

Computerized cognitive training (CCT) is one approach to intervention design with a focus on improving specific aspects of cognition. Feasibility studies on CCT have been shown to improve memory, self-control, reasoning, attention bias, and processing speed (Lampit et al., 2014); CCT has successfully been applied in clinical research to combat mental illness and cognitive deficits in disorders such as dementia (Ngandu et al., 2015), depression (Coyle et al., 2015; Motter et al., 2015, 2016), neurodegenerative diseases (Coyle et al., 2015), attention-deficit hyperactivity disorder (ADHD) (Cortese et al., 2015), and brain injury (Bogdanova et al., 2016). The most common CCT tasks are the go/no-go task (Gomez & Perea, 2009), memory training (Klingberg et al., 2005), and the Attention Bias Modification Training (ABMT) Task (Hakamata et al., 2010). For example, go/no-go tasks require participants to inhibit responses under changing conditions (e.g., in a fast-paced task, press "L" when a red box appears, but inhibit pressing "L" when a green box appears). Go/no-go paradigms have been applied to the training of executive function (Thorell et al., 2009), and research suggests that the paradigm can improve hyperactivity symptoms for children with ADHD (Prins et al., 2011) and reduce undesirable food intake when applied in the context of eating behaviour (Van Koningsbruggen et al., 2014).

Another approach to cognitive retraining is through attention bias modification (Hakamata et al., 2010), which exposes participants simultaneously to negative and neutral stimuli, but reinforces an attentional shift towards neutral stimuli by presenting target probes behind only the neutral stimuli. ABMT has been shown to be an effective technique to shift a participant's attention away from negative stimuli, to decrease self-reported anxiety, and to decrease the response to negative stimuli (Barry et al., 2015; Hakamata et al., 2010; Jones & Sharpe, 2017). While lab-based ABMT training has been shown to be effective (Jones & Sharpe, 2017), Internet-based ABMT has generally not been shown to be effective, suggesting that the training task itself might require adjustment before dissemination over the Internet (Carlbring et al., 2012). In the case of training tasks that require the full attention of the patient, the lack of control over distractions and attention in the environment present in Internet-based interventions may compromise treatment efficacy when delivered in-the-wild: for CCT to be fully effective, participants need to be vigilant, psychologically present, and engaged in the task.

CCTs have shown moderate-to-large effect sizes for improving attention, working memory, and global functioning (Motter et al., 2016). However, to show effects in-the-wild, CCTs need to be designed in ways that maximize user engagement in the moment. Being inattentive, unfocused, or distracted will diminish the efficacy of attention-based training (Chu & Kendall, 2004; Glenn et al., 2014). When CCT is applied in studies, or during a session with a therapist, participants are externally regulated to focus on the task; however, this external regulation is drastically lessened when people engage in cognitive training during a commute, at home, or while they have a few minutes waiting in line. To support the success of Internet-based mental health interventions in-the-wild, researchers need to ask, *how can we increase in-the-moment engagement to compensate for inattentiveness and distraction in-the-wild and subsequently improve training efficacy?*

5.4.3 Engagement

While Internet-based mental health applications have increased the accessibility of treatment, their use still requires participant engagement and effort (Ryan et al., 2008). Theories of motivation provide guidance on how to design applications to maximize engagement. Self-Determination Theory (SDT), a well-established theory of human motivation (Ryan & Deci, 2000b), postulates that competence, relatedness, and autonomy are three predictors of motivation, which is expressed in terms of enjoyment, engagement, and effort. Competence—experiencing mastery over a task, Autonomy—volitionally engaging in a task, and Relatedness–experiencing connectedness to others—predict engagement and have been shown to be positively related to treatment outcome (Ryan et al., 2008). For example, clients who engage in therapy out of personal choice, i.e., autonomy, are more likely to benefit from therapy. Increased volitional engagement has been shown to increase adherence and treatment efficacy (Joosten et al., 2008).

5.4.3.1 Designing for Engagement

Multiple strategies have been applied to foster volitional engagement in a digital context. For example, gamification—the application of game elements in a non-game context (Deterding et al., 2011)—has successfully been applied to increase volitional engagement in a variety of contexts (Connolly et al., 2012). Game-based training has been shown to improve working memory capacity (Basak et al., 2008), task switching ability (Ibid), visual-short term memory (Ibid), verbal reasoning (Ibid; Subrahmanyam & Greenfield, 1994), visuospatial reasoning (Boot et al., 2008; Dorval & Pépin, 1986), response selection (Clark et al., 1987), visual attention (Drew & Waters, 1986; Green & Bavelier, 2003), reaction time(Dustman et al., 1992; Orosy-Fildes & Allan, 1989), and choice reaction time (Dye et al., 2009; Goldstein et al., 1997).

In a similar vein, persuasive technology uses strategies to bring about change by shaping or reinforcing behaviors or attitudes (Fogg, 2002). Two strategies commonly employed in persuasive technologies are personalization (Kaptein et al., 2015; Oinas-Kukkonen & Harjumaa, 2008)—which is system-initiated tailoring that offers content or services personalized to an individual; and customization (Kaptein et al., 2015; Orji et al., 2014)—which is a system that supports user-initiated tailoring of content or services. Sundar and Marathe (2010) argue that although personalization will increase the relevance of content for individuals using an interactive system, customization yields systems and content that are not only relevant, but also boost the agency and self-determination of the individual because it is they themselves who perform the tailoring. Personalisation and customization have both been discussed as techniques to increase long-term engagement with Internet-based interventions for mental health (Birk & Mandryk, 2018; Doherty et al., 2012).

Customization fosters autonomy, a sense of control, and a sense of identity, making the person feel relevant in the context of their interaction (Sundar & Marathe, 2010). As previously described, a series of studies employing avatar customization (which facilitated avatar identification) showed increased enjoyment of and effort invested in a game for training (Birk, Mandryk, & Atkins, 2016), increased the time spent (free-choice) in a game (Birk, Atkins, et al., 2016), and combatted attrition in an Internet-based daily breathing exercise deployed over three weeks (Birk & Mandryk, 2018). Together, these studies show improvements in subjectively-measured motivation and objectively measured motivation in terms of the time spent engaged with the application and number of return sessions in a three-week intervention; however, there was no attempt to demonstrate improved treatment efficacy as a result of the increased engagement with the application in the moment of use. There was no differentiation between how the motivational design resulted in greater exposure to treatment and how the motivational design fostered task engagement in the moment (e.g., more effort invested, better attention and focus, or reduced response to distractions), which should improve treatment efficacy without an accompanying increase in exposure.

Current digital mental health interventions require the attention and motivation of patients (Mohr, Lyon, et al., 2017), which are unfortunately also characteristics that are in short supply for people who suffer from depression and could benefit most from treatment (Ibid). Previous work showed that avatar customization can increase motivation to engage with a training application and thus increase the time spent in training. In this paper, we suggest that employing motivational design principles—specifically the use of avatar customization—can increase the efficacy of an attention bias modification training task delivered online by increasing task engagement *in the moment of training*, without requiring additional exposure to the treatment.

5.4.4 Methods

We conducted an online study in which we asked half of the participants to customize an avatar; the other half were assigned a generic avatar. Each avatar group performed an attention bias modification training (ABMT) task—half of the participants in each group were trained to preferentially attend to neutral over negative stimuli; the other half were not. Trained participants should be more resilient to subsequent negative stimuli; thus participants were all subjected to a negative mood induction (viewing gruesome images) after completing the ABMT task and completed state anxiety scales prior to training and following the stimulus.

Our experiment was a 2 (avatar: customized, generic) by 2 (attentional training, no-training) between-subjects design; see Figure 5.1. Being inattentive, unfocused, or distracted will diminish the efficacy of attention-based training, thus we ask whether the increased in-the-moment engagement as a result of avatar customization can improve training efficacy.

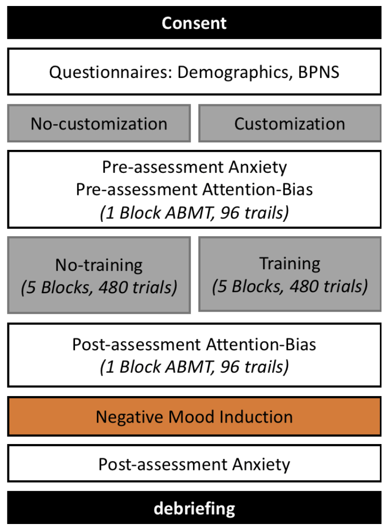

Figure 5.1: Experimental flow from consent (top) to debriefing. The experimental conditions are highlighted in grey. The negative mood induction is highlighted in orange.

The following research questions guided our analyses:

RQ1. Does customization improve the efficacy of attentional retraining?

RQ2. Does customization increase identification for trained and non-trained participants?

RQ3. Does avatar identification increase the efficacy of attentional training?

RQ4. Does training efficacy vary depending on basic needs satisfaction?

5.4.4.1 Customization using an Avatar Creator

To introduce customization, we used an avatar creator that has been shown to facilitate intrinsic motivation and invested effort in a game (Birk, Atkins, et al., 2016). Participants were asked to create an avatar, choose its gender, and adjust its appearance, personality, and attributes (characteristics) in the same manner as described in (Birk, Atkins, et al., 2016); see Figure 5.2. A minimum of four minutes in the character creator were required, but participants could take longer if they wished. After customizing their avatar, participants were shown a summary of their character.

We presented half of our participants with the avatar creator; the other half were assigned an avatar that had generic features, and a medium hair and skin colour (see Figure 5.4). Participants were asked their gender in a demographics survey; those who answered male or female were assigned an avatar of the same gender. Those who answered ‘other’ ($n = 3$) were then asked to choose a gender for their digital representation. For both generic and customized avatars, we created and stored two facial expressions: one neutral face and one angry face (see Figure 5.4). The faces were created by adjusting the 3D model of the face using an algorithm based on Ekman's Facial Action Coding System (FACS, Ekman & Friesen, 1977). Because of the differences in facial geometry between male and female faces, there were two algorithms (one for male, one for female) and all avatars of the same gender had the same algorithm applied. Once the avatars were customized or assigned, participants completed the ABMT task.

Figure 5.2: Picture of the avatar creator displaying the customization tool box on the left and a female avatar on the right.

5.4.4.2 Attention Bias Modification Training Task

In each trial of the ABMT task (Hakamata et al., 2010), a fixation cross was presented for 500ms centered in the screen. Following the presentation of the fixation cross, two avatar faces were displayed: one with a neutral facial expression and one with an angry expression. One face was displayed above the cross and one was displayed below the cross. Depending on the condition the avatar faces were either based on the customized avatar or the assigned avatar. After the faces were presented for 500ms, they disappeared and a probe was displayed behind one of the faces. The probe indicated either a right arrow or left arrow and participants were asked to press the corresponding arrow key on the keyboard as quickly as possible. After pressing the left or right arrow key, an inter-trial interval, showing a white screen, was displayed for 500ms. The next trial started immediately with the presentation of the fixation cross. See Figure 5.3.

Figure 5.3: ABMT trials. From left to right: Fixation cross (500ms), neutral/angry face (500ms), probe (participants response), and inter-trial interval (500ms).

5.4.4.3 ABMT: Training/No-Training

Before beginning the ABMT task, participants were guided through a tutorial of 10 trials of neutral faces to learn the mechanics of the task. Participants were prompted to focus on the fixation cross and could only proceed if they pressed the correct arrow key as indicated by the probe.

Following the tutorial, the ABMT was presented in seven blocks: one pre-assessment block, five consecutive training/no-training blocks, and a post-assessment block. During each block, participants completed 96 trials (672 total). After each block, participants had a 6-second break, while being notified that the next block was about to start. This was done to indicate progress through the task.

We used two image sets of four avatar faces (two male, two female) for the pre-assessment and post-assessment blocks; the order of the presentation of the two image sets was fully counterbalanced to avoid order effects. The avatars were selected from user-generated avatars from a previous study (Birk, Atkins, et al., 2016). The presentation of sex (male, female), position (top, bottom), probe location (angry, neutral), and probe direction (left, right) was fully balanced over the 96 trials.

In the ABMT training condition, probes only appeared behind neutral faces, with the intent to shift participants' attention towards neutral stimuli. In the no-training condition, participants were exposed to a fully-balanced probe presentation similar to the assessment condition in which 50% of the probes appear behind the neutral faces and 50% behind the angry faces. See Hakamata et al. (2010) for a detailed description of the ABMT task.

Before beginning the ABMT task, participants were guided through a tutorial of 10 trials of neutral faces to learn the mechanics of the task. Participants were prompted to focus on the fixation cross and could only proceed if they pressed the correct arrow key as indicated by the probe. After each of the seven blocks of trials, participants completed a mood scale with 7-point agreement to four states (I feel: relaxed, happy, depressed, anxious) representing the four corners of arousal-valence space (Mandryk & Atkins, 2007).

5.4.4.4 Trial Removal and Logging

Incorrect responses to the probes were removed from subsequent analysis because they show that the participant did not pay attention in the trial. Individual trials were also removed when participant's response time was greater than three standard deviations of their own mean performance over both assessment blocks.

Figure 5.4: Customized (top) and generic (bottom) male and female avatars with neutral and angry facial expressions.

We logged the position (top, bottom), gender (male, female), and expression (angry, neutral) of the probe location, the response time (in ms), and the correctness of the probe response (true, false).

5.4.4.5 Negative Mood Induction

To measures resilience to a negative mood induction, we presented twenty negative images from the International Affective Picture System (IAPS, Lang et al., 2008)—ID: 2703, 3010, 3015, 3225, 3230, 3350, 3530, 3550.1, 9040, 9265, 9301, 9410, 9420, 9433, 9490, 9500, 9570, 9611, 9635.1, 9901. Images were selected based on valence ($mean = 2.01, SD = 0.5, min = 1.51, max = 3.60$), and arousal ($mean = 5.92, SD = 0.63, min = 4.34, max = 7.16$). To ensure that participants looked at the images, we asked them to rate each image using the valence and arousal scales of the visual Self-Assessment Manikin (SAM, Bradley & Lang, 1994)—valence and arousal scales were sequential to increase the time spent looking at each image. IAPS images have previously been used as a negative mood induction in the context of the ABMT (Hakamata et al., 2010). Descriptively, participants show a similar response to the negative images as with normative IAPS ratings (valance: $mean = 1.91, SD = .62, min = 1, max = 4.05$; arousal: $mean = 4.97, SD = 1.215, min = 1, max = 7$), indicating that the images were perceived as expected.

5.4.4.6 Participants and Deployment Platform

We recruited 317 participants through Amazon's Mechanical Turk (MTurk). MTurk acts as a broker between parties offering a range of Human Intelligence Tasks (HITs) and paid workers. Although MTurk has been shown to be reliable as a recruitment tool for research (Crump et al., 2013; Gillan & Daw, 2016; Shapiro et al., 2013), we excluded participants from analysis if they were not performing the task with care, which we determined in multiple steps. We removed 33 participants based on missing trials (indicating technical difficulties), or too many trials (indicating that they reloaded the task part way through). Then we calculated variance within each survey subscale and removed participants ($n = 8$) from subsequent analyses who demonstrated response variance greater than three standard deviations above mean variance on three or more questionnaire subscales. Having high variance within a subscale is indicative of not paying attention to the survey questions and the reverse-coded items. We also removed participants from subsequent analyses who completed two or more questionnaires with an average time below 1 SD of the average response time ($n = 10$). Finally, we removed participants who spent more than two minutes answering the state anxiety questionnaire after the negative mood induction ($n = 4$), as it would indicate that they were taking time to recover while answering the questions.

After the outlier participants were excluded, 262 participants remained. Because we controlled our analysis for age and gender, we also excluded the 3 participants who identified their gender as 'other', leaving 259 participants ($51\% female$,*mean-age*$= 35.3, SD = 11.5$) in all of our analyses. Participants received compensation of $10 for their participation. Ethical approval was obtained from the University of Saskatchewan behavioural research ethics board, and participants were asked to give informed consent at the beginning of the task. To comply with ethical guidelines, the task was only available to workers who were older than 18. Additionally, only workers from the USA with an approval rate above 90% were offered the task as a means of quality control and a trigger warning was provided at multiple points prior to the negative images being presented.

5.4.4.7 Measures

Identification was measured using the avatar-related sub-scales of similarity identification, embodied identification, and wishful identification from the Player Identification Scale (PIS, Van Looy et al., 2012). Participants rated their agreement to identification-related statements—"My character is like me in many ways"—on a 7-pt Likert scale. Identification has been shown before to be an important construct factor when customizing avatars (Birk, Atkins, et al., 2016).

State-Anxiety was measured using the state scale from State-Trait-Anxiety Inventory (STAI, Spielberger et al., 1970). Participants rated how well statements—e.g., "I'm calm"—described their current state on a 4-point scale from "Not at all" to "Very much". STAI has been successfully been used before in research on resilience to negative affect (Davidson et al., 2003).

Basic Needs Satisfaction was measured using the Basic Psychological Need Satisfaction scale (BPNS, Gagne, 2003). BPNS includes subscales for the basic satisfaction of competence, autonomy, and relatedness, as three ongoing needs that people need satisfied to optimally develop and function. Participants rated their agreement to statements—"People I know tell me I am good at what I do"—on a 7-point Likert-scale. The scale has been used before in research on resilience to negative affect (Gagne, 2003).

5.4.4.8 Procedures

Participants were informed about time (60 min) and payment ($10), procedure, and the fact that the study included gruesome imagery. After giving consent, participants were asked to fill in questionnaires on their demographics, basic needs satisfaction, and the questionnaire for baseline state anxiety. Participants were then assigned to one of four conditions: customized avatar/ABMT training, customized avatar/no ABMT training, generic avatar/ABMT training, generic avatar/no ABMT training. For descriptive statistics, see Table 5.1.

Depending on the condition, participants either customized an avatar or were assigned a generic avatar and started immediately with the ABMT tutorial. After the tutorial, all participants did a block of pre-assessment trials with one of the previously-described image sets. Following this initial block, half of the participants received five blocks of attentional retraining with probes only behind neutral faces, while the other half completed five blocks with the probe appearing equally behind neutral and angry faces. All participants performed one block of post-assessment after training/no-training with the other image set. Following assessment, all participants were exposed to the negative mood induction using the gruesome IAPS images. Following the negative mood induction, participants filled in the post state-anxiety questionnaire. Finally, participants provided information about how much they identified with the avatar face that was used for training/no-training. Participants were then debriefed about the purpose of the experiment and were directed to adorable images of baby animals if they wanted to combat the negative mood induced by the IAPS images. See Figure 5.1 for the experimental flow.

5.4.4.9 Data Collection and Analysis

All data were logged to a database on a server at the University of Saskatchewan and were analyzed using SPSS 24 with the Process macro for moderation and mediation analyses (Hayes, 2012).

Prior to analyzing the effects of training and customization on anxiety, we analyzed the block-based mood ratings to ensure that the different conditions did not directly induce different vulnerability to the negative mood induction. There were no significant interactions with either avatar customization or training, implying that the conditions did not directly influence mood (see Macleod et al., 2002).

We computed four models. First, we investigated the role of avatar customization and attentional retraining on state anxiety after the negative mood induction. We use an ANCOVA with avatar (customized, generic) and attentional retraining (training, no-training) as factors on the dependent measure of state anxiety measured after the negative mood induction. The ANCOVA allowed us to control for levels of state anxiety prior to attentional retraining, age, and sex by including these variables as covariates. ANCOVAs have been shown to have higher power in randomized studies with a baseline compared to using repeated measures ANOVAs (Van Breukelen, 2006).

The next three regression models were all controlled for baseline state anxiety, age, and gender. In our second model, we considered whether avatar customization (X) predicts identification (Y), and whether

	No-training		Training	
Avatar customization	*M*	*SD*	*M*	*SD*
	STAI (pre)			
No-customization	1.95	0.56	2.05	0.64
Customization	1.94	0.58	1.98	0.57
	STAI (post)			
No-customization	2.37	0.59	2.60	0.67
Customization	2.55	0.58	2.37	0.59
	Identification			
No-customization	2.63	1.35	2.62	1.29
Customization	4.11	1.33	3.76	1.27
	Relatedness			
No-customization	4.97	1.29	4.97	1.08
Customization	5.15	1.02	5.02	1.13
	Age			
No-customization	37.75	10.83	37.75	10.83
Customization	32.20	9.29	32.20	9.29
	Frequency	*Percent*	*Frequency*	*Percent*
	Female/Male (n)			
No-customization	33/31 (64)	51.6/48.4	27/31 (58)	46.6/53.4
Customization	39/32 (71)	54.9/45.1	34/32 (66)	51.5/48.5

Table 5.1: Descriptive statistics for dependent and control variables displayed by avatar customization and training condition.

training (M) moderates this relationship (model=1, Hayes, 2012). The link between avatar customization and identification has been previously established (Birk, Atkins, et al., 2016; Trepte & Reinecke, 2010).

Third, we considered whether identification (X) predicts anxiety (Y) (post negative mood induction) and whether training (M) moderates this relationship (model=1, Hayes, 2012).

Finally, we conducted a moderated moderation (model=3, Hayes, 2012). We considered whether avatar customization (X) predicts anxiety post negative mood induction (Y), whether this relationship is moderated by baseline need satisfaction (M), and whether that moderation is moderated by training (W). A moderated moderation model is similar to a 3-way interaction in an ANOVA, but allows for the inclusion of continuous variables as factors.

5.4.5 Results

We created four statistical models, as described above, and use these models to answer a series of research questions.

5.4.5.1 RQ1: Does customization improve the efficacy of attentional retraining?

Participants in the ABMT training group who customized their avatar should show increased resilience to negative stimuli. An ANCOVA controlling for pre-training anxiety, age, and gender, showed a significant interaction between avatar customization and attentional retraining on anxiety measured after the negative mood induction ($F_{1,252} = 6.86, p = .009, \eta_p^2 = .027$). Bonferroni-corrected post-hoc tests showed that training reduces state anxiety only when participants used a customized avatar ($p = .038$); when a generic avatar was used, training was not more effective than no-training ($p = .103$). This result suggests that attentional retraining is more effective when participants were allowed to create a customized avatar and were presented with personalized stimuli (see Figure 5.5).

While the interaction of avatar and training shows a significant effect, the partial eta-squared value ($\eta_p^2 = .027$) implies a small effect. The large effect on post-stimulus anxiety ($R^2 = .401$), is mostly explained by baseline anxiety ($\eta_p^2 = .353$). However, when looking at the model by individual effect sizes of the included variables (age, gender, training, avatar, and the interaction of training and avatar), we find that age ($\eta_p^2 = .032$) and gender ($\eta_p^2 = .030$) both show small effects. Training and avatar (both not significant), explain almost no variance ($\eta_p^2 < .001$, and $\eta_p^2 = .001$, respectively). The interaction of avatar and training ($\eta_p^2 = .027$) shows an effect similar in size to the effects of major demographic variables (i.e., age and gender) on anxiety.

These results are in-line with previous research, showing that customization increased identification with a digital representation, and as a result increased positive affect, task enjoyment, invested effort, and motivated behavior measured as time-spent in a free-choice game (Birk, Atkins, et al., 2016), and that these effects are of a similar size as the effects of major demographic variables.

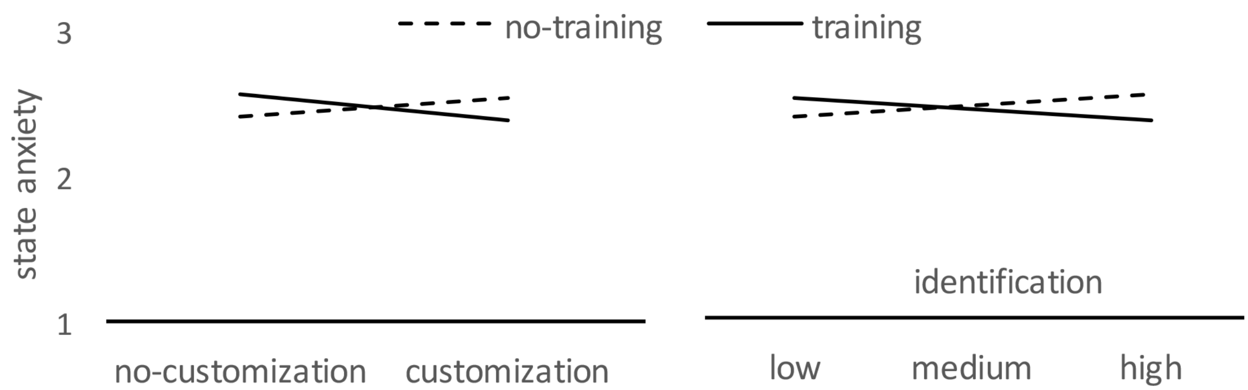

Figure 5.5: Left: State anxiety for the interaction of training and avatar customization (RQ1). Right: State anxiety for the interaction of training and identification (RQ3). Displayed data is controlled for age, gender, and baseline state anxiety.

5.4.5.2 RQ2. Does customization increase identification for trained and non-trained participants?

We investigated the prediction of avatar customization on avatar identification, moderated by training. The model was significant ($F_{5,252} = 12.46, p < .001, R^2 = .23$). Customization predicts identification ($\beta = .651, p < .001$); however the non-significant interaction with training ($F_{1,252} = .79, p = .378$) shows that the prediction does not depend on whether the probe appeared under angry or both angry and neutral faces (see Figure 5.6).

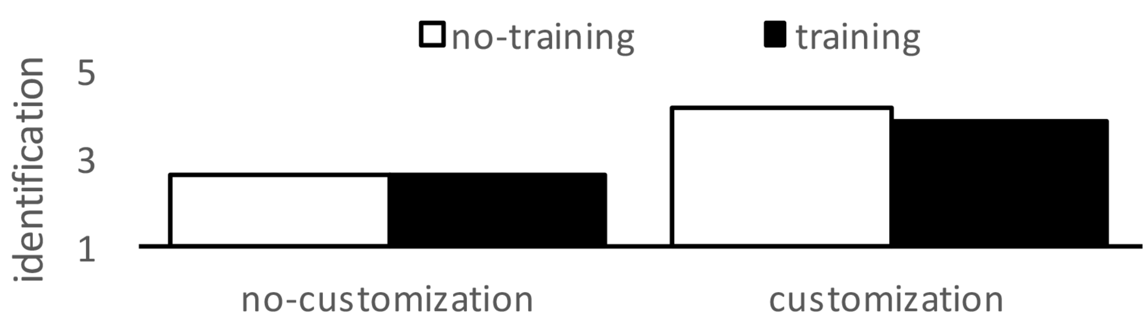

Figure 5.6: Effect of training and customization on identification (RQ2). Displayed data is controlled for age, and gender.

5.4.5.3 RQ3. Does avatar identification increase the efficacy of attentional training?

We investigated the prediction of avatar identification on anxiety measured after the negative mood induction, as moderated by training. The model was significant ($F_{6,252} = 28.49, p < .0001, R^2 = .40$). There was neither a significant main effect for identification ($p = .990$), nor for training ($p = .990$). However, the interaction between identification and training ($F_{1,252} = 6.19, p = .014$) shows that for participants who underwent training, identification tends to reduce anxiety, i.e., increases training efficacy; however, for participants who did not undergo training, the trend is in the other direction—greater identification increases anxiety (see Figure 5.5).

5.4.5.4 RQ4: Does training efficacy vary depending on basic needs satisfaction?

We showed in R1 that avatar customization and training interact to yield lower anxiety. Now we ask whether this effect is being driven by participants with varying levels of basic psychological needs satisfaction. We used basic satisfaction of relatedness as it has been shown to be an important predictor of depression, addiction, and other mental disorders (Blatt & Zuroff, 1992). A moderation model with the predictor avatar (X) on state anxiety (Y) moderated by satisfaction of relatedness (M), and then moderated by training (W) was significant ($F_{10,248} = 18.5, p < .001, R^2 = .43$); we controlled for baseline state anxiety, gender, and age. As expected, the interaction between avatar and training was significant ($p = .007$). Most interestingly, the interaction between satisfaction of relatedness, avatar, and training also was significant ($p = .021$). This 3-way interaction showed that the interaction between avatar customization and training is more pronounced for those lower in satisfaction of relatedness (see Figure 5.7).

Participants with lower relatedness satisfaction were driving the significant interaction between avatar customization and training, which is a meaningful result, as these are the participants who are most likely to be in need of a mental health intervention in the first place (Blatt & Zuroff, 1992).

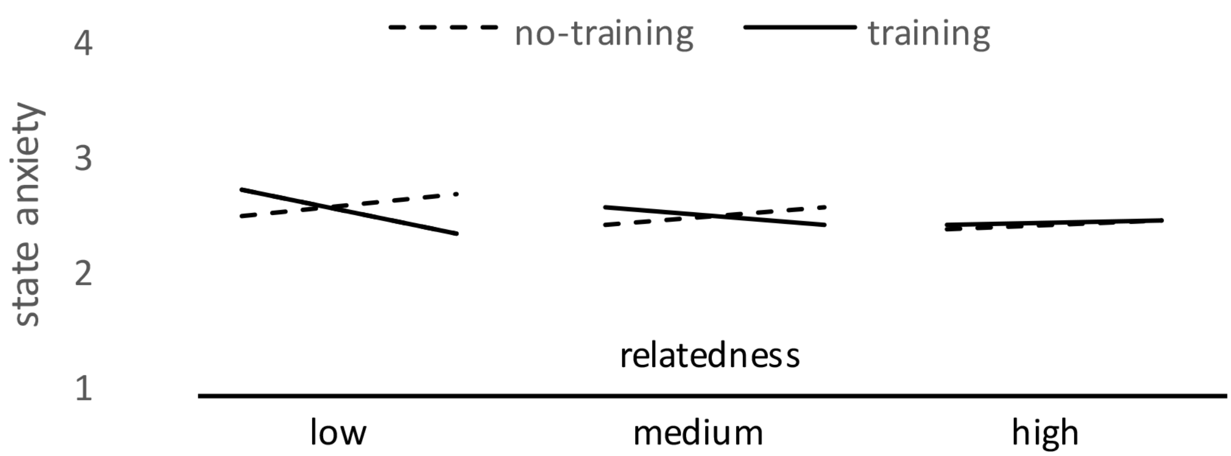

Figure 5.7: Moderation model showing avatar on state anxiety moderated by satisfaction of relatedness, and then moderated by training (RQ4). Displayed data is controlled for age, gender, and baseline state anxiety.

5.4.6 Discussion

We summarize our findings, contextualize them in theory, and discuss the implications, limitations and opportunities for future work.

5.4.6.1 Principal Results

The results revealed four main findings:

- First, avatar customization increased resilience to a negative mood induction for participants who trained attentional bias, compared to those who did not train attentional bias. Thus, avatar customization improved training efficacy, presumably as a result of increased in-the-moment engagement.

- Second, avatar customization increased avatar identification, regardless of whether or not participants trained attentional bias.

- Third, avatar identification tended to reduce anxiety after the negative mood induction for participants who underwent training but tended to increase post-stimulus anxiety for participants who did not undergo training, suggesting that customization increases in-the-moment engagement.

- Fourth, the beneficial effect of avatar customization on training is being driven by participants who are low in their basic satisfaction of relatedness, which is important because these are the participants who are most likely in need of digital interventions for mental health.

5.4.6.2 Explanation of Findings

Our findings suggest that attentional retraining is improved when participants are more engaged and invested in the task. Based on prior research (Birk, Atkins, et al., 2016), we assume that increased motivation and willingness to invest effort increases attention to the task, and efficacy as a result. We further explain potential mechanisms of why avatar customization increased treatment efficacy.

Self-Determined Experience The results presented in this study, and previous research using avatar customization (Birk, Atkins, et al., 2016), show that customization increases identification and that identification positively predicts task engagement. Engagement can be understood as an increase in participants' invested effort; a central construct in describing intrinsic motivation and task engagement (McAuley et al., 1989). Research on motivation (Gagné & Deci, 2005; Ryan & Deci, 2000a) suggests that invested effort depends on factors related to the experience itself; in particular, whether or not the interaction satisfies our needs for competence (i.e., feeling a sense of mastery over challenges), autonomy (i.e., experiencing the choice to engage under our own volition), and relatedness (i.e., feeling connected to others) (Ryan et al., 2006). In our experiment, we aimed to increase motivation by manipulating choice as a means of increasing autonomy, i.e., participants either customized an avatar, or they were assigned a generic avatar. Exercising choice has motivational benefits, increasing the sense of volitional control and ownership over analog and digital objects (Beggan, 1992; Huang et al., 2009).

Self-Determination Theory further suggests that intentionally engaging in a task is an important predictor of outcome—presumably, because of the increased attention and focus paid to the task at hand. A more positive attitude toward the task—facilitated for example by feeling volitional control over the content or ownership over objects—may lower cognitive resistance to engagement. Participants in the customization condition may have been less likely to reject engaging with the task simply because it did not appeal to their preferences. The ABMT—like other computerized cognitive training tasks—is not a particularly entertaining task, so frustration and boredom have the potential to undermine the positive effects of the training; however, we assume potential boredom and frustration are partially counteracted by the positive effects of customization.

Personalized Experience We have argued that avatar customization increased efficacy by increasing autonomy and invested effort, or by improving attitude and feelings of control. In addition, the customized avatar condition presents tailored stimuli, i.e., customized avatar faces with neutral or negative expressions, during the training phase. Tailoring (through customization or personalization) is a persuasive strategy that has been used for example, to increase motivation to play serious games (Göbel et al., 2010), and to improve the efficacy of serious games for changing attitudes, intentions, and self-efficacy around healthy eating behaviour (Orji et al., 2017). The personalization of the stimuli in our experiment may have directly affected participants' perception of the stimuli, making the stimuli's emotions (neutral, negative) more salient to them.

5.4.6.3 Individual Differences

Our results suggest that the customized avatar increases engagement in the moment of use and subsequently training efficacy; however, we also show that the improvements resulting from avatar customization were most pronounced for those participants who experience low satisfaction of relatedness, i.e., people who feel less connected to others, experience less support, or even feel lonely. Our intervention facilitates choice, but enhances the experience of relatedness in the application. Although we did not explicitly measure relatedness satisfaction during the task, prior research has shown that customization can facilitate feelings of connectedness with a digital representation (Ryan et al., 2006). Increased satisfaction of relatedness might have improved engagement with the training task involving an avatar. This is important, because loneliness, feelings of social exclusion, and feeling disconnected from supportive social groups are predictors of vulnerability for many mental health conditions, such as depression and anxiety (Hagerty et al., 1992).

Increased Engagement with no Training Avatar customization increased resilience to negative stimuli after the attentional bias modification training (i.e., saw probes appear only under neutral faces); however, we also observed that the group who used avatar customization without training (i.e., saw probes appear under both angry and neutral faces) were actually more susceptible to the negative mood induction. These results suggest that avatar customization increased task engagement, independent of the training condition. For participants who trained, this resulted in better resilience to the negative stimuli than those who trained with a generic avatar face; however, for participants who did not train, this likely resulted in them investing more attention overall and this being more susceptible to the negative mood induction than those who did not train using the generic avatar faces. This interaction result is not surprising, considering prior evidence of the relationship between avatar customization, identification, and motivation (Birk, Atkins, et al., 2016) in increasing task engagement and further supports our arguments for customization as a motivational design strategy which can increase in-the-moment engagement.

Blending and Extending Existing Therapeutic Approaches There are a variety of existing approaches to maximize the efficacy of Internet-based interventions in mental health. For example, blended interventions, i.e., interventions that blend Internet-based forms of therapy with the interleaved presence of a therapist, have been shown to decrease the load on therapists and are similar in effect to traditionally-delivered Cognitive Behavioral Therapy (CBT) programs (Mohr et al., 2011; Schueller et al., 2017).

Computerized Cognitive Training (CCT), such as ABMT, has been applied adjacent (Shechner et al., 2014) to Internet-based CBTs, showing potential in blending approaches. Considering that different groups of clientele have different demands towards treatments (e.g., treatment intensity) and follow different care pathway around the world, there is value in examining how motivational interface design strategies might enhance exposure and outcome to different therapeutic approaches.

Although our study only considers customization in the context of bias modification, there are other approaches to Internet-based intervention design that may benefit from the increased task engagement that we demonstrate. For example, including customized avatars to guide a patient through a CBT application, or including other personalized stimuli could potentially increase engagement in the moment or result in increased adherence, as suggested by (Birk & Mandryk, 2018). Or consider digital phobia treatments that expose patients to fear-inducing stimuli (e.g., Carlin et al., 1997; Klinger et al., 2005; Parsons & Rizzo, 2008), which could be even more effective if patients are able to customize the presented stimuli, thereby increasing salience in an individual patient's personal context. And finally consider narrative-based therapeutic applications for people who suffer from post-traumatic stress disorder (PTSD), which walk a patient through their experiences and help them to reframe the traumatic event (e.g., Rizzo et al., 2011; Rothbaum et al., 2001). Supporting patients to personalize the narrative, graphical objects, and other intervention elements could increase the efficacy of this important type of mental health intervention.

Until we validate innovative methods, adapting well-evaluated approaches to be delivered at scale is a safe and promising way forward. Exposing people to techniques that are not ready for use as treatment has associated dangers; whereas investigating subtle adaptations to existing approaches—such as interface customization applied to existing treatments—may improve and optimize established interventions.

Interface customization is not an intervention in itself, but rather an enhancement that can be applied across a range of existing interventions. From the perspective of implementation medicine—the branch of medicine that asks if a research result should be implemented in practice (Warner & Obrecht, 2015)—it is a requirement to have evidence-based proof that a technique is either more effective than prior techniques, or at least provides equal effectiveness with a lower investment of time and/or money (Grol & Grimshaw, 1999). Our research shows, that tweaks to existing interventions, such as adding customization, can significantly improve in-the-moment engagement.

Supporting Motivation in the Immediate and Longer Term In this research, we show how avatar customization can increase task engagement and focused attention in the moment of application use. Motivational theories—such as SDT—describe how fostered autonomy can increase the effort invested in a task, which we suggest results in subsequent task engagement during intervention use. In contrast, our previous work on avatar customization has shown increased adherence to a daily training regimen over a medium-term (3-week) breathing intervention (Birk & Mandryk, 2018). It is important to distinguish the motivational benefits that accrue from increased exposure to treatment (i.e., through increased logins or more persistent usage) and those that result from increased attention in the moment of intervention use, without any accompanying increase in treatment exposure. Customizing avatars has been shown to increase motivation both in terms of increased treatment exposure through retention (Birk & Mandryk, 2018) and—in the present paper—through focused attention in the moment. Additionally, personalization has been explored as an approach to enhance longer-term engagement in the context of digital interventions for mental health (Doherty et al., 2012). Further work is needed to explore the potential additive effects of these two approaches in increasing motivation, and to explore the mediating motivational factors (e.g., attention, enjoyment, effort) that could explain the improved outcome observed in the present study.

5.4.6.4 Limitations and Future Work

Our study has limitations that we intend to address in future work.

First, identification occurs in multiple ways, i.e., wishful, similarity, and embodied identification (VanLooy et al., 2012); in this paper, we did not manipulate the different aspects of identification, which potentially could enhance or diminish the efficacy of customization. Using, for example, non-humanoid representations in a customization procedure might differentially affect identification and efficacy of a subsequent task. We plan to investigate different types of customization to explore the differential effects of identification types on intervention efficacy and to determine whether customization of other interface elements (beyond faces) can have the same benefits.

Second, the experimental context needs to be considered for the interpretation of our results; participants were recruited using a crowdsourcing platform and were paid for their participation, which creates a different experience than being exposed to attentional retraining as part of a therapy for participants who are in need of treatment.

Third, the attention bias modification task is only one potential digital intervention. Enhancing efficacy through customization needs to be tested across various interventions. Enhanced engagement through customization will particularly benefit interventions that rely on focused attention, such as cognitive tasks like the ABMT. How customization can integrate with other therapeutic approaches (e.g., CBT-based interventions) remains to be investigated. Our technique requires very little effort to implement but shows significant changes of efficacy. How to leverage volitional engagement to best increase the efficacy of a variety of interventions is a promising direction; however, more systematic research on the limits of customization for in-the-moment task engagement needs is required.

Fourth, our study does not distinguish whether the effects are a result of participants coping better with the negative mood induction post-training, or if the effects are a result of the training itself. The former would suggest that the participants developed better strategies to disconnect from maladaptive thought processes, whereas the latter suggests that training helps participants protectively shift attention away from negative cognitions.

Fifth, the presented study focuses on the effects of in-the-moment engagement. However, many tasks (including ABMT) require frequent repetitions, e.g., being used for 30 minutes daily. While previous work suggests that avatar customization has positive effects on long-term engagement (Birk & Mandryk, 2018), we did not specifically investigate the effects of avatar customization over the long-term in this study.

5.4.7 Conclusions

Computerized interventions delivered at scale have the potential to ease clinical demand and interventions accessible for those who do not qualify for traditional therapies, or cannot access or afford them. However, the efficacy of computerized training cannot be sacrificed in service of a wider reach. Our results suggest that increasing in-the-moment engagement through interface customization and personalization can increase training efficacy.

We asked participants to complete online attentional bias modification training (ABMT) with a customized avatar or an assigned generic avatar. ABMT helps people shift their attention away from negative stimuli and has been shown to increase affective resilience to a subsequent negative mood induction (e.g., rating gruesome images). Our results showed that a version of the ABMT using customized interface elements generated through avatar customization increased resilience to a subsequent negative mood induction, suggesting that avatar customization increases in-the-moment engagement, and subsequently training efficacy. Furthermore, the customization benefits were particularly pronounced for participants with low satisfaction of relatedness, who are most at risk for developing mental illness.

Digital interventions delivered at scale offer a promise of increasing the reach of mental health treatment to a greater number of people in a wider range of places. Our work shows that avatar customization may help to improve the efficacy of existing and future training programs delivered in-the-wild.

5.5 Summary of Manuscript D

Whereas the previously-presented research was focused on sustained engagement—be it in the short-term as presented in Manuscript A or in the medium-term as presented in Manuscript B and C—the study I presented in this Chapter was focused on designing for immediate improvements of training efficacy. I showed that avatar customization effects training efficacy—participants who customized an avatar and received ABMT training showed greater resilience to the negative mood induction. Taken together with previous chapters, I can now infer that avatar customization increased engagement and subsequent training efficacy. The finding that customization also increased the negative effect of the mood induction for participants who did not receive training supports this argument. However, a study designed to measure increases in engagement, e.g., by measuring changes in pupil dilation (Kahneman & Beatty, 1966), would be required to show the currently inferred link between avatar customization and engagement in the moment of use.

The results support the notion that it is important to study individual differences in the context of training—I could show that especially people with low satisfaction of relatedness benefited from the customization. Imagining for example better knowledge about how design strategies and individual differences are best combined, one could imagine applying motivational design elements based on individual need satisfaction, e.g., competence-satisfying interactions for those low in competence, and interactions that allow for connecting to a digital representation for those experiencing low satisfaction of relatedness.

There were several lessons to be learned from studying the efficacy of avatar customization on immediate task efficacy, i.e., lessons around using computerized cognitive tasks in a crowdsourced environment; lessons around inferring increased engagement from task-efficacy; and lessons around the use of negative mood inductions in crowdsourced studies. The results and the research design spawn additional research questions around the measurement of engagement in the moment of use. Simultaneously, my findings are encouraging, because they show that studying design iterations of computerized cognitive tasks is possible in crowdsourced online studies.

5.5.1 Lessons Learned

First, I learned that Computerized Cognitive Tasks can be studied online. Amazon Mechanical Turk workers stay engaged even in a boring and repetitive training task, such as the ABMT. Measures of control do not suggest that participants disengage and showing training effects in the expected direction also supports the appropriateness of studying CCTs in an online crowdsourced context.

Second, I learned that engagement might be a double-edged sword that increases task efficacy, but also heightens negative experiences. A perspective relevant for designing games, but even more relevant when studying the effects of engagement.

Third, I learned that individual differences matter for efficacy. Understanding the individual context of a person might have dramatic effects on efficacy in the future. Especially, the power of big data and machine learning might allow for precise predictions of user needs and tailoring of digital interventions to individual needs. Approaches that can, for example, be seen when observing how big technology companies such as Microsoft[1] or Google's Verily[2] address individual health.

Fourth, the manipulation of facial expressions was very subtle, but the resulting effects from the ABMT were in line with my expectations, suggesting that a) humans respond to avatar facial expressions similar to pictures of human facial expressions, and b) that such a subtle manipulation translates into the expected outcome. A finding that is important when envisioning, for example, embedded in-game training such as receiving good quests only from neutral or positive, but never from angry quest givers in an online role-playing game such as Elder Scrolls (Bethesda, 1994).

Fifth, in hindsight, not directly measuring attention and engagement allows me only to infer that avatar customization increased attention and engagement, and subsequent increases in task efficacy. This argument is supported by the increase in negative response to the mood induction by the group that customized an avatar but did receive attentional retraining. My previous work, however, gave enough grounding for the presented experimental design, but with respect to an iterative research approach, an interim step to investigate engagement in this particular context, as done in Manuscript C for engagement over time would have been beneficial. Developing a model which, for example, would make predictions about attention, flow, or immersion in the context of training and subsequent training effects would improve and advance research on the interface design of computerized cognitive tasks.

5.6 Relevance in Context

In the context of my thesis, the presented research addressed the effects a game-based design strategy (i.e., avatar customization) on training efficacy. I assume that avatar customization induced increased in-the-moment engagement, which was the mechanism that led to the observed improvement in training efficacy.

The work also extends my own research horizon by looking into potential design improvements of computerized cognitive therapy and provides insights into individual differences in training efficacy. A third novel aspect of this particular study are experiences with online induction methods, which are a fundamental instrument to research human behaviour under controlled conditions.

In summary, the presented results are of relevance and interesting for advancing the efficacy of CCTs, but in the context and the future of my own work, the experiences gathered during this project provided insights that allow me to engage further with design challenges affecting CCTs, gather a more detailed understanding of online mood inductions, and provided the grounds to work on design problems of CCTs iteratively in an online context—an outcome which will allow me to address the efficacy of technology-enabled services from a different perspective than the perspective taken in Manuscript A to Manuscript C.

[1] https://www.microsoft.com/en-us/research/research-area/medical-health-genomics/
[2] https://verily.com/

Chapter 6

Discussion & Conclusion

In my work, I investigated whether and how avatar customization fosters volitional engagement in Technology-enabled Services, with the goal of combating waning motivation, and increasing in-the moment training efficacy. Hence, my research contributes to overcoming motivational limitations of technology-enabled services for health and helps to unlock the potential of technology-enabled services to benefit users and reduce strain on health care systems.

In four manuscripts I have investigated four different aspects of the effects of avatar customization on volitional engagement and motivation. The studies I presented go beyond investigating self-reported experience or prospective behaviour but look into changes in actual behaviour with the goal to increase the external validity of my findings and to provide additional insights that go beyond self-reported changes in experience when providing design recommendations and further thoughts on the motivational advantages of avatar customization.

6.1 Summary by Manuscript

To provide an overview of the findings of each paper, I briefly present each paper's problem statement and a summary of the most important findings. I close each Manuscript summary with a brief assessment of the contribution of my finding to the problem.

6.1.1 Manuscript A

The problem addressed in Manuscript A is that researchers and designers of technology-enabled services lack well-evaluated inherently-motivating design strategies that have been shown to translate into behaviour. By evaluating the effect of avatar customization on player engagement, I presented evidence that identification with a representation is associated with increased play experience and behavioural engagement in the short-term. I have shown that the motivational effects of avatar customization translate into increased behaviour in an online scenario, suggesting that avatar customization has potential to be a suitable motivational design strategy to alter behavioral engagement with technology-enabled services.

While there are still plenty of different design strategies to be evaluated, the presented research addresses the problem in three different ways. First, the research shows by example how a game-based motivational design strategy and its translation into behaviour can be researched online. Second, the research highlights that game-based motivational design strategies might not have a direct effect on behaviour but are mediated by a third variable—in this case identification. Third, the research provides evaluatedidence that behavioural changes resulting from design can be researched online—an important message for researchers and designers interested in iterative approaches or in ways of reducing time of recruitment and data processing.

6.1.2 Manuscript B

The problem addressed in Manuscript B is that the stability of effects of identification resulting from avatar customization are only understood in the short-term and it is unknown whether external rewards thwart or diminish the effect. Building on the results presented in Manuscript A, I used identification with an avatar as a proxy for motivation in an 11-day long study. I aimed to investigate the effect of participants' identification on waning motivation and the effect of an anticipated reward—delivered on day 8—on people differentiated by their motivation to engage in a daily training game. The findings are in-line with a known psychological effect, i.e., the overjustification effect, and show that motivation induced through avatar customization is responsive to reward delivery. While in Chapter 2, I established that studying game-based design strategies online is feasible, in Manuscript B I was able to address the effects that payment has on motivation in an online task. The conducted research provided valuable insights which are by themselves interesting, but really came to shine in subsequent studies. When designing the study described in Manuscript B, I would have lacked experience with the effects of payment and likely not considered important factors when delivering online payment while assuring the integrity of the research results. The research presented in Manuscript B also gave me first insights into how to design for the medium-term; I gained for example insights into instructions writing for multiple days, the overall feasibility to perform tasks over time on Amazon Mechanical Turk, and insights into technical requirements—for example, I gained experience on how to design the data collection system when presenting a task daily.

6.1.3 Manuscript C

The problem addressed in Manuscript C is that avatar customization shows promising effects when avatars are displayed either embedded into a game or into a game-based training application, but it remains unclear if avatar customization can facilitate sustained engagement with a non-game-based training task over time. Based on the lessons I learned presented in Chapters 2 and 3, I investigated the effects of customization on identification and engagement over time in a non-game-based training exercise. The results show that customizing an avatar decreases attrition in a non-game-based training exercise, especially pre- and post-notification. The notification phase shows a strong effect on login behaviour, however even during the notification phase, we see a tendency emerging over time, suggesting that participants who customized their avatar show less attrition under notifications compared to those who participated with an assigned avatar.

While Manuscript A and B provide fundamental insights into establishing the effects of avatar customization on engagement and look into the effects of rewards on motivation in an online context, Manuscript C connects previous findings with practically relevant changes in usage of a training task delivered online. The research shows that avatar customization indeed increases sustained engagement, and provides evidence that customization is not only effective in game-based tasks, but the motivational effects also translate into non-game-based training tasks—an important consideration when asking if game-based motivational design strategies such as avatar customization can improve the efficacy of technology-enabled services for health at scale.

6.1.4 Manuscript D

The problem I addressed in Manuscript D is that avatar customization shows promising effects on engagement in the short-term and long-term, but the effects of avatar customization on immediate training efficacy remain unclear. While research presented in Chapters 2 to 3 was focused on engagement—be it in the short-term as presented in Manuscript A or in the medium-term as presented in Manuscript B and C—the study presented in Manuscript D is focused on designing for immediate improvements in training efficacy. I showed that avatar customization affects training efficacy—participants who customized an avatar and received ABMT training showed greater resilience to a negative mood induction.

In the work presented in Manuscript D, I explored unfamiliar terrain by applying motivational design strategies to directly affect task efficacy. I have gained insights into the process of measuring task efficacy, interpreting pre-post effects after a mood induction, and about the effects of manipulating facial expression of avatars. In summary, I could show that efficacy can be increased, and that engagement needs to be carefully considered, because not only positive, but also negative effects are amplified.

6.1.5 Contextualization in Theory

While my findings might be relevant in the context of other theories such as goal-setting (i.e., avatars as agents taking on tasks like quest in World of Warcraft) or the development of habits (i.e., increasing habitual behaviour through avatar customization as suggested in Section 4.4.1), the exhibited experience and behaviour can best be explained by Self-Determination Theory.

Acting under our own volition, experiencing choice, and exploration (i.e., trying out what we like) are all experiences provided when customizing a representation and consequently result in increased engagement, as theory would suggest, and as I can demonstrate in my research. Within Self-Determination Theory, it is important to note that play experience and displayed behaviour closely co-vary as the findings in Chapter 2, 3, and 4 suggest. These findings imply that when applying Self-Determination Theory in the context of game, we can to a fair degree infer behaviour based on self-reported experience.

Considering that the studies presented in Chapter 4 and 5 have a high level of external validity, it is fair to assume that the theoretical consideration around Self-Determination Theory as presented in Section 1.4.1, translates into a practically relevant context, i.e., technology-enabled services. Finding a high convergence between theoretical knowledge and practical applicability renders research on game-based motivational design strategies relevant in the context of the described global healthcare gap, because it shows that individual strategies matched to theoretical constructs can be extracted and applied out of context.

That game-based motivational strategies are a viable tool to increase engagement in, for example, tasks for training, is further supported by the gradually increasing abstraction in my research from a full game to a tedious attentional retraining task—i.e., moving from an infinite runner game to the ABMT task. These findings show that SDT-based game-design might have potential to be motivationally relevant beyond the limits of video games, potentially making game-based design strategies even applicable in the clinical context (see for example Akili (Akili Interactive Labs, Inc., 2018)).

6.1.5.1 Levels of Motivation

In the process of my research, the relevance of the motivational level affected by avatar customization became more and more relevant. Hence, I introduced the SIMS in Manuscript C. From the data I have presented it is not clear how avatar customization affects motivation. Does avatar customization genuinely increase intrinsic motivation or does the technique increase lower levels of motivation, allowing a person who usually would experience their task engagement as externally regulated shift the experience more towards a regulation a person can identify with?

While there are open questions about how levels of motivation are affected by avatar customization, my research also provides insights by showing that enjoyment and effort are affected by avatar customization—both constructs suggest that avatar customization indeed might increase intrinsic motivation independent of regulatory factors, but further research would be required to come to a conclusive answer.

In summary, my findings contribute directly to our understanding of Self-Determination Theory within user experience research. In addition, my work reveals open questions about the motivational structure underlying the engaging effects of avatar customization. However, my findings are encouraging that SDT-based interface modifications are a promising approach to further investigate game-based interface modifications and game-based design strategies.

6.1.6 Conclusion of Manuscript Summaries

In combination, the four presented manuscripts contribute to a better understanding of the use of game-based strategies. They specifically contribute to a better understanding of how avatar customization can be used to foster volitional engagement in technology-enabled services. My work investigates problems of engagement from different temporal perspectives, i.e., short-term, long-term, and in-the-moment, gradually moving from game-based environments to tasks, with increasingly more realistic study conditions for technology-enabled services. While my work contributed directly to the research problem I intended to address, there are many indirect lessons I have learned from conducting research on game-based motivational design strategies in a crowdsourcing context.

6.2 Lessons Learned

Across all manuscripts there are several lessons learned about internal versus external validity, avatar customization and identification, payment plans, working with Amazon Mechanical Turk workers, and the validity of Amazon Mechanical Turk research, that are applicable beyond the individual studies and potentially inform future research geared towards the investigation of the effect of motivational design strategies on online behaviour.

6.2.1 Internal vs. External Validity

Addressing the limitation of experimental scenarios, Jim Blascovich talks about the mundane realism of experimental situations (Blascovich et al., 2002), i.e., the similarity of the experiment to an everyday situation. He argues that ideally the experimental situation increases participant engagement and therefore a participant's sensitivity to the experimental manipulation. Researchers commonly have to face a trade-off between the realism of the experimental context and experimental control. Blascovich, however, argues that immersive virtual environments provide the technology to simulate realistic situations in virtual environments with high levels of experimental control. Using game-based research paradigms allowed me to balance the mundane realism of the situation, i.e., playing a game, with the need for experimental control.

Because my research is geared to improve the real-world application of technology-enabled services for health, balancing internal and external validity was most important to be relevant. While the results needed to be internally valid, they also needed to be applicable outside of the experimental situation, i.e., as an interface manipulation applied within technology-enabled services.

To achieve potential applicability of results, one of the most challenging aspects when researching game-based designs strategies is to balance realistic play experience with the requirements of a rigidly designed study protocol. Because my research aimed to improve engagement with technology-enabled services and measurable improvements when applied outside of the research context, it was crucial for me to investigate fundamental aspects of avatar customization and motivation, while assuring that these findings translate into practical relevance.

Manuscript A to C established a clear path by reproducing the same effect in increasingly more complex research scenarios: 1) by investigating the effect of avatar customization on behaviour directly, 2) by gathering additional insights about a crucial component, i.e., reward delivery, and finally 3) by investigating the effects of avatar customization in a research approximation to application use. Manuscript D branches off by beginning a new line of research that investigates the direct impact of avatar customization on task efficacy. However, in all contributions, it is visible that research on the effects of interface manipulations on user experience benefits from gradual investigation moving from fundamental questions such as "Does avatar customization increase engagement?" to more applied questions such as "Can avatar customization be applied to combat attrition in digital training tasks?".

In summary, video games provide an environment that allows for the balancing of experimental control with the mundane realism of play. In addition to being concerned about the realism of the games used in my study and the context in which the studies were performed, I aimed to emphasize the tension between experimental control and realism also in my overarching design by increasing the realism of the experimental situation from study to study with the goal of showing that a fundamental effect translates into a research approximation of actual application use.

6.2.2 Avatar Customization and Identification

Across all studies, I learned several lessons about avatar customization as an induction method for volitional engagement. As my research demonstrates, avatar customization is a viable method to increase user engagement. Across studies, I could demonstrate that identification also appears when just watching a video of someone creating a representation. Therefore, it seems to be experimentally sounder to compare against a control condition that doesn't facilitate identification at all, e.g., by just assigning an avatar (see Manuscript A). Next to increases in autonomy, I also found increases in relatedness, suggesting that participants who created an avatar felt more connected to their representation—a result that might be of interest in future research investigating relatedness as a design principle.

There are also open questions about avatar customization that might solicit future research. For example, it is unclear how the graphical fidelity of a representation might affect engagement. Research on graphical fidelity suggests that in casual games, low graphical fidelity is appropriate, but more complex and demanding mechanics require a higher resolution (Gerling et al., 2013). It is also unclear how important the shape of the avatar itself is. In my research, I only investigated humanoid representations, but shapes of fantasy characters like dwarfs, lizards, werewolves, or even dragons might also have effects of players experience of customization and subsequent engagement. Finally, in all presented manuscripts, the avatar is part of the presented paradigm in one or the other way. In Manuscript A, the avatar is presented as the player's agent in an infinite runner. In Manuscript B, the avatar enacts the responses to go-no-go stimuli in a game setting. In Manuscript C, the breathing indicator is displayed on the avatar to suggest that the avatar enacts the breathing and the participants should follow with the same pace. In Manuscript D, the avatar's facial expressions are integrated into the Attention Bias Modification Training—arguably the player is here most removed from directly interacting with the avatar, but the avatar is still embedded in the task.

In the future, it would be of value to investigate if presenting customized avatars adjacent to training tasks would increase user engagement and task efficacy, because being able to decouple task and avatar presentation would increase the number of use cases in which avatars could be added as an interface manipulation.

6.2.3 Payment Plans

Incentivizing participants is common practice when researching human experience and behaviour. While the time of participants needs to be valued, attracting participants is also important to assure the success of a research project. In motivational research, "the need to incentivize" creates an experimental situation which is difficult to disentangle—preferably we would investigate motivation without a secondary source of motivation; on the other hand, it is difficult to justify having a participant perform a task without being compensated for their time.

Online crowdsourcing platforms further change the research situation. In research in universities, research institutions, or online panels, participants are commonly recruited specifically to take on research tasks; whereas crowd workers can choose from a variety of different tasks and perform online tasks (not necessarily online experiments) frequently. The participants who partook in the presented studies have all at least participated in 300 tasks—some of them are simple image classification tasks, other are more complex studies, but all studies are computerized and are disseminated via Amazon Mechanical Turk. As a result of frequently engaging in online work, workers might have formed pre-assumptions about compensation levels, e.g., $5 per 30min, and compensation timing, e.g., 1-day after finishing a task. The result of forming presumptions about how tasks need to be incentivized and when they need to be paid for having completed them might potentially alter user behaviour.

While randomly assigning participants to conditions and statistically evaluating group differences compensates for many effects that might be caused by online incentives, there are additional considerations that need to be undertaken when designing for incentives in crowdsourcing platforms. For example, the system keeps track of outstanding payments and might function as a reminder of a specific task. Or the meta-chat about a specific experiment might come up on the forums workers gather on to inquire information about the quality of newly-posted tasks. There could also be cross chatter between MTurk workers about enjoyable or well-paid tasks. While all these factors should potentially balance out between conditions, there is a benefit of considering these hidden factors that can affect a study—in the worst case systematically.

Because payment plays a big role in considering which task to take, and because MTurk workers communicate with each other, it is beneficial to have as little time as possible between HIT creation, completion, and payment. If a task is open for a long time, the likelihood of chatter about the task increases and might result in negative effects for the study.

While daily HITs and daily payments are the most secure way of assuring adherence, in studies in which an investigation of attrition is required, a daily task would function as a notification and artificially increase participation as shown in Manuscript C. Unfortunately, the payment reminders of Amazon Mechanical Turk render a solution in which participants are paid at the very end of a multi-day study inadequate.

In summary, I learned over my thesis that while payment seems to be very straightforward, there are circumstances in which carefully considering the user's circumstances, the interface specifics on the user's end, and the research goal, help to design experiments with a higher probability of coming to valid conclusions to answer a research question.

6.2.4 Working with Amazon Mechanical Turk Workers

In my dissertation alone, I worked with more than 1000 unique workers. In total, I have worked with more than 5000 workers at the time of writing my dissertation. Because of the number of experiments conducted and the similarity between some of the experiments—e.g., the use of an avatar customizer—participants were often excluded from future experiments. While there are ways of circumventing a ban from certain experiments, e.g., by using a different account, the likelihood of deviant behaviour is relatively low—in general, my experience working with MTurk workers was very positive and I was often surprised by how reliable many workers are.

There are obvious advantages of working with the largest crowdsourcing platform, such as the mere number of potential workers, resulting in extremely high response times for a task. The experience of the workers themselves, who have often completed hundreds of tasks, shows to be another source to improve quality and avoid larger issues. Experienced workers have experience with interpreting instructions, identifying issues, and assuring that their technical setup is not a source for any major issues.

When running more complex experiments online, there are several things besides the number of available workers and how experienced the workers are that matter for the quality of the results: 1) A good standing with the MTurk community, which is for example captured through tools like Turkopticon or through discussion about popular studies. Ironically, failure can affect a rating of a requester positively when handled adequately. During my time as a PhD student, I made several mistakes like publishing a task on MTurk with the wrong web address. As a result, people could not access the task, but spent time trying. Communicating quickly about the issue, fixing the issue, and compensating participants with a bonus for their wasted time were responses that were respected in the worker community and consequently increased our requester rating. How valid the collected data would be is a different question, because the initial frustration might—depending on the research—affect the subsequent study. 2) Adequate payment increases task pick-up as well, specifically when tasks are more difficult to accomplish, require special hardware such as a web camera, or in tasks that confront workers with uncomfortable material. 3) Workers are willing to follow complex instructions, which can be used to improve the quality of collected data. We could show for example that we can improve the quality of heart rate extracted from video feeds gathered via MTurk when workers were instructed about best practices of video recording (Muender et al., 2016).

While a good standing with the community, adequate payment, and clear instructions help to get good results fast, there are also a few pitfalls that one might want to avoid when working with MTurk workers: 1) The number of tasks and the percentage of approved work give workers access to different tasks, because requesters can define what levels of prior performance are acceptable for the workers they hire. As a result, workers are very concerned about their approval rating. Workers are aware that requesters monitor their performance, but get rightfully alarmed when their work gets rejected, especially when they felt they stayed within the performance margin, e.g., time or number or errors. Consequently, it is important to be clear about rejection reasons, to document why work was rejected, and to provide workers with means to assure that they fulfill expectations. It is frustrating for workers and requesters to have spent time on a task or on arguing a reject. 2) Not being clear with instructions has consequences for the experimental quality. For example, when one condition is not explained properly, and workers get in touch with the researcher to ask for qualification. Especially when the clarification question is systematic, there is a good chance that the effects of contacting the researcher will be reflected in the collected data—if there are potential ramifications depends on the research question. 3) File sizes can be an issue depending on how large the sizes are, and which region is targeted. My experiments were all conducted in the US and single files larger than 100mb were shown to cause frustration in participants due to slow load times.

In summary, working with workers on Amazon Mechanical Turk has been a pleasant experience and when workers are valued for their work, even difficult tasks can be easily completed online. However, researchers need to be aware of the environment workers are operating in to avoid frustration on both ends.

6.2.5 Implications for the Design of TES

My research suggests that integrating avatar customization as a game-based interface design strategy has positive effects on engagement over time and in the moment. My findings also show that avatar customization positively affects attitudes towards an online exercise (see Table 4.1). The findings across my studies show mainly beneficial effects of avatar customization on engagement, suggesting that TES designers should consider avatar customization as a viable strategy to increase retention and combat attrition.

My research is also encouraging for designers interested in other customization strategies, e.g., program interface customization. Beyond customization, my research suggests that theory-based research on interface elements has applicable implications that might change how the design of TES is approached.

Going beyond avatar customization, my research should encourage designers of TES to explore other strategies such as the implementation of competition or social closeness in games. The crowdsourcing approach I took highlights how these strategies could be explored iteratively with quick turnaround and a high throughput.

I hope that my research will enable TES designers to believe in the motivational potential of avatar customization and envision novel approaches to improve engagement and adherence with their effective applications. For example, taking avatar as a central element in a digital health ecosphere such as Intellicare (Mohr, Tomasino, et al., 2017) might reveal that interacting with the same avatar across applications might have added value for the individual applications.

6.3 Enabling Future Research

While there are several ways in which my dissertation might spawn future research, it directly enables future research on game-based motivational design strategies and might expend the applicability of online crowdsourcing paradigms.

6.3.1 Game-based Motivational Design Strategies

My work enables theory-driven research on game-based motivational design strategies, potentially enabling two direct lines of research. Future research could, for example, focus on exploring autonomy-based design strategies and investigating the effects of other choice paradigms such as path selection in a navigational game or narrative choice in a story-driven game. Alternatively, interactions that emphasize volitional engagement—for example by resistance, e.g., only after five attempts the user can go the path that proposes a specific training—could be investigated.

Another viable path would be to explore other options to facilitate engagement. Following self-determination theory, game-based motivational strategies that foster relatedness or competition would be obvious choices. However, following the research path described in my thesis, it would be crucial to establish behavioral effects first, because early research on self-reported desire for behavioural change has been shown to rarely translate into action (Ajzen, 1991).

6.3.2 Expanding Application Areas for Online Crowdsourcing Study Paradigms

The main contribution of my research is in establishing avatar customization as a game-based design strategy and showing how the motivational gain is relevant in an applied context, in particular in the context of decreasing motivation over time in technology-enabled services and increased task efficacy through increasing in-the-moment engagement. However, the research framework, i.e., extracting game-based motivational strategies, investigating their direct effect on behaviour, and investigating their effect on a relevant outcome variable in a more applied context, has potential for future research by enabling quick iterative design studies and procedures that allow researchers to quickly come to conclusions and cross validate findings while avoiding the usual overhead of lab-based research.

Because online studies require predefined instructions, follow the same protocols reliably, and create datasets of the same internal structure, they can be created as research blue prints that can be tested and evaluated for their validity in self-produced replications or as end-to-end open source research paradigms by other research groups. Because the produced data is also standardized, scale processing and analysis protocols can be created a priori, which allows researchers with minor levels of time investment to re-run an experiment with the goal of replicating prior findings. Open-source analysis protocols could be reviewed similarly to open-source software.

Another way to look at online studies is from an iterative design perspective. For example, the research paradigm described in Manuscript A shows increases in behavioural engagement for those who identify with a representation. However, no group differences are shown. Iterating the control group to produce group differences would have been feasible to quickly find a clear comparison condition to customizing an avatar. Or if the goal would have been to maximize identification, iterating the avatar customizer's body shapes would have required few resources to implement and test.

6.4 Causal Effect Chain: From Customization to Treatment Outcomes

In my thesis I present four manuscripts that investigate different aspects of a causal effect chain providing evidence and argument that avatar customization increases self-reported identification with an avatar, which increases self-reported motivation, leading to increased behavioural engagement (i.e., an increase in behaviour), and presumable increased engagement in the moment (i.e., increased focus), leading to heighted exposure over time or qualitatively higher exposure through increased focused in the moment of use. Both pathways positively affect treatment outcomes of technology-enabled services for health. The effect chain itself and where each manuscript contributes to results that support this chain are displayed in Figure 1.

6.4.1 Contribution Manuscript A

Manuscript A establishes that customization affects identification and shows that *customization* increases self-reported *motivation*. Participants who identified with their avatar showed increased enjoyment and effort. Participants showed increased play behaviour, and the combination of increased immersion and prolonged playtime for those who identified more with their avatar suggests that these participants were engaged more. However, a direct measure of engagement besides immersion would have been a valuable addition to understand if participants were also more engaged in the moment of playing the game. Future work could analyze the keystroke dynamics (i.e., 'M' and 'N' presses) to get a more in-depth understanding of participants' engagement in-the-moment of play.

6.4.2 Contribution Manuscript B

Manuscript B replicates findings from Manuscript A, which show that avatar *customization* facilitates *identification* and supports that identifying with an avatar increases *motivation*. Participants who identified more showed increased effort and enjoyment after customizing an avatar. While the manuscript replicates results found in Manuscript A, it also extends our understanding of avatar customization as a motivational design-strategy by taking a close look into the effects of identification on experience and performance over time. The manuscript further investigates how external rewards affect sustained *interactions* in our study design by deploying a game-based task (i.e., go-no/go) and therefore provides further insights into the pathway between avatar customization, identification, motivation, and interactions over time. Similar to Manuscript A, engagement can only be inferred, but is additionally supported by showing expected differential effects of introducing an external reward on motivation and task performance.

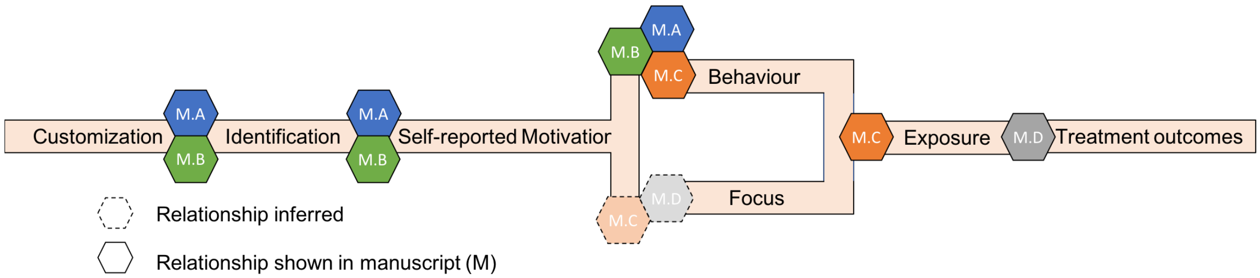

Figure 6.1: Diagram displaying the investigated causal effect chain; links investigated by Manuscripts A-D are indicated by coloured hexagons.

6.4.3 Contribution Manuscript C

Manuscript C builds on the findings of Manuscript A and B, specifically on the effects of avatar customization on motivation. However, in Manuscript C, we further investigate the effects of avatar customization on behavior and experience by studying login behavior and self-reported engagement; engagement can only be inferred through the check-all-that apply questions (e.g., increases in interest in stress management) presented at the final day of the three week-long study. The study further shows increased *exposure* to a technology-enabled service (i.e., breathing exercise) during the study as a causal result of customizing an avatar.

6.4.4 Contribution Manuscript D

While Manuscript A, B, and C provide support and explanation for the pathway between avatar customization and exposure, there is, however, no direct investigation of potential treatment outcomes. Manuscript D is a first investigation of potential treatment effects in the short time. While engagement was not directly measured, we argue based on the presented results that avatar customization increases in-the-moment engagement and the quality of *exposure* resulting in improve *treatment outcomes*.

6.4.5 Summary of Contributions

Investigating my contribution in the context of the causal effect chain, I made progress on solidifying the relationship between several constructs. First, my work established the link between customization and identification (Manuscript A and B) and provides support for the link between motivation and frequency of interactions (Manuscript A and B). I further investigated the effects of customization on interaction, engagement, and resulting exposure in Manuscript C. Manuscript D presents a first investigation of a single computerized cognitive task (i.e., ABMT).

Taking a high-level perspective, the presented work provides several points of evidence for the relationship between customization and motivation and how motivation translates into increased interactions (Manuscript A, B, C). The support for increased in-the-moment engagement is only by argument and leaves room for improvement. Studies measuring engagement, e.g., through pupillometry or electroencephalography, would allow us to strengthen the argumentative link between avatar customization motivation and engagement.

Manuscript C provides support for the relationship between customization and increased interaction, resulting in increased exposure. However, it remains unclear if increased exposure is also linked to improved treatment outcomes. Embedding avatar customization within the context of a established technology-enabled service for health, such as SilverCloud (Richards et al., 2015), would allow us to test for the effects on the link between exposure and treatment outcomes. Manuscript D investigates treatment outcome effects, but in a very specific context (i.e., ABMT) and the in-the-moment engagement is only inferred; additional work could be done on the benefits of avatar customization in different service contexts to generate knowledge about where applying avatar customization as an interface manipulation would be most beneficial. While Manuscript A, B, and C investigate interaction profiles and Manuscript C and D provide a foundation to talk about engagement, it is unclear whether increased exposure or exposure quality (i.e., focus) affect treatment outcomes differently and which strategy is more beneficial or if both strategies could be combined.

6.5 Conclusion

Current technology-enabled services for health underutilize inherently motivating design strategies, thwarting the sustainable and effective implementation of such services in health care systems. To provide a research foundation with the aim of increasing the potential utilization of motivational design strategies in technology-enabled services for health, I investigated whether and how avatar customization fosters volitional engagement in technology-enabled services. I showed that avatar customization fosters volitional self-reported and behavioral engagement. I further showed that behavioural engagement over time resulting from identifying with an avatar is responsive to financial rewards. Further, I could show that avatar customization reduces attrition in a 3-week long online task and in-the-moment engagement with a computerized cognitive task. My thesis shows that avatar customization is a viable and effective motivational design strategy to improve user engagement with technology-enabled services for health.

References

Abuhamdeh, S., Csikszentmihalyi, M., & Jalal, B. (2014). Enjoying the possibility of defeat: Outcome uncertainty, suspense, and intrinsic motivation. *Motivation and Emotion*, *39*(1), 1–10. doi: 10.1007/s11031-014-9425-2

Ahtinen, A., Mattila, E., Välkkynen, P., Kaipainen, K., Vanhala, T., Ermes, M., . . . Lappalainen, R. (2013, jul). Mobile Mental Wellness Training for Stress Management: Feasibility and Design Implications Based on a One-Month Field Study. *JMIR mhealth and uhealth*, *1*(2), 1–13. doi: 10.2196/mhealth.2596

Ajzen, I. (1991). The theory of planned behavior. *Orgnizational Behavior and Human Decision Processes*, *50*, 179–211. doi: 10.1016/0749-5978(91)90020-T

Almuhimedi, H., Schaub, F., Sadeh, N., Adjerid, I., Acquisti, A., Gluck, J., . . . Agarwal, Y. (2015). Your Location has been Shared 5,398 Times!: A Field Study on Mobile App Privacy Nudging. In *Proceedings of the 33rd annual acm conference on human factors in computing systems - chi '15* (pp. 787–796). New York, New York, USA: ACM Press. doi: 10.1145/2702123.2702210

Althoff, T., White, R. W., & Horvitz, E. (2016, dec). Influence of Pokémon Go on Physical Activity: Study and Implications. *Journal of Medical Internet Research*, *18*(12), e315. doi: 10.2196/jmir.6759

Amabile, T. M. (1993). Motivational synergy: Toward new conceptualizations of intrinsic and extrinsic motivation in the workplace. *Human Resource Management Review*, *3*(3), 185–201. doi: 10.1016/1053-4822(93)90012-S

Andersson, G., & Cuijpers, P. (2009). Internet-based and other computerized psychological treatments for adult depression: A meta-analysis. *Cognitive Behaviour Therapy*, *38*(4), 196–205. doi: 10.1080/16506070903318960

Armitage, C. J., & Conner, M. (2001). Efficacy of the Theory of Planned Behaviour: a meta-analytic review. *The British journal of social psychology / the British Psychological Society*, *40*(Pt 4), 471–499. doi: 10.1348/014466601164939

Atkinson, J. W., & Litwin, G. H. (1960). Achievement motive and test anxiety conceived as motive to approach success and motive to avoid failure. *Journal of Abnormal and Social Psychology*, *60*(1), 52–63. doi: 10.1037/h0041119

Bailey, B. P., & Konstan, J. A. (2006). On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. *Computers in Human Behavior*, *22*(4), 685–708.
doi: 10.1016/j.chb.2005.12.009

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology*, *2*(1), 21–41.
doi: 10.1111/1467-839X.00024

Baranowski, T., Baranowski, J., Cullen, K. W., Marsh, T., Islam, N., Zakeri, I., ... DeMoor, C. (2003, jan). Squire's Quest! *American Journal of Preventive Medicine*, *24*(1), 52–61.
doi: 10.1016/S0749-3797(02)00570-6

Barry, T. J., Vervliet, B., & Hermans, D. (2015). An integrative review of attention biases and their contribution to treatment for anxiety disorders. *Frontiers in Psychology*, *6*(July), 1–15.
doi: 10.3389/fpsyg.2015.00968

Basak, C., Boot, W. R., Voss, M. W., & Kramer, A. F. (2008). Can Training in a Real-Time Strategy Videogame Attenuate cognitive decline in older adults? *Psychology and aging*, *23*(4), 765–777.
doi: 10.1037/a0013494.Can

Baumeister, H., Reichler, L., Munzinger, M., & Lin, J. (2014, oct). *The impact of guidance on Internet-based mental health interventions - A systematic review* (Vol. 1) (No. 4). Elsevier.
doi: 10.1016/j.invent.2014.08.003

Baumeister, R. F., & Leary, M. R. (1995, may). The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychological bulletin*, *117*(3), 497–529.
doi: 10.1037/0033-2909.117.3.497

Bean, J. P., & Eaton, S. B. (2000). A psychological model of college student retention. *Reworking the student departure puzzle*, *1*, 48–61.

Beggan, J. K. (1992). On the Social Nature of Nonsocial Perception: The Mere Ownership Effect. *Journal of Personality and Social Psychology*, *62*(2), 229–237.

Bell, C. (2012). Designing for Friendship: Shaping Player Relationships with Rules and Freedom. In *Games developer conference 2012.*

Bennett, G. G., & Glasgow, R. E. (2009). The Delivery of Public Health Interventions via the Internet: Actualizing Their Potential. *Annual Review of Public Health*, *30*(1), 273–292.
doi: 10.1146/annurev.publhealth.031308.100235

Berge, Z. L., & Huang, Y.-P. (2004). A Model for Sustainable Student Retention: A Holistic Perspective on the Student Dropout Problem with Special Attention to e-Learning. *Deosnews*, *13*(5), 26.

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). *Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk* (Vol. 20). Oxford University PressSociety for Political Methodology.

doi: 10.2307/23260322

Bessière, K., Seay, A. F., & Kiesler, S. (2007, aug). The ideal elf: identity exploration in World of Warcraft. *CYBERPSYCHOLOGY & BEHAVIOR*, *10*(4), 530–536. doi: 10.1089/cpb.2007.9994

Birk, M., & Mandryk, R. L. (2013). Control your game-self: effects of controller type on enjoyment, motivation, and personality in game. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, 685–694. doi: 10.1145/2470654.2470752

Birk, M. V., Atkins, C., Bowey, J. T., & Mandryk, R. L. (2016). Fostering Intrinsic Motivation through Avatar Identification in Digital Games. In *Proceedings of the 2016 chi conference on human factors in computing systems - chi '16* (Vol. 55, pp. 2982–2995). doi: 10.1145/2858036.2858062

Birk, M. V., Fries, M. A., & Mandryk, R. L. (2017). Age-Based Preferences and Player Experience: A Crowdsourced Cross-sectional Study. *Proceedings of the ACM SIGCHI Annual Symposium on Computer-Human Interaction in Play 2017*.

Birk, M. V., & Mandryk, R. (2018). Combating Attrition in Digital Self-Improvement Programs using Avatar Customization. In *Proceedings of the 2018 chi conference on human factors in computing systems - chi '18* (Vol. 2018-April, pp. 1–15). New York, New York, USA: ACM Press. doi: 10.1145/3173574.3174234

Birk, M. V., Mandryk, R. L., & Atkins, C. (2016). The Motivational Push of Games: The Interplay of Intrinsic Motivation and External Rewards in Games for Training. In *Proceedings of the 2016 annual symposium on computer-human interaction in play (chi play)* (pp. 291–303). doi: 10.1145/2967934.2968091

Birk, M. V., Mandryk, R. L., Miller, M. K., & Gerling, K. M. (2015). How Self-Esteem Shapes our Interactions with Play Technologies. In *Proceedings of the 2015 annual symposium on computer-human interaction in play - chi play '15* (pp. 35–45). doi: 10.1145/2793107.2793111

Birk, M. V., Toker, D., Mandryk, R. L., & Conati, C. C. (2015). Modeling motivation in a social network game using player-centric traits and personality traits. In *International conference on user modeling, adaptation, and personalization – umap'15* (Vol. 9146, pp. 18–30). doi: 10.1007/978-3-319-20267-9$_2$

Blascovich, J., Loomis, J., Beall, A. C., Swinth Kimberly R., Hoyt, C. L., & Bailenson, J. N. (2002). TARGET ARTICLE: Immersive Virtual Environment Technology as a Methodological Tool for Social Psychology. *Psychological Inquiry*, *13*(2), 103–124. doi: 10.1207/S15327965PLI1302$_0$1

Blatt, S. J., & Zuroff, D. C. (1992, jan). Interpersonal relatedness and self-definition: Two prototypes for depression. *Clinical Psychology Review*, *12*(5), 527–562. doi: 10.1016/0272-7358(92)90070-O

Bogdanova, Y., Yee, M. K., Ho, V. T., & Cicerone, K. D. (2016). Computerized Cognitive Rehabilitation of

Attention and Executive Function in Acquired Brain Injury. *Journal of Head Trauma Rehabilitation*, *31*(6), 419–433. doi: 10.1097/HTR.0000000000000203

Bogost, I. (2011a). *Gamification is Bullshit.*

Bogost, I. (2011b). *Persuasive Games: Exploitationware.*

Boot, W. R., Kramer, A. F., Simons, D. J., Fabiani, M., & Gratton, G. (2008, nov). The effects of video game playing on attention, memory, and executive control. *Acta Psychologica*, *129*(3), 387–398. doi: 10.1016/j.actpsy.2008.09.005

Bowey, J. T., Birk, M. V., & Mandryk, R. L. (2015). Manipulating Leaderboards to Induce Player Experience. In *Proceedings of the 2015 annual symposium on computer-human interaction in play - chi play '15* (pp. 115–120). doi: 10.1145/2793107.2793138

Bradley, M. M., & Lang, P. J. (1994, mar). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49–59. doi: 10.1016/0005-7916(94)90063-9

Braithwaite, A. (2018). WoWing Alone: The Evolution of "Multiplayer" in World of Warcraft. *Games and Culture*, *13*(2), 119–135. doi: 10.1177/1555412015610246

Brockmyer, J. H., Fox, C. M., Curtiss, K. A., McBroom, E., Burkhart, K. M., & Pidruzny, J. N. (2009). The development of the Game Engagement Questionnaire: A measure of engagement in video game-playing. *Journal of Experimental Social Psychology*, *45*(4), 624–634 ST – The development of the Game Engageme. doi: 10.1016/j.jesp.2009.02.016

Buchanan, T. W., Al'Absi, M., & Lovallo, W. R. (1999). Cortisol fluctuates with increases and decreases in negative affect. *Psychoneuroendocrinology*, *24*(2), 227–241. doi: 10.1016/S0306-4530(98)00078-X

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk. *Perspectives on Psychological Science*, *6*(1), 3–5. doi: 10.1177/1745691610393980

Burns, M. N. M., Begale, M., Duffecy, J., Gergle, D., Karr, C. J., Giangrande, E., & Mohr, D. C. (2011, aug). Harnessing context sensing to develop a mobile intervention for depression. *Journal of Medical Internet Research*, *13*(3), e55. doi: 10.2196/jmir.1838

Cacioppo, S., Grippo, A. J., London, S., Goossens, L., & Cacioppo, J. T. (2015). Loneliness: Clinical import and interventions. *Perspectives on Psychological Science*, *10*(2), 238–249.

Canadian Mental Health Association. (2016). *Suicide Statistics - Canadian Mental Health Association, Toronto BranchCanadian Mental Health Association, Toronto Branch.*

Carlbring, P., Apelstrand, M., Sehlin, H., Amir, N., Rousseau, A., Hofmann, S. G., & Andersson, G. (2012). Internet-delivered attention bias modification training in individuals with social anxiety disorder - a double blind randomized controlled trial. *BMC Psychiatry*, *12*, 66. doi: 10.1186/1471-244X-12-66

Carlin, A. S., Hoffman, H. G., & Weghorst, S. (1997, feb). Virtual reality and tactile augmentation in the treatment of spider phobia: a case report. *Behaviour Research and Therapy*, *35*(2), 153–158. doi: 10.1016/S0005-7967(96)00085-X

Casey, B. J., Trainor, R. J., Orendi, J. L., Schubert, A. B., Nystrom, L. E., Giedd, J. N., ... Rapoport, J. L. (1997). A Developmental Functional MRI Study of Prefrontal Activation during Performance of a Go-No-Go Task. *Journal of Cognitive Neuroscience*, *9*(6), 835–847. doi: 10.1162/jocn.1997.9.6.835

Casler, K., Bickel, L., & Hackett, E. (2013, nov). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior*, *29*(6), 2156–2160. doi: 10.1016/j.chb.2013.05.009

Castronova, E. (2004, may). The price of bodies: A hedonic pricing model of avatar attributes in a synthetic world. *Kyklos*, *57*(2), 173–196. doi: 10.1111/j.0023-5962.2004.00249.x

Casual Games Association. (2012). *Social Network Games 2012, Casual Games Sector Report* (Tech. Rep.).

Chen, Z.-h., Liao, C. C. Y., Cheng, H. N. H., Yeh, C. Y. C., Chan, T.-W., Chen, Z.-H., ... Chan, T.-W. (2012). Influence of Game Quests on Pupils' Enjoyment and Goal-pursuing in Math Learning. *Journal of Educational Technology & Society*, *15*(2), 317–327.

Childers, T. L., Carr, C. L., Peck, J., & Carson, S. (2001, dec). Hedonic and utilitarian motivations for online retail shopping behavior. *Journal of Retailing*, *77*(4), 511–535. doi: 10.1016/S0022-4359(01)00056-2

Chisholm, D., Sweeny, K., Sheehan, P., Rasmussen, B., Smit, F., Cuijpers, P., & Saxena, S. (2016, may). Scaling-up treatment of depression and anxiety: a global return on investment analysis. *The Lancet Psychiatry*, *3*(5), 415–424. doi: 10.1016/S2215-0366(16)30024-4

Christensen, H., Griffiths, K. M., & Farrer, L. (2009). Adherence in internet interventions for anxiety and depression. *Journal of Medical Internet Research*, *11*(2), 1–16. doi: 10.2196/jmir.1194

Christensen, H., Griffiths, K. M., & Jorm, A. F. (2004). Delivering interventions for depression by using the internet: randomised controlled trial. *BMJ*, *328*(7434), 265–0. doi: 10.1136/bmj.37945.566632.EE

Christensen, H., Griffiths, K. M., & Korten, A. (2002). Web-based cognitive behavior therapy: Analysis of site usage and changes in depression and anxiety scores. *Journal of Medical Internet Research*, *4*(1), 29–40. doi: 10.2196/jmir.4.1.e3

Christensen, H., Griffiths, K. M., & Korten, A. (2004). Comparison of changes in anxiety and depression symptoms of spontaneous users and trial participants of a CBT website. *Journal of Medical Internet Research*, *6*(4), 46.

Christensen, H., & Mackinnon, A. (2006). The law of attrition revisited. *Journal of medical Internet research*, *8*(3).

Christoph, K., Dorothe, H., & Peter, V. (2009). The video game experience as ”true” identification: A theory of enjoyable alterations of players’ self-perception. *Communication Theory*, *19*(4), 351–373. doi: 10.1111/j.1468-2885.2009.01347.x

Chu, B. C., & Kendall, P. C. (2004). Positive association of child involvement and treatment outcome within a manual-based cognitive-behavioral treatment for children with anxiety. *Journal of Consulting and Clinical Psychology*, *72*(5), 821–829. doi: 10.1037/0022-006X.72.5.821

Clark, J. E., Lanphear, A. K., & Riddick, C. C. (1987, jan). The effects of videogame playing on the response selection processing of elderly adults. *Journal of gerontology*, *42*(1), 82–5.

Cohen, J. (2001). Defining Identification: A Theoretical Look at the Identification of Audiences with Media Characters. *Mass Communication & Society*, *4*(3), 245–264. doi: 10.1207/S15327825MCS0403

Collins, F. S., & Varmus, H. (2015, jan). A New Initiative on Precision Medicine. *New England Journal of Medicine*, *372*(9), 793–795. doi: 10.1056/NEJMp1500523

Connolly, T. M., Boyle, E. a., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education*, *59*(2), 661–686. doi: 10.1016/j.compedu.2012.03.004

Cooke, R., & Sheeran, P. (2004). Moderation of cognition – intention and cognition – behaviour relations : A meta-analysis of properties of variables from the theory of planned behaviour. *Journal of Social Psychology*, *43*, 159–186.

Cooper, S., Khatib, F., Treuille, A., Barbero, J., Lee, J., Beenen, M., . . . Players, F. (2010, aug). Predicting protein structures with a multiplayer online game. *Nature*, *466*(7307), 756–760. doi: 10.1038/nature09304

Cortese, S., Ferrin, M., Brandeis, D., Buitelaar, J., Daley, D., Dittmann, R. W., . . . Sonuga-Barke, E. J. (2015). Cognitive training for attention-deficit/hyperactivity disorder: Meta-analysis of clinical and neuropsychological outcomes from randomized controlled trials. *Journal of the American Academy of Child and Adolescent Psychiatry*, *54*(3), 164–174. doi: 10.1016/j.jaac.2014.12.010

Coyle, H., Traynor, V., & Solowij, N. (2015, apr). Computerized and Virtual Reality Cognitive Training for Individuals at High Risk of Cognitive Decline: Systematic Review of the Literature. *The American Journal of Geriatric Psychiatry*, *23*(4), 335–359. doi: 10.1016/j.jagp.2014.04.009

Crenshaw, N., & Nardi, B. (2014). What's in a name? Naming practices in online video games. In *Proceedings of the first acm sigchi annual symposium on computer-human interaction in play - chi play '14* (pp. 67–76). New York, New York, USA: ACM Press. doi: 10.1145/2658537.2658685

Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013, mar). Evaluating Amazon's Mechanical Turk as a Tool for Experimental Behavioral Research. *PLoS ONE*, *8*(3), e57410. doi: 10.1371/journal.pone.0057410

Cuijpers, P., Kleiboer, A., Karyotaki, E., & Riper, H. (2017, jul). *Internet and mobile interventions for depression: Opportunities and challenges* (Vol. 34) (No. 7). Wiley-Blackwell. doi: 10.1002/da.22641

Curran;, G. M., Bauer, M., Mittman, B., Pyne, J. M., & Stetler, C. (2012). Effectiveness- Implementation Hybrid Design: Combining Elements of Clinical Effectiveness and Implementation Research to Enhance Public Health. *Med Care*, *50*(3), 217–226. doi: 10.1097/MLR.0b013e3182408812.Effectiveness-implementation

Cutrell, E., Czerwinski, M., & Horvitz, E. (2001). Notification, Disruption, and Memory: Effects of Messaging Interruptions on Memory and Performance. *Conference on Human-Computer Interaction Interact 2001*(1999), 263–269.

Czerwinski, M., Horvitz, E., & Wilhite, S. (2004). A diary study of task switching and interruptions. In *Proceedings of the 2004 conference on human factors in computing systems - chi '04* (pp. 175–182). doi: 10.1145/985692.985715

Dale, L. P., White, L., Mitchell, M., & Faulkner, G. (2018, apr). *Smartphone app uses loyalty point incentives and push notifications to encourage influenza vaccine uptake.* Elsevier. doi: 10.1016/j.vaccine.2018.04.018

Daly, T. M., & Nataraajan, R. (2015, dec). Swapping bricks for clicks: Crowdsourcing longitudinal data on Amazon Turk. *Journal of Business Research*, *68*(12), 2603–2609. doi: 10.1016/j.jbusres.2015.05.001

Davidson, R. J., Kabat-Zinn, J., Schumacher, J., Rosenkranz, M., Muller, D., Santorelli, S. F., ... Sheridan, J. F. (2003, jul). Alterations in Brain and Immune Function Produced by Mindfulness Meditation. *Psychosomatic Medicine*, *65*(4), 564–570. doi: 10.1097/01.PSY.0000077505.67574.E3

De Grove, F., Breuer, J., Hsueh Hua Chen, V., Quandt, T., Ratan, R., & Van Looy, J. (2017, jan). Validating the Digital Games Motivation Scale for Comparative Research Between Countries. *Communication Research Reports*, *34*(1), 37–47. doi: 10.1080/08824096.2016.1250070

Deci, E. L. (1975). *Intrinsic Motivation.* Boston, MA: Springer US. doi: 10.1007/978-1-4613-4446-9

Deci, E. L., Koestner, R., & Ryan, R. M. (2001, jan). Extrinsic Rewards and Intrinsic Motivation in Education: Reconsidered Once Again. *Review of Educational Research*, *71*(1), 1–27. doi: 10.3102/00346543071001001

Deci, E. L., Koestner, R., Ryan, R. M., & Koestner, R. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, *125*(6), 627–668. doi: 10.1037/0033-2909.125.6.627

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of Goal Pursuits: of Behavior Human Needs and the Self-determination. *Psychological Inquiry*, *11*(4), 227–268. doi: 10.1207/S15327965PLI1104

Dennis, C.-L. (2003). Peer support within a health care context: a concept analysis. *International Journal of Nursing Studies*, *40*(April 2002), 321–332. doi: 10.1016/S0020-7489(02)00092-5

Dennis, T. A., & O'Toole, L. J. (2014, sep). Mental health on the go: Effects of a gamified attention-bias modification mobile application in trait-anxious adults. *Clinical Psychological Science*, *2*(5), 576–590. doi: 10.1177/2167702614522228

Depping, A. E., Mandryk, R. L., Johanson, C., Bowey, J. T., & Thomson, S. C. (2016). Trust Me. In *Proceedings of the 2016 annual symposium on computer-human interaction in play - chi play '16* (pp. 116–129). New York, New York, USA: ACM Press. doi: 10.1145/2967934.2968097

Deterding, S. (2016). Eudaimonic Design, or: Six Invitations to Rethink Gamification. In M. Fuchs, S. Fizek, P. Ruffino, & N. Schrape (Eds.), *Rethinking gamification* (pp. 305–331). meson press.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th international academic mindtrek conference on envisioning future media environments - mindtrek '11* (pp. 9–15). Tampere, Finland: ACM. doi: 10.1145/2181037.2181040

Dinakar, K., Lieberman, H., Chaney, A. J. B., & Blei, D. M. (2014). Real-time Topic Models for Crisis Counseling. *KDD DSSG Workshop*, 1–4.

Dishman, R. K. (1991). Increasing and maintaining physical activity and exercise. *Behavior Therapy*, *22*, 345–378. doi: 10.1016/S0005-7894(05)80371-5

Doherty, G., Coyle, D., Sharry, J., & St, J. J. (2012). Engagement with online mental health interventions: an exploratory clinical study of a treatment for depression. In *Proceedings of the 2012 acm annual conference on human factors in computing systems - chi '12* (pp. 1421–1430). doi: 10.1145/2207676.2208602

Donkin, L., Christensen, H., Naismith, S. L., Neal, B., Hickie, I. B., & Glozier, N. (2011). A systematic review of the impact of adherence on the effectiveness of e-therapies. *Journal of Medical Internet Research*, *13*(3), 1–19. doi: 10.2196/jmir.1772

Dorval, M., & Pépin, M. (1986, feb). Effect of Playing a Video Game on a Measure of Spatial Visualization. *Perceptual and Motor Skills*, *62*(1), 159–162.
doi: 10.2466/pms.1986.62.1.159

Dowsett, S. M., & Livesey, D. J. (2000). The development of inhibitory control in preschool children: Effects of "executive skills" training. *Developmental Psychobiology*, *36*(2), 161–174.
doi: 10.1002/(SICI)1098-2302(200003)36:2¡161::AID-DEV7¿3.0.CO;2-0

Drew, B., & Waters, J. (1986). Video games: Utilization of a novel strategy to improve perceptual motor skills and cognitive functioning in the non-institutionalized elderly. *Cognitive Rehabilitation*, *4*(2), 26–31.

Dring, C. (2015). *More money is spent on games than movies and music combined, says IHS — Games industry news — MCV.*

Ducheneaut, N., Wen, M., Yee, N., & Wadley, G. (2009). Body and Mind: A Study of Avatar Personalization in Three Virtual Worlds. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1151–1160.
doi: 10.1145/1518701.1518877

Dustman, R. E., Emmerson, R. Y., Steinhaus, L. A., Shearer, D. E., & Dustman, T. J. (1992, may). The Effects of Videogame Playing on Neuropsychological Performance of Elderly Individuals. *Journal of Gerontology*, *47*(3), P168–P171.
doi: 10.1093/geronj/47.3.P168

Dye, M. W. G., Green, C. S., & Bavelier, D. (2009, jul). The development of attention skills in action video game players. *Neuropsychologia*, *47*(8-9), 1780–9.
doi: 10.1016/j.neuropsychologia.2009.02.002

Ekman, P., & Friesen, W. V. (1977). Facial action coding system.

Entertainment Software Association. (2014). *Essential Facts About the Computer and Video Game Industry* (Tech. Rep.).

Entertainment Software Association. (2017). *2017 Essential Facts About the Computer and Video Game Industry* (Tech. Rep.).

Entertainment Software Association, E. (2015). *Essential Facts about the Canadian Video Game Industry* (Tech. Rep.). Entertainment Software Association.

Etter, R., Costa, P. D., & Broens, T. (2006). A Rule-Based Approach Towards Context-Aware User Notification Services. *2006 ACS/IEEE International Conference on Pervasive Services*, 281–284.
doi: 10.1109/PERSER.2006.1652242

Eysenbach, G. (2005). *The law of attrition* (Vol. 7) (No. 1).
doi: 10.2196/jmir.7.1.e11

Festinger, L. (1954). A Theory of Social Comparison Processes. *Human Relations*, *7*(2), 117–140.
doi: 10.1177/001872675400700202

Field, A. (2013). *Discovering statistics using IBM SPSS statistics.* Sage.

Firth, J., Torous, J., Carney, R., Newby, J., Cosco, T. D., Christensen, H., & Sarris, J. (2018, jun). *Digital Technologies in the Treatment of Anxiety: Recent Innovations and Future Directions* (Vol. 20) (No. 6). Springer US.
doi: 10.1007/s11920-018-0910-2

Firth, J., Torous, J., Nicholas, J., Carney, R., Pratap, A., Rosenbaum, S., & Sarris, J. (2017). The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. *World Psychiatry*, *16*(3), 287–298.
doi: 10.1002/wps.20472

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research.* Reading, MA: Addison-Wesley.

Fogg, B. J. (2002). Persuasive Technology: Using Computers to Change What We Think and Do. *Ubiquity*, 5. doi: 10.1145/764008.763957

Franke, N., Schreier, M., & Kaiser, U. (2009). The "I Designed It Myself" Effect in Mass Customization. *Management Science*, *56*(1), 125–140.
doi: 10.1287/mnsc.1090.1077

Franke, N., Schreier, M., & Kaiser, U. (2010). The "I Designed It Myself" Effect in Mass Customization. *Management Science*, *56*(1), 125–140.
doi: 10.1287/mnsc.l090.1077

Frederick, C. M., & Ryan, R. M. (1995). Self-determination in sport: A review using cognitive evaluation theory. *International Journal of Sport Psychology*, *26*(1), 5–23.

Fullerton, T. (2014). *Game design workshop: a playcentric approach to creating innovative games.* Leiden, Netherlands: Taylor & Francis Ltd.

Gagne, M. (2003, oct). Autonomy Support and Need Satisfaction in the Motivation and Well-Being of Gymnasts. *Journal of Applied Sport Psychology*, *15*(4), 372–390.
doi: 10.1080/714044203

Gagné, M., & Deci, E. L. (2005, jun). Self-determination theory and work motivation. *Journal of Organizational Behavior*, *26*(4), 331–362.
doi: 10.1002/job.322

Gerling, K. M., Birk, M., Mandryk, R. L., & Doucette, A. (2013). The Effects of Graphical Fidelity on Player Experience. In *Proceedings of international conference on making sense of converging media - academicmindtrek '13* (pp. 229–236).
doi: 10.1145/2523429.2523473

Gerling, K. M., Mandryk, R. L., Birk, M. V., Miller, M. K., Orji, R., Mandryk, R. L., . . . Smeddinck, J. D. (2014). The effects of embodied persuasive games on player attitudes toward people using wheelchairs. In *Proceedings of the 32nd annual acm conference on human factors in computing systems - chi '14* (pp. 2201–2210). New York, New York, USA: ACM Press.

doi: 10.1145/2556288.2556962

Gilbody, S., Littlewood, E., Hewitt, C., Brierley, G., Tharmanathan, P., Araya, R., ... White, D. (2015). Computerised cognitive behaviour therapy (cCBT) as treatment for depression in primary care (REEACT trial): large scale pragmatic randomised controlled trial. *Bmj*, h5627. doi: 10.1136/bmj.h5627

Gillan, C. M., & Daw, N. D. (2016, jul). Taking Psychiatry Research Online. *Neuron*, *91*(1), 19–23. doi: 10.1016/j.neuron.2016.06.002

Glenn, D., Golinelli, D., Rose, R. D., Roy-Byrne, P., Stein, M. B., Sullivan, G., ... Craske, M. G. (2014). Who gets the most out of cognitive-behavioral therapy for anxiety disorders? The role of treatment dose and patient engagement. *Journal of Consulting and Clinical Psychology*, *81*(4), 639–649. doi: 10.1037/a0033403.Who

Göbel, S., Hardy, S., Wendel, V., Mehm, F., & Steinmetz, R. (2010). Serious games for health. In *Proceedings of the international conference on multimedia - mm '10* (p. 1663). New York, New York, USA: ACM Press.
doi: 10.1145/1873951.1874316

Goldstein, J., Cajko, L., Oosterbroek, M., Michielsen, M., Van Houten, O., & Salverda, F. (1997, jan). VIDEO GAMES AND THE ELDERLY. *Social Behavior and Personality: an international journal*, *25*(4), 345–352.
doi: 10.2224/sbp.1997.25.4.345

Gomez, P., & Perea, M. (2009). A model of the go/no go task. *Journal of experimental psychology*, *136*(3), 389–413. doi: 10.1037/0096-3445.136.3.389.A

Gonnerman, M. E., Parker, C. P., Lavine, H., & Huff, J. (2000, sep). The Relationship between Self-Discrepancies and Affective States: The Moderating Roles of Self-Monitoring and Standpoints on the Self. *Personality and Social Psychology Bulletin*, *26*(7), 810–819.
doi: 10.1177/0146167200269006

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003, dec). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, *37*(6), 504–528.
doi: 10.1016/S0092-6566(03)00046-1

Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature*, *423*(6939), 534–537.

Green, C. S., & Bavelier, D. (2006, dec). Effect of action video games on the spatial distribution of visuospatial attention. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(6), 1465–1478.
doi: 10.1037/0096-1523.32.6.1465

Greitemeyer, T., & Osswald, S. (2010). Effects of prosocial video games on prosocial behavior. *Journal of personality and social psychology*, *98*(2), 211–221. doi: 10.1037/a0016997

Griffiths, K. M., Farrer, L., & Christensen, H. (2010). The efficacy of internet interventions for depression and

anxiety disorders: a review of randomised controlled trials. *The Medical journal of Australia*, *192*(11 Suppl), S4–S11.
doi: gri10844$_f m[pii]$

Grol, R., & Grimshaw, J. (1999). Evidence-based implementation of evidence-based medicine. *The Joint Commission journal on quality improvement*, *25*(10), 503–13.
doi: 10.1016/S1070-3241(16)30464-3

Grossberg, S. (1999, mar). The Link between Brain Learning, Attention, and Consciousness. *Consciousness and Cognition*, *8*(1), 1–44.
doi: 10.1006/ccog.1998.0372

Guay, F., Vallerand, R. J., & Blanchard, C. (2000). On the assessment of situational motivation scale. *Motivation and Emotion*, *24*(3), 175–213.

Hagerty, B. M., Lynch-Sauer, J., Patusky, K. L., Bouwsema, M., & Collier, P. (1992, jun). Sense of belonging: A vital mental health concept. *Archives of Psychiatric Nursing*, *6*(3), 172–177.
doi: 10.1016/0883-9417(92)90028-H

Hakamata, Y., Lissek, S., Bar-Haim, Y., Britton, J. J. C. J., Fox, N. A., Leibenluft, E., … Pine, D. S. (2010). Attention bias modification treatment: A meta-analysis toward the establishment of novel treatment for anxiety. *Biological Psychiatry*, *68*(11), 982–990.
doi: 10.1016/j.biopsych.2010.07.021

Hanks, A. S., Just, D. R., Smith, L. E., & Wansink, B. (2012). Healthy convenience: Nudging students toward healthier choices in the lunchroom. *Journal of Public Health (United Kingdom)*, *34*(3), 370–376.
doi: 10.1093/pubmed/fds003

Hanus, M. D., & Fox, J. (2015). Persuasive avatars: The effects of customizing a virtual salespersons appearance on brand liking and purchase intentions. *International Journal of Human Computer Studies*, *84*, 33–40.
doi: 10.1016/j.ijhcs.2015.07.004

Hars, A., & Shaosong Ou. (2001). Working for free? Motivations of participating in open source projects. In *Proceedings of the 34th annual hawaii international conference on system sciences* (p. 9). IEEE Comput. Soc.
doi: 10.1109/HICSS.2001.927045

Hauser, D. J., & Schwarz, N. (2016, mar). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods*, *48*(1), 400–407.
doi: 10.3758/s13428-015-0578-z

Haw, J. (2008, jun). Random-ratio schedules of reinforcement: The role of early wins and unreinforced trials. *Journal of Gambling Issues*, *21*, 56–67.
doi: 10.4309/jgi.2008.21.6

Hayes, A. F. (2012). *PROCESS: A Versatile Computational Tool for Observed Variable Mediation, Modera-*

tion, and Conditional Process Modeling (Tech. Rep.).

Hayes, A. F. (2013). Introduction to mediation, moderation, and conditional process analysis. *New York, NY: Guilford*, 3–4.
doi: 978-1-60918-230-4

Heber, E., Lehr, D., Ebert, D. D., Berking, M., & Riper, H. (2016). Web-Based and Mobile Stress Management Intervention for Employees: A Randomized Controlled Trial. *Journal of medical Internet research*, *18*(1), e21. doi: 10.2196/jmir.5112

Hernandez, H. A., Ketcheson, M., Schneider, A., Ye, Z., Fehlings, D., Switzer, L., ... Graham, T. C. N. (2014). Design and Evaluation of a Networked Game to Supportsocial Connection of Youth with Cerebral Palsy. *Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility*, 161–168.
doi: 10.1145/2661334.2661370

Hernandez Alvarado, H. A. (2015). *Exergames for Youth with Cerebral Palsy: Designing for Gameplay and Social Accessibility* (Unpublished doctoral dissertation).

Hitlin, P. (2016). *Research in the Crowdsourcing Age, a Case Study* (Tech. Rep.).

Hitt, D. D., Marriott, R. G., & Esser, J. K. (1992). Effects of Delayed Rewards and Task interest on Intrinsic Motivation. *Basic and Applied Social Psychology*, *13*(4), 405–414. doi: 10.1207/s15324834basp1304$_2$

Hoffner, C. (1996, jun). Children's Wishful Identification and Parasocial Interaction with Favorite Television Characters. *Journal of Broadcasting & Electronic Media*, *40*(3), 389–402.
doi: 10.1080/08838159609364360

Hoffner, C., & Buchanan, M. (2005, nov). Young Adults' Wishful Identification With Television Characters: The Role of Perceived Similarity and Character Attributes. *Media Psychology*, *7*(4), 325–351.
doi: 10.1207/S1532785XMEP0704$_2$

Horvitz, E., Jacobs, A., & Hovel, D. (1999). Attention-sensitive alerting. In *Proceedings of the fifteenth conference on uncertainty in artificial intelligence* (pp. 305–313).

http://www.duolingo.com. (n.d.).

Huang, Y., Wang, L., Shi, J., Chan, S., Clark, K., Jennings, K., ... Huang, Y. (2009). When do objects become more attractive? The individual and interactive effects of choice and ownership on object evaluation. *Personality and social psychology bulletin*, *35*(6), 713–722. doi: 10.1177/0146167209333046

Huckvale, K., Morrison, C., Ouyang, J., Ghaghda, A., & Car, J. (2015). The evolution of mobile apps for asthma: An updated systematic assessment of content and tools. *BMC Medicine*, *13*(1), 1–15. doi: 10.1186/s12916-015-0303-x

Huff, C., & Tingley, D. (2015, aug). "Who are these people?" Evaluating the demographic characteristics and political preferences of MTurk survey respondents. *Research & Politics*, *2*(3), 205316801560464.
doi: 10.1177/2053168015604648

Huillery, E., & Seban, J. (2014). Performance-Based Financing, Motivation and Final Output in the Health

Sector: Experimental Evidence from the Democratic Republic of Congo. *Sciences Po Economics Discussion Papers.*, 1–57.

Hunicke, R., Leblanc, M., & Zubek, R. (2004). MDA: A Formal Approach to Game Design and Game Research. San Jose, California.

Hussain, Z., & Griffiths, M. D. (2008, feb). Gender Swapping and Socializing in Cyberspace: An Exploratory Study. *CyberPsychology & Behavior*, *11*(1), 47–53.
doi: 10.1089/cpb.2007.0020

Insel, T. R. (2014, apr). *The nimh research domain criteria (rdoc) project: Precision medicine for psychiatry* (Vol. 171) (No. 4). American Psychiatric AssociationArlington, VA.
doi: 10.1176/appi.ajp.2014.14020138

Iqbal, S. T., & Bailey, B. P. (2008). Effects of intelligent notification management on users and their tasks. *Proceedings of the SIGCHI Conference on Human*, 93–102. doi: 10.1145/1357054.1357070

Iqbal, S. T., & Horvitz, E. (2007). Disruption and recovery of computing tasks: field study, analysis, and directions. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 677–686).
doi: 10.1145/1240624.1240730

Iqbal, S. T., & Horvitz, E. (2010). Notifications and awareness: A field study of alert usage and preferences. *Proceedings of the 2010 ACM conference on Computer supported cooperative work - CSCW '10*, 27–30.
doi: 10.1145/1718918.1718926

Jeni. (2017). *The right app rewards to boost motivation – Google Play Apps & Games.*

Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., & Walton, A. (2008, sep). Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies*, *66*(9), 641–661.
doi: 10.1016/j.ijhcs.2008.04.004

Johnson, D., Gardner, J., & Sweetser, P. (2016). Motivations for videogame play: Predictors of time spent playing. *Computers in Human Behavior*, *63*, 805–812.
doi: 10.1016/j.chb.2016.06.028

Jones, E. B., & Sharpe, L. (2017). Cognitive bias modification: A review of meta-analyses. *Journal of Affective Disorders*, *223*(June), 175–183.
doi: 10.1016/j.jad.2017.07.034

Joosten, E., DeFuentes-Merillas, L., de Weert, G., Sensky, T., van der Staak, C., & de Jong, C. (2008). Systematic Review of the Effects of Shared Decision-Making on Patient Satisfaction, Treatment Adherence and Health Status. *Psychotherapy and Psychosomatics*, *77*(4), 219–226.
doi: 10.1159/000126073

Kahneman, D., & Beatty, J. (1966, dec). Pupil Diameter and Load on Memory. *Science*, *154*(3756), 1583–1585.

doi: 10.1126/science.154.3756.1583

Kane, M. J., Conway, A. R., Miura, T. K., & Colflesh, G. J. (2007). Working memory, attention control, and the N-back task: a question of construct validity. *J Exp Psychol Learn Mem Cogn*, *33*(3), 615–622. doi: 2007-06096-010 [pii]̊10.1037/0278-7393.33.3.615

Kaplowitz, M. D., Hadlock, T. D., & Levine, R. (2004, mar). A comparison of web and mail survey response rates. *Public Opinion Quarterly*, *68*(1), 94–101. doi: 10.1093/poq/nfh006

Kaptein, M., Markopoulos, P., De Ruyter, B., & Aarts, E. (2015). Personalizing persuasive technologies: Explicit and implicit personalization using persuasion profiles. *International Journal of Human Computer Studies*, *77*, 38–51. doi: 10.1016/j.ijhcs.2015.01.004

Kasser, T., Davey, J., & Ryan, R. M. (1992). Motivation and employee- supervisor discrepancies in psychiatric vocational rehabilitation settings. *Rehabilitation Psychology*, *37*(3), 175–188.

Kast, F., Meier, S., & Pomeranz, D. (2012). Under-savers anonymous: Evidence on self-help groups and peer pressure as a savings commitment device. *Journal of Chemical Information and Modeling*, *53*(3), 556–581. doi: 10.1017/CBO9781107415324.004

Kaufmann, N., Schulze, T., & Veit, D. (2011). More than fun and money. Worker Motivation in Crowdsourcing – A Study on Mechanical Turk. *Proceedings of the Seventeenth Americas Conference on Information Systems*, *4*(2009), 1–11. doi: 10.1145/1979742.1979593

Kazdin, A. E., & Blase, S. L. (2011, jan). Rebooting Psychotherapy Research and Practice to Reduce the Burden of Mental Illness. *Perspectives on Psychological Science*, *6*(1), 21–37. doi: 10.1177/1745691610393527

Kees, J., Berry, C., Burton, S., & Sheehan, K. (2017, jan). An Analysis of Data Quality: Professional Panels, Student Subject Pools, and Amazon's Mechanical Turk. *Journal of Advertising*, *46*(1), 141–155. doi: 10.1080/00913367.2016.1269304

Kelders, S. M., Kok, R. N., Ossebaard, H. C., & Van Gemert-Pijnen, J. E. (2012, nov). Persuasive system design does matter: A systematic review of adherence to web-based interventions. *Journal of Medical Internet Research*, *14*(6), 1–24. doi: 10.2196/jmir.2104

Kenney, E. L., Gortmaker, S. L., Carter, J. E., Howe, M. C. W., Reiner, J. F., & Cradock, A. L. (2015, sep). Grab a Cup, Fill It Up! An Intervention to Promote the Convenience of Drinking Water and Increase Student Water Consumption During School Lunch. *American journal of public health*, *105*(9), 1777–83. doi: 10.2105/AJPH.2015.302645

Kerber, F., Hirtz, C., Gehring, S., Löchtefeld, M., & Krüger, A. (2016). Managing smartwatch notifications through filtering and ambient illumination. In *Proceedings of the 18th international conference on human-computer interaction with mobile devices and services adjunct* (pp. 918–923). ACM.
doi: 10.1145/2957265.2962657

Khaled, R., Barr, P., Noble, J., Fischer, R., & Biddle, R. (2007). Fine Tuning the Persuasion in Persuasive Games. In *Persuasive technology* (pp. 36–47). Springer.
doi: 10.1007/978-3-540-77006-0_5

Khaled, R., Fischer, R., Noble, J., & Biddle, R. (2008). A Qualitative Study of Culture and Persuasion in a Smoking Cessation Game. In *Persuasive technology* (pp. 224–236). Berlin, Heidelberg: Springer Berlin Heidelberg.
doi: 10.1007/978-3-540-68504-3_20

Killikelly, C., He, Z., Reeder, C., & Wykes, T. (2017, jul). Improving Adherence to Web-Based and Mobile Technologies for People With Psychosis: Systematic Review of New Potential Predictors of Adherence. *JMIR mHealth and uHealth*, *5*(7), e94.
doi: 10.2196/mhealth.7088

Kim, Y., & Sundar, S. S. (2012). Visualizing ideal self vs. actual self through avatars: Impact on preventive health outcomes. *Computers in Human Behavior*, *28*(4), 1356–1364.
doi: 10.1016/j.chb.2012.02.021

Kittur, A., Chi, E. H., & Suh, B. (2008). Crowdsourcing User Studies With Mechanical Turk. *CHI '08 Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, 453–456.
doi: 10.1145/1357054.1357127

Klimmt, C., Hefner, D., Vorderer, P., Roth, C., & Blake, C. (2010, nov). Identification With Video Game Characters as Automatic Shift of Self-Perceptions. *Media Psychology*, *13*(4), 323–338.
doi: 10.1080/15213269.2010.524911

Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlström, K., ... Westerberg, H. (2005, feb). Computerized Training of Working Memory in Children With ADHD-A Randomized, Controlled Trial. *Journal of the American Academy of Child & Adolescent Psychiatry*, *44*(2), 177–186.
doi: 10.1097/00004583-200502000-00010

Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Training of working memory in children with ADHD. *Journal of Clinical and Experimental Neuropsychology*, *24*(6), 781–791.
doi: 10.1076/jcen.24.6.781.8395

Klinger, E., Bouchard, S., Légeron, P., Roy, S., Lauer, F., Chemin, I., & Nugues, P. (2005, feb). Virtual Reality Therapy Versus Cognitive Behavior Therapy for Social Phobia: A Preliminary Controlled Study. *CyberPsychology & Behavior*, *8*(1), 76–88.
doi: 10.1089/cpb.2005.8.76

Königbauer, J., Letsch, J., Doebler, P., Ebert, D., Baumeister, H., Josephine, K., ... Harald, B. (2018). Internet-and mobile-based depression interventions for people with diagnosed depression: a systematic review and meta-analysis. *Journal of Affective Disorders*, *223*(April), 2–5. doi: 10.1016/j.jad.2017.07.021

Konijn, E. A., Nije Bijvank, M., & Bushman, B. J. (2007). I wish I were a warrior: The role of wishful identification in the effects of violent video games on aggression in adolescent boys. *Developmental Psychology*, *43*(4), 1038–1044.
doi: 10.1037/0012-1649.43.4.1038

Koster, R. (2013). *Theory of fun for game design.* O'Reilly Media, Inc.

Krause, M., Mogalle, M., Pohl, H., & Williams, J. J. (2015). A Playful Game Changer. In *Proceedings of the second (2015) acm conference on learning @ scale - l@s '15* (pp. 95–102). New York, New York, USA: ACM Press.
doi: 10.1145/2724660.2724665

Kuhl, J. (2013). *Auswertungsmanual für den Operanten Multi-Motiv-Test OMT: basierend auf Julius Kuhl & David Scheffer (2009).* Sonderpunkt-Wissen-Verlag.

Lally, P., Van Jaarsveld, C. H., Potts, H. W., & Wardle, J. (2010, oct). How are habits formed: Modelling habit formation in the real world. *European Journal of Social Psychology*, *40*(6), 998–1009.
doi: 10.1002/ejsp.674

Lampit, A., Hallock, H., & Valenzuela, M. (2014). Computerized Cognitive Training in Cognitively Healthy Older Adults: A Systematic Review and Meta-Analysis of Effect Modifiers. *PLoS Medicine*, *11*(11). doi: 10.1371/journal.pmed.1001756

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual.* (Tech. Rep.).

Lin, D. T., Park, J., Liebert, C. A., & Lau, J. N. (2015, jan). Validity evidence for Surgical Improvement of Clinical Knowledge Ops: a novel gaming platform to assess surgical decision making. *The American Journal of Surgery*, *209*(1), 79–85.
doi: 10.1016/j.amjsurg.2014.08.033

Lin, Y.-G., McKeachie, W. J., & Kim, Y. C. (2003). College student intrinsic and/or extrinsic motivation and learning. *Learning and Individual Differences*, *13*(3), 251–258.
doi: 10.1016/S1041-6080(02)00092-4

Livingston, I. I. J. I., Gutwin, C., Mandryk, R. R. L. R., & Birk, M. (2014). How players value their characters in world of warcraft. *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing - CSCW '14*(February), 1333–1343.
doi: 10.1145/2531602.2531661

Locke, E. A. (1968, may). Toward a theory of task motivation and incentives. *Organizational Behavior and Human Performance*, *3*(2), 157–189.

doi: 10.1016/0030-5073(68)90004-4

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, *57*(9), 705–717. doi: 10.1037//0003-066X.57.9.705

Locke, E. A., & Latham, G. P. (2006, oct). New Directions in Goal-Setting Theory. *Current Directions in Psychological Science*, *15*(5), 265–268.
doi: 10.1111/j.1467-8721.2006.00449.x

Lorig, K., Ritter, P. L., Laurent, D. D., Plant, K., Green, M., Jernigan, V. B. B., & Case, S. (2010). Online diabetes self-management program: A randomized study. *Diabetes Care*, *33*(6), 1275–1281. doi: 10.2337/dc09-2153

Macleod, C., Rutherford, E., Campbell, L., Ebsworthy, G., & Holker, L. (2002). Selective attention and emotional vulnerability: Assessing the causal basis of their association through the experimental manipulation of attentional bias. *Journal of Abnormal Psychology*, *111*(1), 107–123.
doi: 10.1037/0021-843X.111.1.107

Maddux, J. E. (1995). Self-efficacy theory. In *Self-efficacy, adaptation, and adjustment* (pp. 3–33). Boston, MA: Springer.

Mallon, B., & Webb, B. (2005, jan). Stand up and take your place. *Computers in Entertainment*, *3*(1), 6.
doi: 10.1145/1057270.1057285

Mandryk, R., & Birk, M. (2017). Habits and preferences of players as a context for the design of game-based digital mental health interventions. *Journal of Medical Internet Research*.

Mandryk, R. L., & Atkins, M. S. (2007, apr). A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. *International Journal of Human-Computer Studies*, *65*(4), 329–347.
doi: 10.1016/j.ijhcs.2006.11.011

Mandryk, R. L., & Birk, M. V. (2017). Toward Game-Based Digital Mental Health Interventions: Player Habits and Preferences. *Journal of Medical Internet Research*, *19*(4), e128. doi: 10.2196/jmir.6906

Mandryk, R. L., Dielschneider, S., Kalyn, M. R., Bertram, C. P., Gaetz, M., Doucette, A., … Keiver, K. (2013). Games as neurofeedback training for children with FASD. *Proceedings of the 12th International Conference on Interaction Design and Children - IDC '13*, 165–172. doi: 10.1145/2485760.2485762

Mandryk, R. L., & Stanley, K. G. (2004). Gemini: Accumulating context for play applications. In *Proceedings of the ubicomp 2004 workshop on playing with sensors.*

Marcus, B. H., Selby, V. C., Niaura, R. S., & Rossi, J. S. (1992, mar). Self-Efficacy and the Stages of Exercise Behavior Change. *Research Quarterly for Exercise and Sport*, *63*(1), 60–66.
doi: 10.1080/02701367.1992.10607557

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, *44*(1), 1–23.
doi: 10.3758/s13428-011-0124-6

McAuley, E., Duncan, T., & Tammen, V. V. (1989, mar). Psychometric Properties of the Intrinsic Motivation Inventory in a Competitive Sport Setting: A Confirmatory Factor Analysis. *Research Quarterly for Exercise and Sport*, *60*(1), 48–58.
doi: 10.1080/02701367.1989.10607413

McAuley, E., Poag, K., Gleason, A., & Wraith, S. (1990). Attrition from exercise programs: Attributional and affective perspectives. *Journal of Social Behavior & Personality*, *5*(6), 591–602.

McDonald, D. G., & Kim, H. (2001, jun). When I Die, I Feel Small: Electronic Game Characters and the Social Self. *Journal of Broadcasting & Electronic Media*, *45*(2), 241–258.
doi: 10.1207/s15506878jobem4502$_3$

Meyer, B., Bierbrodt, J., Schröder, J., Berger, T., Beevers, C. G., Weiss, M., … Klein, J. P. (2015). Effects of an Internet intervention (Deprexis) on severe depression symptoms: Randomized controlled trial. *Internet Interventions*, *2*, 48–59. doi: 10.1016/j.invent.2014.12.003

Michael, D. R., & Chen, S. L. (2005). *Serious games: Games that educate, train, and inform.* Cengage Learning PTR.

Miller, M. K., Mandryk, R. L., Birk, M. V., Depping, A. E., & Patel, T. (2017). Through the Looking Glass : Effects of Feedback on Self-Awareness and Conversation during Video Chat.
doi: 10.1145/3025453.3025548

Mittal, B. (2006). I, me, and mine—how products become consumers' extended selves. *Journal of Consumer Behaviour*, *5*, 550–562.
doi: 10.1002/cb

Mohr, D. C., Cuijpers, P., & Lehman, K. (2011). Supportive accountability: A model for providing human support to enhance adherence to eHealth interventions. *Journal of Medical Internet Research*, *13*(1), 1–16.
doi: 10.2196/jmir.1602

Mohr, D. C., Ho, J., Duffecy, J., Sokol, L., Burns, M. N., Jin, L., & Siddique, J. (2012). Effect of Telephone-Administered vs Face-to-face Cognitive Behavioral Therapy on Adherence to Therapy and Depression Outcomes Among Primary Care Patients. *Jama*, *307*(21), 2278–2285.

Mohr, D. C., Lyon, A. R., Lattie, E. G., Reddy, M., & Schueller, S. M. (2017). Accelerating digital mental health research from early design and creation to successful implementation and sustainment. *Journal of Medical Internet Research*, *19*(5), 1–18. doi: 10.2196/jmir.7725

Mohr, D. C., Schueller, S. M., Montague, E., Burns, M. N., & Rashidi, P. (2014). The behavioral intervention technology model: An integrated conceptual and technological framework for ehealth and mhealth interventions. *Journal of Medical Internet Research*, *16*(6). doi: 10.2196/jmir.3077

Mohr, D. C., Tomasino, K. N., Lattie, E. G., Palac, H. L., Kwasny, M. J., Weingardt, K., … Schueller, S. M. (2017). Intellicare: An eclectic, skills-based app suite for the treatment of depression and anxiety. *Journal of Medical Internet Research*, *19*(1). doi: 10.2196/jmir.6645

Mohr, D. C., Weingardt, K. R., Reddy, M., & Schueller, S. M. (2017). Three Problems With Current Digital Mental Health Research . . . and Three Things We Can Do About Them. *Psychiatric Services*, *68*(5), 427–429. doi: 10.1176/appi.ps.201600541

Morris, R. R., Schueller, S. M., & Picard, R. W. (2015, mar). Efficacy of a web-based, crowdsourced peer-to-peer cognitive reappraisal platform for depression: Randomized controlled trial. *Journal of Medical Internet Research*, *17*(3), e72.
doi: 10.2196/jmir.4167

Motter, J. N., Devanand, D. P., Murali Doraiswamy, P., & Sneed, J. R. (2015). Computerized cognitive training for major depressive disorder: What's next? *Frontiers in Psychiatry*, *6*(OCT), 1–5. doi: 10.3389/fpsyt.2015.00137

Motter, J. N., Pimontel, M. A., Rindskopf, D., Devanand, D. P., Doraiswamy, P. M., & Sneed, J. R. (2016). Computerized cognitive training and functional recovery in major depressive disorder: A meta-analysis. *Journal of Affective Disorders*, *189*, 184–191. doi: 10.1016/j.jad.2015.09.022

Moyer-Gusé, E., Chung, A. H., & Jain, P. (2011, jun). Identification With Characters and Discussion of Taboo Topics After Exposure to an Entertainment Narrative About Sexual Health. *Journal of Communication*, *61*(3), 387–406.
doi: 10.1111/j.1460-2466.2011.01551.x

Moyer-Gusé, E., & Nabi, R. L. (2010). Explaining the effects of narrative in an entertainment television program: Overcoming resistance to persuasion. *Human Communication Research*, *36*(1), 26–52. doi: 10.1111/j.1468-2958.2009.01367.x

Muender, T., Miller, M. K., Birk, M. V., & Mandryk, R. L. (2016). Extracting Heart Rate from Videos of Online Participants. In *Proceedings of the 2016 chi conference on human factors in computing systems - chi '16* (pp. 4562–4567).
doi: 10.1145/2858036.2858064

Mullen, S. P., Wójcicki, T. R., Mailey, E. L., Szabo, A. N., Gothe, N. P., Olson, E. A., . . . McAuley, E. (2013). A Profile for Predicting Attrition from Exercise in Older Adults. *Prevention Science*, *14*(5), 489–496.
doi: 10.1007/s11121-012-0325-y

Murphy, S. C. (2004, aug). 'Live in Your World, Play in Ours': The Spaces of Video Game Identity. *Journal of Visual Culture*, *3*(2), 223–238.
doi: 10.1177/1470412904044801

Murray, H. A. (1943). Thematic apperception test.

Neustaedter, C., & Fedorovskaya, E. (2009). Presenting identity in a virtual world through avatar appearances. *Proceedings of Graphics Interface 2009*, 183–190.

Nevin, J. A. (1969). Signal detection theory and operant behavior: A review of David M. Green and John A. Swets' signal detection and psychophysics. *Journal of the Experimental Analysis of Behavior*, *12*(3),

475–480.
doi: 10.1901/jeab.1969.12-475

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In *Cognitive skills and their acquisition* (pp. 1–55). doi: 10.1.1.910.5264

Ngandu, T., Lehtisalo, J., Solomon, A., Levälahti, E., Ahtiluoto, S., Antikainen, R., . . . Kivipelto, M. (2015). A 2 year multidomain intervention of diet, exercise, cognitive training, and vascular risk monitoring versus control to prevent cognitive decline in at-risk elderly people (FINGER): A randomised controlled trial. *The Lancet*, *385*(9984), 2255–2263. doi: 10.1016/S0140-6736(15)60461-5

Nicholson, S. (2015). A RECIPE for Meaningful Gamification. In *Gamification in education and business* (pp. 1–20). Cham: Springer International Publishing.
doi: 10.1007/978-3-319-10208-5_1

Nouchi, R., Taki, Y., Takeuchi, H., Hashizume, H., Nozawa, T., Kambara, T., . . . Kawashima, R. (2013, feb). Brain Training Game Boosts Executive Functions, Working Memory and Processing Speed in the Young Adults: A Randomized Controlled Trial. *PLoS ONE*, *8*(2), e55518.
doi: 10.1371/journal.pone.0055518

Oinas-Kukkonen, H., & Harjumaa, M. (2008). A Systematic Framework for Designing and Evaluating Systems. *Persuasive Technology, Third International Conference, PERSUASIVE 2008, Oulu, Finland, June 4-6, 2008. Proceedings*, *5033*, 164–176.
doi: 10.1007/978-3-540-68504-3

Orji, R., Mandryk, R. L., & Vassileva, J. (2017, oct). Improving the Efficacy of Games for Change Using Personalization Models. *ACM Transactions on Computer-Human Interaction*, *24*(5), 1–22.
doi: 10.1145/3119929

Orji, R., Vassileva, J., & Mandryk, R. L. (2014, dec). Modeling the efficacy of persuasive strategies for different gamer types in serious games for health. *User Modeling and User-Adapted Interaction*, *24*(5), 453–498.
doi: 10.1007/s11257-014-9149-8

Orosy-Fildes, C., & Allan, R. W. (1989, aug). Psychology of Computer Use: XII. Videogame Play: Human Reaction Time to Visual Stimuli. *Perceptual and Motor Skills*, *69*(1), 243–247.
doi: 10.2466/pms.1989.69.1.243

Padgett, L. S., Strickland, D., & Coles, C. D. (2006). Case Study : Using a Virtual Reality Computer Game to Teach Fire Safety Skills to Children Diagnosed with Fetal Alcohol Syndrome. *Journal of pediatric*, *31*(1), 65–70.

Parsons, T. D., & Rizzo, A. A. (2008, sep). Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: A meta-analysis. *Journal of Behavior Therapy and Experimental Psychiatry*, *39*(3), 250–261.
doi: 10.1016/j.jbtep.2007.07.007

Passmore, C. J., Birk, M. V., & Mandryk, R. L. (2018). The Privilege of Immersion: Racial and Ethnic Experiences, Perceptions, and Beliefs in Digital Gaming. *CHI '18: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. doi: 10.1145/3173574.3173957

Patall, E. A., Cooper, H., & Robinson, J. C. (2008). The Effects of Choice on Intrinsic Motivation and Related Outcomes: A Meta-Analysis of Research Findings. *Psychological Bulletin*, *134*(2), 270–300. doi: 10.1037/0033-2909.134.2.270

Pears, M. (2017). *Design Pillars – The Core of Your Game.*

Pearson, S. D., & Lieber, S. R. (2009). Financial penalties for the unhealthy? Ethical guidelines for holding employees responsible for their health. *Health Affairs*, *28*(3), 845–852.

Pierce, J. L., Kostova, T., & Dirks, K. T. (2003). The state of psychological ownership: Integrating and extending a century of research. *Review of General Psychology*, *7*(1), 84–107. doi: 10.1037/1089-2680.7.1.84

Prins, P. J., Dovis, S., Ponsioen, A., ten Brink, E., & van der Oord, S. (2011). Does Computerized Working Memory Training with Game Elements Enhance Motivation and Training Efficacy in Children with ADHD? *Cyberpsychology, Behavior, and Social Networking*, *14*(3), 115–122. doi: 10.1089/cyber.2009.0206

Przybylski, A. K., Rigby, C. S., & Ryan, R. M. (2010). A motivational model of video game engagement. *Review of General Psychology*, *14*(2), 154–166. doi: 10.1037/a0019440

Przybylski, A. K., Weinstein, N., Murayama, K., Lynch, M. F., & Ryan, R. M. (2012, jan). The ideal self at play: the appeal of video games that let you be all you can be. *Psychological Science*, *23*(1), 69–76. doi: 10.1177/0956797611418676

Rammstedt, B., & John, O. P. (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), 203–212. doi: 10.1016/j.jrp.2006.02.001

Rapp, A. (2017). Drawing Inspiration from World of Warcraft: Gamification Design Elements for Behavior Change Technologies. *Interacting with Computers*, *29*(5). doi: 10.1093/iwc/iwx001

Reeves, B., & Read, J. L. (2009). *Total Engagement: How games and virtual worlds are changing the way people work and businesses compete.* Harvard Business Press.

Richards, D., Richardson, T., Timulak, L., & Mcelvaney, J. (2015). The efficacy of internet-delivered treatment for generalized anxiety disorder: A systematic review and meta-analysis. *INVENT*, *2*, 272–282. doi: 10.1016/j.invent.2015.07.003

Richter, G., Raban, D. R., Rafaeli, S., Daphne R. Raban, & Sheizaf Rafaeli. (2015). Studying gamification: the effect of rewards and incentives on motivation. *Gamification in education and business*, 21–46.

doi: 10.1007/978-3-319-10208-5$_2$

Rigby, C. S., Deci, E. L., Patrick, B. C., & Ryan, R. M. (1992, sep). Beyond the intrinsic-extrinsic dichotomy: Self-determination in motivation and learning. *Motivation and Emotion*, *16*(3), 165–185. doi: 10.1007/BF00991650

Rigby, C. S., & Przybylski, A. K. (2009, jun). Virtual worlds and the learner hero: How today's video games can inform tomorrow's digital learning environments. *Theory and Research in Education*, *7*(2), 214–223. doi: 10.1177/1477878509104326

Rigby, S., & Ryan, R. M. (2011). *Glued to games: How video games draw us in and hold us spellbound: How video games draw us in and hold us spellbound.* ABC-CLIO.

Rizzo, A., Parsons, T. D., Lange, B., Kenny, P., Buckwalter, J. G., Rothbaum, B., ... Reger, G. (2011, jun). Virtual reality goes to war: a brief review of the future of military behavioral healthcare. *Journal of clinical psychology in medical settings*, *18*(2), 176–87. doi: 10.1007/s10880-011-9247-2

Ross, J., Irani, L., Silberman, M. S., Zaldivar, A., & Tomlinson, B. (2010). Who are the crowdworkers? In *Proceedings of the 28th of the international conference extended abstracts on human factors in computing systems - chi ea '10* (p. 2863). New York, New York, USA: ACM Press. doi: 10.1145/1753846.1753873

Rothbaum, B. O., Hodges, L. F., Ready, D., Graap, K., & Alarcon, R. D. (2001, aug). Virtual Reality Exposure Therapy for Vietnam Veterans With Posttraumatic Stress Disorder. *The Journal of Clinical Psychiatry*, *62*(8), 617–622. doi: 10.4088/JCP.v62n0808

Rousseau, D. (1995). *Psychological contracts in organizations: Understanding written and unwritten agreements.* Thousand Oaks, CA: Sage Publications.

Ryan, R., & Deci, E. (2002). *Self-determination theory: An organismic dialectical perspective.*

Ryan, R., & Lynch, M. (2011). *Motivation and automony in counseling psychothrapy and behavior change-a look at theory and practice.*

Ryan, R. M., & Deci, E. L. (2000a, jan). Intrinsic and extrinsic motivation: Classic definitions and new directions. *Contemporary Educational Psychology*, *25*(1), 54–67. doi: doi:10.1006/ceps.1999.1020

Ryan, R. M., & Deci, E. L. (2000b, jan). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *The American Psychologist*, *55*(1), 68–78. doi: 10.1037/0003-066X.55.1.68

Ryan, R. M., & Deci, E. L. (2008). A self-determination theory approach to psychotherapy: The motivational basis for effective change. In *Canadian psychology* (Vol. 49, pp. 186–193). doi: 10.1037/a0012753

Ryan, R. M., Koestner, R., & Deci, E. L. (1991, sep). Ego-involved persistence: When free-choice behavior is not intrinsically motivated. *Motivation and Emotion*, *15*(3), 185–205. doi: 10.1007/BF00995170

Ryan, R. M., Patrick, H., Deci, E. L., & Williams, G. C. (2008). Facilitating health behaviour change and its maintenance: Interventions based on Self-Determination Theory. *The European Health Psychologist*, *10*(1), 2–5. doi: 10.1080/17509840701827437

Ryan, R. M., Rigby, C. S., & Przybylski, A. (2006, nov). The Motivational Pull of Video Games: A Self-Determination Theory Approach. *Motivation and Emotion*, *30*(4), 344–360. doi: 10.1007/s11031-006-9051-8

Salen, K., & Zimmerman, E. (2004). Rules of play: Game design fundamentals. MIT Press.

Sansone, C., Weir, C., Harpster, L., & Morgan, C. (1992). Once a Boring Task Always a Boring Task? Interest as a Self-Regulatory Mechanism. *Journal of Personality and Social Psychology*, *63*(3), 379–390. doi: 10.1037//0022-3514.63.3.379

Schilling, T. A., & Hayashi, C. T. (2001). Achievement Motivation among High School Basketball and Cross-Country Athletes: A Personal Investment Perspective. *Journal of Applied Sport Psychology*, *13*(1), 103–128. doi: 10.1080/10413200109339006

Schlosser, D., Campellone, T., Kim, D., Truong, B., Vergani, S., Ward, C., & Vinogradov, S. (2016). Feasibility of PRIME: A Cognitive Neuroscience-Informed Mobile App Intervention to Enhance Motivated Behavior and Improve Quality of Life in Recent Onset Schizophrenia. *JMIR Research Protocols*, *5*(2), e77. doi: 10.2196/resprot.5450

Schneider, E. F., Lang, A., Shin, M., & Bradley, S. D. (2004). Death with a Story. *Human Communication Research*, *30*(3), 361–375. doi: 10.1111/j.1468-2958.2004.tb00736.x

Schueller, S. M., Tomasino, K. N., & Mohr, D. C. (2017). Integrating Human Support Into Behavioral Intervention Technologies: The Efficiency Model of Support. *Clinical Psychology: Science and Practice*, *24*(1), 27–45. doi: 10.1111/cpsp.12173

Seaborn, K., & Fels, D. I. (2015). Gamification in theory and action: A survey. *International Journal of Human-Computer Studies*, *74*, 14–31. doi: 10.1016/j.ijhcs.2014.09.006

Seaverson, E. L. D., Grossmeier, J., Miller, T. M., & Anderson, D. R. (2009, may). The Role of Incentive Design, Incentive Value, Communications Strategy, and Worksite Culture on Health Risk Assessment Participation. *American Journal of Health Promotion*, *23*(5), 343–352. doi: 10.4278/ajhp.08041134

Seligman, M. E. P. (1975). *Helplessness: On depression, development, and death.* New York, NY, US: W H Freeman/Times Books/ Henry Holt & Co.

Seligman, M. E. P., & Csikszentmihalyi, M. (2014). Positive Psychology: An Introduction. In *Flow and the foundations of positive psychology* (pp. 279–298). Dordrecht: Springer Netherlands. doi: 10.1037//0003-066x.55.1.5

Shapiro, D. N., Chandler, J., & Mueller, P. A. (2013, apr). Using Mechanical Turk to Study Clinical Populations. *Clinical Psychological Science*, *1*(2), 213–220. doi: 10.1177/2167702612469015

Shechner, T., Rimon-Chakir, A., Britton, J. C., Lotan, D., Apter, A., Bliese, P. D., ... Bar-Haim, Y. (2014, jan). Attention Bias Modification Treatment Augmenting Effects on Cognitive Behavioral Therapy in Children With Anxiety: Randomized Controlled Trial. *Journal of the American Academy of Child and Adolescent Psychiatry*, *53*(1), 61–71. doi: 10.1016/j.jaac.2013.09.016

Sheinin, M., & Gutwin, C. (2014). Exertion in the small. In *Proceedings of the 32nd annual acm conference on human factors in computing systems - chi '14* (pp. 1845–1854). New York, New York, USA: ACM. doi: 10.1145/2556288.2557385

Silberman, S. (2016). *Paying minimum wage on online labor platforms.*

Skinner, B. F. (1972). *Beyond freedom and dignity.* Indianapolis, IN: Hackett Pub Co.

Smahel, D., Blinka, L., & Ledabyl, O. (2008). Playing MMORPGs: Connections between Addiction and Identifying with a Character. *CyberPsychology & Behavior*, *11*(6), 715–718. doi: 10.1089/cpb.2007.0210

Smeddinck, J. D., Herrlich, M., & Malaka, R. (2015). Exergames for Physiotherapy and Rehabilitation. In *Proceedings of the 33rd annual acm conference on human factors in computing systems - chi '15* (pp. 4143–4146). New York, New York, USA: ACM Press. doi: 10.1145/2702123.2702598

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: applications to dementia and amnesia. *Journal of experimental psychology. General*, *117*(1), 34–50. doi: 10.1037/0096-3445.117.1.34

Spielberger, C. D., Gorsuch, R., & Lushene, R. E. (1970). *Manual for the state-trait anxiety inventory.* Sunnyvale, CA: Consulting Psychologists Press.

Staw, B. M., Calder, B. J., Hess, R. K., & Sandelands, L. E. (1980, mar). Intrinsic Motivation and norms about payment. *Journal of Personality*, *48*(1), 1–14. doi: 10.1111/j.1467-6494.1980.tb00961.x

Steinemann, S. T., Mekler, E. D., & Opwis, K. (2015). Increasing Donating Behavior Through a Game for Change. In *Proceedings of the 2015 annual symposium on computer-human interaction in play - chi play '15* (pp. 319–329). New York, New York, USA: ACM Press. doi: 10.1145/2793107.2793125

Subrahmanyam, K., & Greenfield, P. M. (1994). Effect of video game practice on spatial skills in girls and boys. *Journal of applied developmental psychology*, *15*(1), 13–32.

Sundar, S. S., & Marathe, S. S. (2010, jun). Personalization versus customization: The importance of agency, privacy, and power usage. *Human Communication Research*, *36*(3), 298–322. doi: 10.1111/j.1468-2958.2010.01377.x

Sweetser, P., & Wyeth, P. (2005). GameFlow: a model for evaluating player enjoyment in games. *Computers in Entertainment (CIE)*, *3*(3), 1–24.

Tate, D. F., Jackvony, E. H., & Wing, R. R. (2006). A randomized trial comparing human e-mail counseling, computer-automated tailored counseling, and no counseling in an Internet weight loss program. *Archives of internal medicine*, *166*(15), 1620–5. doi: 10.1001/archinte.166.15.1620

Thaler, R., & Sunstein, C. (2008). *Nudge: Improving Decisions about Health, Wealth, and Happiness.* Yale University Press.

Thiart, H., Lehr, D., Ebert, D. D., Berking, M., & Riper, H. (2015, mar). Log in and breathe out: Internet-based recovery training for sleepless employees with work-related strain – results of a randomized controlled trial. *Scandinavian Journal of Work, Environment and Health*, *41*(2), 164–174. doi: 10.5271/sjweh.3478

Thorell, L. B. L., Lindqvist, S., Nutley, S. B., Bohlin, G., Klingberg, T., Bergman Nutley, S., ... Klingberg, T. (2009, jan). Training and transfer effects of executive functions in preschool children. *Developmental Science*, *12*(1), 106–113. doi: 10.1111/j.1467-7687.2008.00745.x

Tomlinson, B. (2005). From Linear to Interactive Animation : How Autonomous Characters Change the Process and Product of Animating. , *3*(1), 1–20.

Torous, J., & Firth, J. (2018). *Bridging the dichotomy of actual versus aspirational digital health* (Vol. 17) (No. 1). Wiley-Blackwell. doi: 10.1002/wps.20464

Trepte, S., & Reinecke, L. (2010, jan). Avatar creation and video game enjoyment. *Journal of Media Psychology*, *22*(4), 171–184. doi: 10.1027/1864-1105/a000022

UCLA. (2016). *Institute for digital research and education.*

Vallerand, R. J., & Losier, G. F. (1999). An integrative analysis of intrinsic and extrinsic motivation in sport. *Journal of Applied Sport Psychology*, *11*(1), 142–169. doi: 10.1080/10413209908402956

Van Breukelen, G. J. P. (2006). ANCOVA versus change from baseline had more power in randomized studies and more bias in nonrandomized studies. *Journal of Clinical Epidemiology*, *59*(9), 920–925. doi: 10.1016/j.jclinepi.2006.02.007

Van Koningsbruggen, G. M., Veling, H., Stroebe, W., & Aarts, H. (2014). Comparing two psychological interventions in reducing impulsive processes of eating behaviour: Effects on self-selected portion size. *British Journal of Health Psychology*, *19*(4), 767–782. doi: 10.1111/bjhp.12075

Van Looy, J., Courtois, C., De Vocht, M., & De Marez, L. (2012, may). Player Identification in Online Games: Validation of a Scale for Measuring Identification in MMOGs. *Media Psychology*, *15*(2), 197–221. doi: 10.1080/15213269.2012.674917

Van Reijmersdal, E. A., Jansz, J., Peters, O., & Van Noort, G. (2013). Why girls go pink: Game character identification and game-players' motivations. *Computers in Human Behavior*, *29*(6), 2640–2649. doi: 10.1016/j.chb.2013.06.046

VandenBerghe, J. (2014). Engines of play. *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play - CHI PLAY '14*, 476–476. doi: 10.1145/2658537.2677953

VanLooy, J., Courtois, C., De Vocht, M., & De Marez, L. (2012). Player Identification in Online Games: Validation of a Scale for Measuring Identification in MMOGs. *Media Psychology Bandura & Huston*, *15*, 197–221. doi: 10.1080/15213269.2012.674917

van Muijden, J., Band, G. P. H., & Hommel, B. (2012, aug). Online games training aging brains: limited transfer to cognitive control functions. *Frontiers in Human Neuroscience*, *6*, 221. doi: 10.3389/fnhum.2012.00221

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

Vasalou, A., & Joinson, A. N. (2009). Me, myself and I: The role of interactional context on self-presentation through avatars. *Computers in Human Behavior*, *25*(2), 510–520. doi: 10.1016/j.chb.2008.11.007

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, *27*(3), 425. doi: 10.2307/30036540

Vermeire, E., Hearnshaw, H., Van Royen, P., & Denekens, J. (2001). Patient adherence to treatment: Three decades of research. A comprehensive review. *Journal of Clinical Pharmacy and Therapeutics*, *26*(5), 331–342. doi: 10.1046/j.1365-2710.2001.00363.x

Vicencio-Moreira, R., Mandryk, R. L., & Gutwin, C. (2015). Now You Can Compete With Anyone: Balancing Players of Different Skill Levels in a First-Person Shooter Game. In *Proceedings of the 33rd annual acm conference on human factors in computing systems - chi '15* (pp. 2255–2264). New York, New York, USA: ACM Press. doi: 10.1145/2702123.2702242

Waltz, E. (2018, jun). Pear approval signals FDA readiness for digital treatments. *NATURE BIOTECH-*

NOLOGY, *36*(6). doi: 10.1038/nbt0618-481

Warner, A. T., & Obrecht, A. (2015). *Standardising humanitarian data for a better response: The Humanitarian eXchange Language* (Tech. Rep.). INNOVATION IN HUMANITARIAN ACTION.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *47*, 1063–1070. doi: http://dx.doi.org/10.1037/0022-3514.54.6.1063

Watson, D., Hancock, M., Mandryk, R. L., & Birk, M. (2013). Deconstructing the touch experience. *Proceedings of the 2013 ACM international conference on Interactive tabletops and surfaces - ITS '13*, 199–208.
doi: 10.1145/2512349.2512819

Webb, T. L., Joseph, J., Yardley, L., & Michie, S. (2010). Using the Internet to promote health behavior change: A systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. *Journal of Medical Internet Research*, *12*(1), 1–23. doi: 10.2196/jmir.1376

Weber, D., Voit, A., Le, H. V., & Henze, N. (2016). Notification dashboard: enabling reflection on mobile notifications. In *Proceedings of the 18th international conference on human-computer interaction with mobile devices and services adjunct - mobilehci '16* (pp. 936–941). New York, New York, USA: ACM Press.
doi: 10.1145/2957265.2962660

Wongsuphasawat, K., Gamburg, A., & Moraveji, N. (2012). You Can't Force Calm: Designing and Evaluating Respiratory Regulating Interfaces for Calming Technology. *the 25th annual ACM symposium on User interface software and technology*, 69–70.
doi: 10.1145/2380296.2380326

World Health Organization. (2017). *Mental Health Atlas* (Tech. Rep.).

World Health Organization, M. o. S. A. U. (2014). *Global status report on alcohol and health* (Tech. Rep.).

Yee, N. (2005). *The Daedalus Project: A Model of Player Motivations.*

Yee, N. (2006, dec). Motivations for play in online games. *Cyberpsychology & Behavior*, *9*(6), 772–5. doi: 10.1089/cpb.2006.9.772

Yee, N., & Bailenson, J. (2007, jul). The Proteus Effect: The Effect of Transformed Self-Representation on Behavior. *Human Communication Research*, *33*(3), 271–290.
doi: 10.1111/j.1468-2958.2007.00299.x

Zendle, D., Cairns, P., & Kudenko, D. (2015). Higher Graphical Fidelity Decreases Players' Access to Aggressive Concepts in Violent Video Games. In *Proceedings of the 2015 annual symposium on computer-human interaction in play - chi play '15* (pp. 241–251). New York, New York, USA: ACM Press.
doi: 10.1145/2793107.2793113

Zhu, B., Hedman, A., Feng, S., Li, H., & Osika, W. (2017). Designing, prototyping and evaluating digital mindfulness applications: A case study of mindful breathing for stress reduction. *Journal of Medical Internet Research*, *19*(6). doi: 10.2196/jmir.6955

Zimmerman, B. J. (2000, jan). Self-Efficacy: An Essential Motive to Learn. *Contemporary Educational Psychology*, *25*(1), 82–91.
doi: 10.1006/ceps.1999.1016

Appendix A

Manuscript A

A.1 Consent Form

Title: Personalizing, adapting, and balancing computer games

Researcher(s): Max Birk, PhD Student, Department of Computer Science, University of Saskatchewan, 306-966-2327, max.birk@usask.ca

Purpose(s) and Objective(s) of the Research: The purpose of this project is to develop models for personalizing and adapting games.

Procedures: In this study, you will be asked to complete a survey, asking you some questions about yourself. Following the survey completion, you will be asked to play a game. Following game play, you will be asked to complete additional questionnaires that ask you questions about your experience.

Funded by: The Natural Sciences and Engineering Research Council of Canada (NSERC).

Potential Risks and Benefits: There are no known or anticipated risks to you by participating in this research. Your participation will help us to design games, including serious games designed to motivate behaviours or support learning.

Confidentiality:

- Confidentiality will be maintained throughout the study. The entire process and data will be anonymized. Data will only be presented in the aggregate and any individual user comments will be anonymized prior to presentation in academic venues.
- Only the principal researcher and her research assistants will have access to the data to ensure that your confidentiality is protected.
- Storage of Data
  - Data (including survey and interview responses, logs of computer use, and videos of interaction) will be stored on a secure password-protected server for 7 years after data collection.
  - After 7 years, the data will be destroyed. Paper data will be shredded and digital data will be wiped from hard disks beyond any possibility for data recovery.

Right to Withdraw:

- Your participation is voluntary. You may withdraw from the research project for any reason, at any time without explanation.
- Should you wish to withdraw, you may do so at any point, and we will not use your data; we will destroy all records of your data.
- Your right to withdraw data from the study will apply until the data have been aggregated (one week after study completion). After this date, it is possible that some form of research dissemination will have already occurred and it may not be possible to withdraw your data.

Follow up: To obtain results from the study, please contact Max Birk (max.birk@usask.ca).

Questions or Concerns:

- Contact the researcher(s) using the information at the top.
- This research project has been approved on ethical grounds by the University of Saskatchewan Research Ethics Board. Any questions regarding your rights as a participant may be addressed to that committee through the Research Ethics Office ethics.office@usask.ca (306) 966-2975. Out of town participants may call toll free (888) 966-2975.

A.2 Instruments

A.2.1 Player Experience of Need Satisfaction Scale

The Player Experience of Need Satisfaction Scale is a 21-item multidimensional instrument that measures competence, autonomy, relatedness, immersion, and intuitive control, and is geared towards video games.

The scale was cited as "Ryan, Richard M., C. Scott Rigby, and Andrew Przybylski. "The motivational pull of video games: A self-determination theory approach." Motivation and emotion 30, no. 4 (2006): 344-360.". To validate PENS, Ryan et al. (2006) compared the PENS scale to Yee's Model of Motivation (Yee, 2005).

A.2.1.1 Instructions

All participants received the following instructions: "Reflect on your experiences and rate your agreement with the following statements."

A.2.1.2 Items

I have agreed to a confidentiality and limited use agreement that prevents me from publishing the PENS items.

A.2.1.3 Answer Format

A seven-point Likert-scale was used to record participants' agreement:

- Strongly disagree
- Disagree
- Slightly disagree
- Neutral
- Slightly agree
- Agree
- Strongly agree

A.2.2 Intrinsic Motivation Inventory

The Intrinsic Motivation Inventory (IMI) is an 18-item multidimensional instrument that measures interest/enjoyment, effort, pressure/tension, and competence when performing a target activity such as playing a game.

The inventory was cited as "McAuley, E., Duncan, T., & Tammen, V. V. (1989). Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: A confirmatory factor analysis. Research Quarterly for Exercise and Sport, 60, 48-58.". McAuley et al. (1989) investigated the validity of the IMI in a basketball free-throw shooting game.

A.2.2.1 Instructions

All participants received the following instructions: "A number of statements which people have used to describe the previous task are given below. Read each statement and indicate your agreement with that statement. There are no right or wrong answers. Do not spend too much time on any one statement. Remember, give the answer which seems to describe how you thought during the game."

A.2.2.2 Items

Interest/Enjoyment

- I enjoyed this game very much.
- Playing the game was fun.
- I would describe this game as very interesting.
- While playing the game, I was thinking about how much I enjoyed it.
- This game did not hold my attention.

Effort

- I put a lot of effort into this game.
- I tried very hard while playing the game.
- I didn't try very hard at playing the game.
- It was important to me to do well at this game.

Pressure/Tension

- I felt tense while playing the game.
- I felt pressured while playing the game.
- I was anxious while playing the game.
- I was very relaxed while playing the game.

Competence

- I am satisfied with my performance at this game.
- I think I am pretty good at this game.
- After playing the game for a while, I felt pretty competent.
- I am pretty skilled at the game.
- I couldn't play this game very well.

A.2.2.3 Answer Format

A seven-point Likert-scale was used to record participants' agreement:

- Strongly disagree
- Disagree
- Slightly disagree
- Neutral
- Slightly agree
- Agree
- Strongly agree

A.2.3 Positive Affect/Negative Affect Schedule

The Positive Affect/Negative Affect Schedule (PANAS) is a 20-item scale that measures positive and negative affect by gauging the extent participants felt a positive or negative emotion or feeling during a specified time frame.

The schedule was cited as "Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. Journal of personality and social psychology, 54(6), 1063.". Watson et al. (1988) developed and validated the PANAS scales over a two-month time period in comparison to established measures at the time of comparison.

A.2.3.1 Instructions

Because PANAS was measured at baseline and after playing a session of the infinite runner I used two sets of instructions.

Baseline Instruction "This scale consists of a number of words and phrases that describe different feelings and emotions. Read each item and then mark the appropriate answer in the space next to that word. Indicate to what extent you feel this way right now."

Post-play Instruction "This scale consists of a number of words and phrases that describe different feelings and emotions. Read each item and then mark the appropriate answer in the space next to that word. Indicate to what extent you felt this way DURING THE GAME."

A.2.3.2 Items

Positive Affect

- active
- interested
- excited
- strong
- hostile
- inspired
- proud
- enthusiastic
- determined
- attentive

Negative Affect

- upset
- irritable
- guilty
- scared
- hostile
- distressed
- ashamed
- alert
- nervous
- jittery
- afraid

A.2.3.3 Answer Format

A five-point Likert-scale was used to record the extent to which participants felt a feeling or emotion:

- very slightly or not at all
- slightly
- moderately
- quite a bit
- extremely

A.2.4 Player Identification Scale

The Player Identification Scale (PIS) is a 17-item multidimensional instrument that measures similarity identification, embodied identification, and wishful identification. In addition to the three identification scales, the instrument also features a 10-step single item slider to measure participants' level of identification.

The PIS scale is identical with the avatar identification scale presented as part of the scale to measure player identification in massively multiplayer online games proposed and validated by VanLooy et al. (2012). The scale was cited in as "Van Looy, J., Courtois, C., De Vocht, M., & De Marez, L. (2012). Player identification in online games: Validation of a scale for measuring identification in MMOGs. Media Psychology, 15(2), 197-221."

A.2.4.1 Instructions

All participants received the following instructions: "Below are a number of statements which people have used to describe how they feel about they character. Read each statement and indicate your agreement with that statement regarding your character."

A.2.4.2 Items

Similarity Identification

- My character is like me in many ways.
- My character resembled me.
- I identify with my character.
- My character is an extension of myself.
- My character is similar to me.
- I resemble my character.

Embodied Identification

- When I am playing, it feels as if I am my character.
- I feel like I am inside my character when playing.
- In the game, it is as if I become one with my character.
- When I am playing I am transported into my character.
- When playing, it feels as if my character's body becomes my own.
- In the game, it is as if I act directly through my character.

Wishful Identification

- If I could become my character, I would.
- I would like to be more like my character.
- My character is an example to me.
- My character is a better me.
- My character has characteristics that I would like to have.

Identification Slider

- Please rate how much you identify with your character.

A.2.4.3 Answer Format

A seven-point Likert-scale was used to record participants' agreement:

- Strongly disagree
- Disagree
- Slightly disagree
- Neutral
- Slightly agree
- Agree
- Strongly agree

The level of identification was measured using a 10-step slider ranging from "not at all" to "very much".

Appendix B

Manuscript B

B.1 Consent Form

Title: Personalizing, adapting, and balancing computer games

Researcher(s): Max Birk, PhD Student, Department of Computer Science, University of Saskatchewan, 306-966-2327, max.birk@usask.ca

Purpose(s) and Objective(s) of the Research: The purpose of this project is to develop models for personalizing and adapting games.

Procedures: Due to procedural differences, I presented two different procedure descriptions: the first description for the procedures on Day 1, and the second describtion for the procedures on Day 2–11.

Procedures Day 1: In this multi-day study, you will be asked to complete a survey, asking you some questions about yourself. Following the survey completion, you will be asked to play a game using an avatar. Following game play, you will be asked to complete additional questionnaires that ask you questions about your experience. We will also reinvite you for five additional studies taking place over the course of this week. The five subsequent studies will be short in time (10 min) and pay $1 each. Of course, you are free not to participate in any of the subsequent studies without any consequences regarding this study.

Procedures Day 2–11: In this study, you will be asked to fill out a short questionnaire and play a game. Following game play, you will be asked to complete additional questionnaires that ask you questions about your experience. After the survey you will be asked to participate in a spatial search task.

Funded by: The Natural Sciences and Engineering Research Council of Canada (NSERC).

Potential Risks and Benefits: There are no known or anticipated risks to you by participating in this research. Your participation will help us to design games, including serious games designed to motivate behaviours or support learning.

Confidentiality:

- Confidentiality will be maintained throughout the study. The entire process and data will be anonymized. Data will only be presented in the aggregate and any individual user comments will be anonymized prior to presentation in academic venues.
- Only the principal researcher and her research assistants will have access to the data to ensure that your confidentiality is protected.
- Storage of Data
  - Data (including survey and interview responses, logs of computer use, and videos of interaction) will be stored on a secure password-protected server for 7 years after data collection.
  - After 7 years, the data will be destroyed. Paper data will be shredded and digital data will be wiped from hard disks beyond any possibility for data recovery.

Right to Withdraw:

- Your participation is voluntary. You may withdraw from the research project for any reason, at any time without explanation.
- Should you wish to withdraw, you may do so at any point, and we will not use your data; we will destroy all records of your data.
- Your right to withdraw data from the study will apply until the data have been aggregated (one week after study completion). After this date, it is possible that some form of research dissemination will have already occurred and it may not be possible to withdraw your data

Follow up: To obtain results from the study, please contact Max Birk (max.birk@usask.ca).

Questions or Concerns:

- Contact the researcher(s) using the information at the top.
- This research project has been approved on ethical grounds by the University of Saskatchewan Research Ethics Board. Any questions regarding your rights as a participant may be addressed to that committee through the Research Ethics Office ethics.office@usask.ca (306) 966-2975. Out of town participants may call toll free (888) 966-2975

B.2 Instruments

B.2.1 Intrinsic Motivation Inventory

Instructions, answer format, and items were identical with the Intrinsic Motivation Inventory described in A.2.2.

B.2.2 Player Identification Scale

Instructions, answer format, and items were identical with the Player Identification Scale described in A.2.4.

Appendix C

Manuscript C

C.1 Consent Form

Title: Personalizing, adapting, and balancing computer games

Researcher(s): Max Birk, PhD Student, Department of Computer Science, University of Saskatchewan, 306-966-2327, max.birk@usask.ca

Purpose(s) and Objective(s) of the Research: The purpose of this project is to develop models for personalizing and adapting games.

Procedures: Due to procedural differences, I presented two different procedural descriptions: the first description for Day 0–19, and the second description for Day 20, i.e., the Exit Survey.

Procedures Day 0–19: In this study, you will be asked to fill in questionnaires about yourself. On the second day of the study, we will ask you to participate in a breathing exercise. We also will invite you to complete the breathing exercise daily for 21 days total.

Procedures Day 20: In this study, you will be asked to fill in questionnaires about yourself. We will then ask you to participate in a one-minute breathing exercise, followed by additional questionnaires.

Funded by: The Natural Sciences and Engineering Research Council of Canada (NSERC).

Potential Risks and Benefits: There are no known or anticipated risks to you by participating in this research. Your participation will help us to design games, including serious games designed to motivate behaviours or support learning.

Confidentiality:

- Confidentiality will be maintained throughout the study. The entire process and data will be anonymized. Data will only be presented in the aggregate and any individual user comments will be anonymized prior to presentation in academic venues.
- Only the principal researcher and her research assistants will have access to the data to ensure that your confidentiality is protected.
- Storage of Data
  - Data (including survey and interview responses, logs of computer use, and videos of interaction) will be stored on a secure password-protected server for 7 years after data collection.
  - After 7 years, the data will be destroyed. Paper data will be shredded and digital data will be wiped from hard disks beyond any possibility for data recovery.

Right to Withdraw:

- Your participation is voluntary. You may withdraw from the research project for any reason, at any time without explanation.
- Should you wish to withdraw, you may do so at any point, and we will not use your data; we will destroy all records of your data.
- Your right to withdraw data from the study will apply until the data have been aggregated (one week after study completion). After this date, it is possible that some form of research dissemination will have already occurred and it may not be possible to withdraw your data

Follow up: To obtain results from the study, please contact Max Birk (max.birk@usask.ca).

Questions or Concerns:

- Contact the researcher(s) using the information at the top.
- This research project has been approved on ethical grounds by the University of Saskatchewan Research Ethics Board. Any questions regarding your rights as a participant may be addressed to that committee through the Research Ethics Office ethics.office@usask.ca (306) 966-2975. Out of town participants may call toll free (888) 966-2975.

C.2 Instruments

C.2.1 Situational Motivation Scale

The Situational Motivation Scale (SIMS) is a 16-item multidimensional instrument that measures intrinsic motivation, introjected, external regulation, and amotivation.

The scale was cited as "Guay, F., Vallerand, R. J., & Blanchard, C. (2000). On the assessment of situational intrinsic and extrinsic motivation: The Situational Motivation Scale (SIMS). Motivation and emotion, 24(3), 175-213.". Guay et al. (2000) conducted five studies to develop and validate the Situational Motivation Scale.

C.2.1.1 Instructions

All participants received the following instructions: "Read each item carefully. Using the provided scale, please indicate how each item corresponds with your opinion."

C.2.1.2 Items

Items were presented in a grid and prefaced with the question "Why are you currently engaged in this activity?"

Intrinsic Motivation

- Because I think that this activity is interesting
- Because I think that this activity is pleasant
- Because this activity is fun
- Because I feel good when doing this activity

Introjected Regulation

- Because I am doing it for my own good
- Because I think that this activity is good for me
- By personal decision
- Because I believe that this activity is important for me

External Regulation

- Because I am supposed to do it
- Because it is something that I have to do
- Because I don't have any choice
- Because I feel that I have to do it

Amotivation

- There may be good reasons to do this activity, but personally I don't see any
- I do this activity but I am not sure if it is worth it
- I don't know; I don't see what this activity brings me
- I do this activity, but I am not sure it is a good thing to pursue in

C.2.1.3 Answer Format

- corresponds not at all
- corresponds very little
- corresponds a little
- corresponds moderately
- corresponds enough
- corresponds a lot
- corresponds exactly

C.2.2 State-Trait Anxiety Inventory

The State-Trait Anxiety Inventory is a 40-item instrument that measures state and trait anxiety.

The inventory was cited as "Spielberger, C. D., Gorsuch, R., & Lushene, R. E. (1970). Manual for the state-trait anxiety inventory. Sunnyvale, CA: Consulting Psychologists Press.". Spielberger et al. (1970) developed the instrument and published the original manual.

C.2.2.1 Instructions

State All participants received the following instructions: "A number of statements which people have used to describe themselves are given below. Read each statement and then, using the scale below, **indicate how you feel right now, that is, at this moment.** There are no right or wrong answers. Do not spend too much time on any one statement but give the answer which seems to describe your present feelings best."

Trait All participants received the following instructions: "A number of statements which people have used to describe themselves are given below. Read each statement and then, using the scale below, **indicate how you generally feel.** There are no right or wrong answers. Do not spend too much time on any one statement, but give the answer which seems to describe how you generally feel."

C.2.2.2 Items

State

- I feel calm
- I feel secure
- I am tense
- I feel strained
- I feel at ease
- I feel upset
- I am presently worrying over possible misfortunes
- I am satisfied
- I feel frightened
- I feel comfortable
- I feel self-confident
- I feel nervous
- I am jittery
- I feel indecisive
- I am relaxed
- I feel content
- I am worried
- I feel confused
- I feel steady
- I feel pleasant

Trait

- I feel pleasant
- I feel nervous and restless
- I feel satisfied with myself
- I wish I could be as happy as others seem to be
- I feel like a failure
- I feel rested
- I am calm, cool, and collected

- I feel that difficulties are piling up so that I cannot overcome them
- I worry too much over something that really doesn't matter
- I have disturbing thoughts
- I lack self-confidence
- I make decisions easily
- I feel inadequate
- I am content
- Some unimportant thought runs through my mind and bothers me
- I take disappointments so keenly that I can't put them out of my mind
- I am a steady person
- I feel confused
- I feel steady
- I get in a state of tension or turmoil as I think over my recent concerns and interest

C.2.2.3 Answer Format

I used a four-point scale to record participants' agreement for the state and trait part of the inventory:

- Strongly disagree
- Disagree
- Agree
- Strongly agree

C.2.3 Ten Item Personality Inventory

C.2.3.1 Items

Items are presented in a grid and prefaced with the statement "I see myself as:"

Extraversion

- Extraverted, enthusiastic.
- Reserved, quiet.

Agreeableness

- Critical, quarrelsome.
- Sympathetic, warm.

Conscientiousness

- Dependable, self-disciplined.
- Disorganized, careless.

Neuroticism

- Anxious, easily upset.
- Calm, emotionally stable.

Openness

- Open to new experiences, complex.
- Conventional, uncreative.

C.2.3.2 Answer Format

- Disagree strongly
- Disagree moderately
- Disagree a little
- Neither agree nor disagree
- Agree a little
- Agree moderately
- Agree strongly

C.2.4 Player Identification Scale

Instructions, answer format, and items were identical with the Player Identification Scale described in A.2.4.

Appendix D

Manuscript D

D.1 Consent Form

Title: Personalizing, adapting, and balancing computer games

Researcher(s): Max Birk, PhD Student, Department of Computer Science, University of Saskatchewan, 306-966-2327, max.birk@usask.ca

Purpose(s) and Objective(s) of the Research: The purpose of this project is to develop models for personalizing and adapting games.

Procedures: In this study, you will be asked to complete some surveys, asking you questions about yourself. Following the survey completion, you will be asked to do several blocks of an experiment task. Following task completion, you will be asked to complete additional questionnaires about your experience. *****Warning: The task includes gruesome imagery. Worker discretion is advised.*****

Funded by: The Natural Sciences and Engineering Research Council of Canada (NSERC).

Potential Risks and Benefits: There are no known or anticipated risks to you by participating in this research. Your participation will help us to design games, including serious games designed to motivate behaviours or support learning.

Confidentiality:

- Confidentiality will be maintained throughout the study. The entire process and data will be anonymized. Data will only be presented in the aggregate and any individual user comments will be anonymized prior to presentation in academic venues.
- Only the principal researcher and her research assistants will have access to the data to ensure that your confidentiality is protected.
- Storage of Data
  - Data (including survey and interview responses, logs of computer use, and videos of interaction) will be stored on a secure password-protected server for 7 years after data collection.
  - After 7 years, the data will be destroyed. Paper data will be shredded and digital data will be wiped from hard disks beyond any possibility for data recovery.

Right to Withdraw:

- Your participation is voluntary. You may withdraw from the research project for any reason, at any time without explanation.
- Should you wish to withdraw, you may do so at any point, and we will not use your data; we will destroy all records of your data.
- Your right to withdraw data from the study will apply until the data have been aggregated (one week after study completion). After this date, it is possible that some form of research dissemination will have already occurred and it may not be possible to withdraw your data

Follow up: To obtain results from the study, please contact Max Birk (max.birk@usask.ca).

Questions or Concerns:

- Contact the researcher(s) using the information at the top.
- This research project has been approved on ethical grounds by the University of Saskatchewan Research Ethics Board. Any questions regarding your rights as a participant may be addressed to that committee through the Research Ethics Office ethics.office@usask.ca (306) 966-2975. Out of town participants may call toll free (888) 966-2975.

D.2 Instruments

D.2.1 Basic Psychological Need Satisfaction

The Basic Psychological Need Satisfaction (BPNS) Scale is a 21-item multidimensional instrument that measures competence, autonomy, and relatedness.

The scale was cited as "Gagne, M. (2003). Autonomy Support and Need Satisfaction in the Motivation and Well-Being of Gymnasts. Journal of Applied Sport Psychology, 15(4), 372–390.". Gagne (2003) investigated the effects of need satisfaction on the well-being of gymnasts in a four week long study.

D.2.1.1 Instructions

All participants received the following instructions: "Please read each of the following items carefully, thinking about how it relates to your life, and then indicate your agreement to each statement."

D.2.1.2 Items

Competence

- Often, I do not feel very competent.
- People I know tell me I am good at what I do.
- I have been able to learn interesting new skills recently.
- Most days I feel a sense of accomplishment from what I do.
- In my life I do not get much of a chance to show how capable I am.
- I often do not feel very capable.

Autonomy

- I feel like I am free to decide for myself how to live my life.
- I generally feel free to express my ideas and opinions.
- In my daily life, I frequently have to do what I am told.
- I feel like I can pretty much be myself in my daily situations.
- I feel pressured in my life.
- There is not much opportunity for me to decide for myself how to do things in my daily life.

Relatedness

- I really like the people I interact with.
- I get along with people I come into contact with.
- I pretty much keep to myself and don't have a lot of social contacts.
- I consider the people I regularly interact with to be my friends.
- People in my life care about me.
- People I interact with on a daily basis tend to take my feelings into consideration.
- There are not many people that I am close to.
- The people I interact with regularly do not seem to like me much.
- People are generally pretty friendly towards me.

D.2.1.3 Answer Format

I used a seven-point Likert-scale to record participants' agreement:

- Strongly disagree
- Disagree
- Slightly disagree
- Neutral
- Slightly agree
- Agree
- Strongly agree

D.2.2 State-Trait Anxiety Inventory

Instructions, answer format, and items were identical with the Player Identification Scale described in C.2.2.

D.2.3 Player Identification Scale

Instructions, answer format, and items were identical with the Player Identification Scale described in A.2.4.

Appendix E

Manuscript A-D

Across all manuscripts I used the same demographic questionnaire and the same questionnaire to measure play behaviour. Because of the variety of answer formats, items and answer formats are presented within one bullet point. Both questionnaires use the following answer formats: Numeric input, text input, single choice, multiple choice, and slider. Numeric input fields only allowed for the input of numbers, e.g., age. Text input fields allowed for the input of numbers and text, e.g., study subject. Single choice items prevented participants from selecting more than one answer, while multiple choice items allowed participants to select between 1 and N items as their answer, where N is the number of total items. The slider response format allowed participants to provide a continuous response between 0 (e.g., 'not at all') and 100 (e.g., 'absolutely') in steps of 10 if not described differently.

E.1 Demographic Questionnaire

The demographic questionnaires gauges age, gender, preferred avatar gender[1], highest degree, employment status, marital status, ethnicity[2].

E.1.1 Items and Answer Format

- What is your age? (Numeric input)
- Indicate your gender: (Single choice)
  - male
  - female
  - other
- Other gender, if applicable: (Text input)
- If you were to create an avatar to represent yourself in digital environments, that digital representation would be: (Single choice)
  - female
  - male
- What is the highest degree or level of school you have completed? If currently enrolled, mark the previous grade or highest degree received: (Single choice)
  - No schooling completed
  - Nursery school to 8th grade
  - 9th, 10th or 11th grade
  - 12th grade, no diploma
  - High school graduate - high school diploma or the equivalent (for example: GED)

[1]The avatar gender was used to assign an avatar when participants indicated 'other' as their sex. While I recognize that the answer format falls short in representing transgender identities and other identities, I felt separating gender identity and avatar identity in my studies would strike a balance between the complex topic of gender identities and implementation demand.

[2]Ethnicity is a complex topic and the used item and answer format leaves room for improvement, i.e., a more fine-grained representation of ethnicity and options for representing dual-identities. See Passmore et al. (2018) for a more nuanced approach of asking for participants' ethnic identity

  - Some college credit, but less than 1 year
  - 1 or more years of college, no degree
  - Associate degree
  - Bachelor's degree
  - Master's degree
  - Professional degree
  - Doctorate degree
  - I prefer not to answer
- If you are a student, please indicate your subject: (Text input)
- Please indicate your employment status: (Single choice)
  - Employed for wages
  - Self-employed
  - Out of work and looking for work
  - Out of work but not currently looking for work
  - A homemaker
  - A student
  - Military
  - Retired
  - Unable to work
  - I prefer not to answer
- Please indicate your marital status: (Single choice)
  - Single, never married
  - Married or domestic partnership
  - Widowed
  - Divorced
  - Separated
  - I prefer not to answer
- Please indicate your household income: (Single choice)
  - Less than $10,000
  - $10,000 to $25,000
  - $25,001 to $45,000
  - $45,001 to $65,000
  - $65,001 to $85,000
  - $85,001 to $100,000
  - $100,001 to $150,000
  - $150,000 or more
  - I prefer not to answer
- Please indicate your ethnicity: (Single choice)
  - American Indian or Alaskan Native

- Asian
- Native Hawaiian or Other Pacific Islander
- Black or African American
- Hispanic/Latino
- White
- Two or more categories
- I prefer not to answer

E.2 Play Behaviour Questionnaire

The Play Behaviour Questionnaire gauges frequency of play in the present, frequency of play in the past, self-identification as a gamer[3], dominant hand, preferred genres, and used devices. identity.

E.2.1 Items and Answer Format

- Please indicate how often (on average) you play games: (Single choice)
  - Every day
  - A few times per week
  - Once per week
  - A few times per month
  - Once a month
  - A few times per year
  - Once per year
  - Not at all
- If you have played games in the past, please indicate how often you have played at peak times: (Single choice)
  - Every day
  - A few times per week
  - Once per week
  - A few times per month
  - Once a month
  - A few times per year
  - Once per year
  - Not at all
- How much do you self-identify as a gamer on the following scale: (Slider)
  - 0: not at all
  - 100: gamer
- Which is your dominant hand?
  - left
  - right

[3]The validation of the single item scale is described in Mandryk and Birk (2017)

- Please indicate the genres that you enjoy playing: (Multiple choice)
  - Action
  - Platform games
  - First Person Shooter
  - Beat 'em up
  - Adventure
  - Role Playing Games
  - Mass Multiplayer Role Playing Games (MMORPG)
  - Simulation
  - Vehicle simulation
  - Strategy
  - Music games
  - Puzzle games
  - Sport games
  - Multiplayer Online Battle Arena (MOBA)
  - Casual games
  - Different genre(s)
- Please indicate on which devices you play: (Multiple choice)
  - Desktop (e.g. Windows, Linux, OS X, etc.)
  - Console (e.g. X-Box, Play Station, etc.)
  - Mobile device (e.g. phone, tablet, PS Portable, etc.)
  - Different device(s)