A Group Recommender System for Avatars in Exergames to Increase Physical Activity

Quentin van Reenen
4611780

November, 2023

Utrecht University
Master Business Informatics
Graduate School of Natural Sciences

Supervisors
Hanna Hauptmann (first)
Sander Bakkes (second)

Abstract

This research investigates the potential effect of a group recommender system for avatars on users' physical activity during the gameplay of an exergame. This study aims to counteract the effects of the increase in sedentary time and decrease of physical activity in people by trying to increase physical activity during the gameplay of an exergame while maintaining or even increasing engagement and satisfaction of the users with the game. The research questions in this research test the influence of a group recommender system on participants' heart rate, perceived experience, perceived fairness, and perceived difference in the ability of the avatars while playing the game Mario Tennis Aces on the Nintendo Switch console. The group recommender system aims to link a group of users with avatars by recommending avatars that best represent the users' abilities. The results indicate that the group recommender system increases participants' heart rate compared to a setting where the users chose the avatar themselves, indicating higher physical activity during gameplay. Moreover, no significant differences in perceived experience, fairness, or difference in ability can be found, which suggests that the recommender system does not affect participants' engagement and satisfaction with the game.

Keywords: Group Recommender Systems, Avatars, Physical Activity, Serious Games, Exergames, Physical Characteristics, IPAQ, Experience, Fairness

Contents

1 Introduction

1.1 Motivation

Many countries initiated nationwide lockdowns when the COVID-19 pandemic hit the world in 2020. This created a standstill for a large part of society. Most research conducted on these lockdowns was on the effectiveness they had on containing the COVID-19 outbreaks (Alfano & Ercolano, 2020; Nafees & Khan, 2020). It was only after a while that researchers started measuring what type of effect these lockdowns have on specific sectors in society, such as the effect lockdowns have on online behaviour, like shopping (Georgiadou et al., 2021; Pathak & Warpade, 2020; Yadav et al., 2020; Zamboni et al., 2021), or even the economy as a whole (Asahi et al., 2021; Joshi et al., 2020). Another focus point of many researchers is finding out what kind of effect a lockdown has on mental health (Adams-Prassl et al., 2020; Banks & Xu, 2020; Morelli et al., 2020; Thakur et al., 2020).

A research direction that is focused on less is the effect that lockdowns have on peoples' lifestyles. Fitness centres and sports clubs are not allowed to open, people are encouraged to work from home and close contact is frowned upon. When the lockdowns initially started, many people used their extra time by going for walks or some other form of physical activity, eating healthier and even trying to fix a bad sleeping rhythm. Unfortunately, as Nguyen et al. (2021) found, these positive changes were short-lived. The COVID-19 lockdown measures ultimately hurt peoples' eating behaviours, dropped their physical activity immensely and, as found by many researchers, their mental health declined (Adams-Prassl et al., 2020; Banks & Xu, 2020; Elmer et al., 2020).

If you take all these negative effects of the COVID-19 lockdowns and take into consideration that, due to the many technological innovations, the sedentary time of most people was already rising, which results in an increased risk of diabetes and other cardiovascular diseases (Hamilton et al., 2008; Wilmot et al., 2012), it is no surprise that the physical state of most people is not, and will not, be in a great place at the moment and will probably worsen in the future.

An idea to counter this effect would be to try to find a way to motivate people in such a way that their physical activity gets a boost (Hills et al., 2010). But being motivated for a longer amount of time is hard when the amount of enjoyment is low (Kilpatrick et al., 2005). That is why something needs to be found that motivates people to perform some physical activity for a longer amount of time while maintaining their levels of enjoyment.

1.2 Problem Statement

Motivation is based on internal and external factors that stimulate the desire and energy of people to commit to reaching a certain goal. Motivation can be split into three categories according to the Self-Determination Theory by Deci and Ryan (2008); Amotivation, Intrinsic motivation and Extrinsic motivation. Amotivation defines the absence of motivation, Intrinsic motivation defines the genuine interest that exists while performing a task and Extrinsic motivation defines the external factors that influence the motivation of a person. As these external factors can be influenced, they might promote someone's Extrinsic motivation. Like Muntean (2011), who found that if a person is rewarded for their actions, the level of Extrinsic motivation increases. Another, more comprehensive practice to improve the degree of Extrinsic motivation could be using other external factors, such as gaming elements.

Considering the previous, the use of gaming elements could be a solution to increase the physical activity of people. This might sound rather counterproductive, as gaming is generally an activity that costs a lot of time and does not really encourage users to carry out any form of physical activity. According to some researchers at the University of Würzburg (2019), there is even a slight correlation between the time that someone plays a video game and their weight. This problem could become even bigger as gaming is on the rise due to the whole COVID-19 pandemic (King et al., 2020), where many people have more free time than usual, which they use to game. Another way to look at this is that, as gaming is already becoming an important part of people's lives, the change needed to be made is in the physical activity of players during the gameplay.

There are already several gaming solutions that urge the user to perform some form of physical activity, these are called exergames (Berkovsky et al., 2010; Douris et al., 2012; Foley & Maddison, 2010; Miller et al., 2014). Exergames can often be played online, which makes it easier for users to socially connect with others (Kaos et al., 2018; Kooiman & Sheehan, 2015), and should therefore be a good candidate to increase the engagement of the users.

As exergames are oftentimes played with other players in an online setting, the exergame section in the gaming world will likely become more multiplayer-oriented. Zagal et al. (2006) already found in 2006 that games that have a collaborative or cooperative nature were becoming more important and this trend has only continued. Because of this, research about increasing the physical activity of users should not focus on the player itself, but on the group of players that is playing the game together.

The optimization of engagement in exergames is extremely important to ensure users' physical activity. One of the possibilities for achieving this is the use of avatars that are representative of the strengths and weaknesses of the users to establish an association between the players and their respective avatars. This has the potential to increase the overall engagement, and therefore create an environment that increases the physical activity of the users.

1.3 Research Questions

How will a group recommender system for avatars, based on the physical characteristics of players, influence the players of an exergame?

1.3.1 Subquestions

1: What kind of circumstances could have an impact on the data?

2: How will a group recommender system for avatars influence the participants' heart rate?

3: Does a correlation exist between a participant's heart rate difference and their perceived experience?

4: How is a participant's perceived experience of the recommender system?

5: Does a correlation exist between a participant's heart rate difference and their perceived fairness?

6: How is a participant's perceived fairness of the recommender system?

7: Does a correlation exist between a participant's heart rate difference and their perceived difference in the ability of the avatars?

8: How is a participant's perceived difference in the ability of the avatars of the recommender system?

2 Literature Review

2.1 Extensive Literature Review

This research will start with an extensive literature review. It will research what recommender systems and group recommender systems are, what avatars and/or games can be used for this research, what the different advantages are between the different avatars and how these avatars, with the help of a group recommender system, can be linked to the physical characteristics of the users.

2.1.1 Protocol

The literature review has been performed in different steps. The first step was selecting the different topics that have been discussed in this review. The second step was systematically searching for literature that concerned these different topics, these sources have been found by using a systematical search in Google Scholar, mainly by selecting sources that have been referenced significantly more compared to similar sources. In the third step, which in reality happened concurrently with the second step, more sources were found by looking at the references of the sources that were found in the second step and by searching for other works of the more common authors found for those topics.

2.2 Recommender Systems

Recommender systems can be classified as a subclass of information filtering approaches. Information filtering approaches aim to reduce information overload by creating systems that automatically remove unwanted and redundant information (Hanani et al., 2001). Recommender systems do this by trying to predict something based on information the users give them (Ricci et al., 2011).

Recommender systems enable users to give some information as input, which the system then processes, computes, and outputs as a recommendation to the appropriate users. The biggest value of a recommender system is the ability to create favourable matches between what is recommended and to whom it is recommended (Resnick & Varian, 1997). The main focus of recommender systems is to turn user data, together with a set of preferences, into a prediction of what the user might like (Lü et al., 2012).

There are many different approaches to a recommender system, but most systems can be classified into one of two different kinds: content-based recommender systems and collaborative filtering recommender systems (Ricci et al., 2011).

2.2.1 Content-Based

The system in a content-based recommender system aims to recommend items to the user that have similarities to the items that the user has preferred in the past. How the similarity of these different items is calculated differs, as there are many different approaches. These approaches mainly focus on the features that belong to the items. For example, if a user has positively rated a song that falls into the dance category, the system could learn to recommend other music from that same category (Ricci et al., 2011).

2.2.2 Collaborative Filtering

The collaborative filtering approach is the best-known and most implemented recommender system that exists. Schafer et al. (2001) called this collaborative filtering approach the "people-to-people correlation". A system that uses the collaborative filtering approach works by aiming to recommend items to users that other users, who look like the user, have preferred in the past. How much a user looks like another user in terms of preference is calculated by finding the similarity between them (Ricci et al., 2011).

Ellenberg (2008) stated that the study of recommender systems was at a crossroads, they stated that it was a field that was originally mainly used by computer scientists and that it is now an enormous field in which the interests of psychologists, physicists and even mathematicians was instigated. Ellenberg was not surprised that, in a recommendation contest organized by Netflix, an approach based on what is known about human behaviour scored very high.

Because the costs concerning data storage and processing are decreasing, recommendation systems are arising everywhere. Generally speaking, if a system has a diverse group of products and the users' preferences are not alike, a recommendation system could be implemented and would help to recommend the preferred products to the right user (Anderson, 2006).

2.2.3 Challenges

Unfortunately, not everything is perfect about recommender systems and researchers in this field do face several challenges which could negatively influence the systems.

The first one is data sparsity, as the number of different items that could be recommended is often so incredibly large that it is hard to find overlapping users. Furthermore, because there are so many items, they do not have enough ratings to be helpful. An algorithm should take this into account (Huang et al., 2004).

The second challenge is scalability, as stated before, there are many (sometimes millions) of items and users, which could heavily increase the system's computational costs. A solution for this would be to use algorithms that, when the data grows, only compute recommendations based on increments of the data (Jin et al., 2009; Sarwar et al., 2002).

The third, and possibly the biggest challenge is the cold start problem. When a new customer starts using the system, the system has no data on this customer. The solution for this would be to ask the customer in advance for some basic information (Lü et al., 2012).

2.3 Group Recommender Systems

Most offline games in the world are games that are collective in nature, people play board games with or against other people. As Zagal et al. (2000) researched, most online games are considered individual. However, a study by that same Zagal, together with others, showed that games that have a collaborative or cooperative nature are becoming more important (Zagal et al., 2006). Kaye and Bryce (2012) suggest that it is important to understand the experience and outcomes of gaming due to its social nature. The results of their research showed that social belonging, opportunity for social networking, and the promotion of social interaction are three things that are very important for game enjoyment. They also found that on the other side, when the social dynamics are poor and the competitiveness is high, it could result in feelings of frustration during the game.

The most common recommender systems are systems that focus on recommending items to single users, like selecting a movie for a user to see by using a model of the preferences of that user. In some cases, as stated above, it is better to not recommend to an individual but to a group, such as a recommendation for a game that is played by multiple people at once.

It is of course harder to recommend to a group than to one user as you often need to deal with different preferences and group dynamics, so the main challenge that needs to be researched is what possible solutions there are to make sure that the recommendation is received best by the whole group. The most obvious strategy would be to take the average of all the individual ratings and give a recommendation of the highest score. This could result in the problem that some recommendations have high averages with some outliers that feel very negative towards the recommendation. There are many possible solutions for this, such as the least misery strategy, where the minimum of the individual ratings is used. Another solution is the multiplicative strategy, where the ratings are multiplied instead of averaged and many more strategies are available (Masthoff, 2011).

2.3.1 Relevance

Focussing on literature about recommender systems is extremely relevant for this study as the main part of this study is based on the development and analysis of a group recommender system. The literature on recommender systems gives an additional understanding of the different techniques used to make recommendations to users, which is important for this study as the group recommender system tries to find the best matching avatars for the players based on their physical characteristics. Furthermore, the choice between content-based and collaborative filtering approaches is important to study. The focus of this research is based on the collaborative filtering approach, as the system in this research recommends avatars to groups of users based on their characteristics.

In conclusion, reviewing recommender systems is relevant as it provides insights into one of the core components of this research: the design and use of a group recommender system for avatars.

2.4 Avatars

An avatar is, generally speaking, the representation of a user. Oftentimes avatars are presented as two-dimensional images when used in an online setting such as social media or other online communities (Blackwood, 2006; Fink, 1999). Avatars can also be presented in a three-dimensional form, this occurs most of the time in games or virtual worlds (Lessig, 2009). Avatars used in a game are often a player's representation of themselves in this virtual world. This representation of one's self could influence the experience of a player.

2.4.1 Experience

This experience is one of the most, if not the most, important part of a game. Research conducted by Lucas et al. (2016) suggests that avatars could play a very important role considering this experience. The research claims that, when an avatar looks like a player, the subjective experience of the player increases. These findings even stated that it was important to note that only the subjective experience of the player increased and that no effect was found on the ability of the players. A somewhat unrelated study invigorates these conclusions. Steinberger et al. (2017) researched if there was a relation to be found between road deaths. They concluded that if somebody's subjective experience is increased, their attention and arousal will increase. Another research by Smallwood et al. (2004) also linked someone's subjective experience to awareness. They researched this by evaluating two different measurements of subjective experience and measuring how they corresponded to different variations in psycho-physiological measures and task performance.

2.4.2 Relevance

This section about avatars and the effects they have on user experience is highly relevant for this research as this research tries to find insights into the impact of a group recommender system for avatars on the physical activity of the users. Avatars are representations of these users in games and are a vital aspect of these games. Gaining insight into the effects that avatars can have on the experience and other aspects such as awareness, arousal and awareness is important. This section provides an extra understanding of the possible effect of the group recommender system for avatars.

In conclusion, the relevance of the avatar section in this literature can be found because it creates an understanding of the potential impact of avatars, with the help of the group recommender system, on the engagement and experience of the users in games, which is an important part of this research.

2.5 Serious games

Serious games are games that are used for other purposes than the main purpose of games, which is entertainment. Serious games could make it possible for users to engage in certain situations that would be impossible in the real world, due to time concerns, safety concerns, costs etc. They could also be used to develop a certain user's skill or skill set (Susi et al., 2007).

The idea of a serious game was introduced by Abt (1970) who was concerned about them as they would have needed to have an "explicit and carefully thought-out educational purpose" and should not have the sole purpose to be played for entertainment.

In 2006, the global market for serious games was estimated at 20 million dollars, and digital gaming was estimated as a 10 billion dollar per year industry (Van Eck, 2006). The assumption was made that this market would only grow exponentially over the coming years.

In the last couple of years, this market has indeed grown exponentially, a market study by Michaud et al. (2010) suggested that the market was already at 1.5 billion dollars and that it had an estimated growth rate of nearly 100% per year.

2.5.1 Exergames

Physical activity in games has been a big talking point in recent times as the problem where increasing amounts of people are getting overweight is getting larger. According to a survey in the United States by the US National Health and Nutrition Institute, more than 60% of adults, and more than 20% of kids are overweight (Berkovsky et al., 2009).

Berkovsky et al. (2009) aimed to design computer games that would increase the engagement and enjoyment of users which should motivate the users to perform (more) physical activity in order to contribute to the decrease of the overweight problem. To incentivize the players to use more physical activity while playing, they proposed that the game should get more challenging over time. When the time would come that the game would be too hard to play, the user could perform some physical activity to help their in-game character. Fujiki et al. (2008) showed how physical activity could be integrated into games, these games are called exergames.

An exergame is a specification of a serious game, they contain all kinds of game elements (goal-driven, rules, challenges, feedback), but they combine these game elements with physical activity. An exergame is, in short, a game that combines exercise with gameplay. Exergames increase the amount of calories burnt, the heart rate and the coordination of the player. Apart from the obvious physical advantages of exergames, they also impact the player's psychosocial and cognitive parts, where they could increase social interaction, motivation, attention span, visualspatial skills and even a player's self-esteem (Staiano & Calvert, 2011).

2.5.2 Relevance

The section on serious games is relevant to this research due to several reasons. Firstly, serious games align perfectly with the main objective of this research, as it tries to increase the physical activity of players with the help of a game. Secondly, exergames share this objective due to their promotion of physical activity through gameplay. This aligns with this study as it involves the use of a recommender system to increase a user's physical activity during the use of the game.

In conclusion, the literature, specifically on exergames, provides a baseline for understanding how games can be used to promote the physical activity of users playing a game.

2.6 Game

The game used in this research will be Mario Tennis Aces. Mario Tennis Aces is a sporting game played on the Nintendo Switch and can be played by a maximum of 4 players at a time, offline and online. As the name suggests this is a tennis game where the players play a game of tennis in a group of 4, 2 versus 2. This game has a specific exergame mode called "Swing mode" which promotes physical activity as the joy-con (the remote controller for the Nintendo Switch) is used to register the movements of the player (Nintendo, 2022).

2.6.1 Relevance

This game is a good choice for this research as it possesses all the requirements for this study. It is a multiplayer-oriented game with an exergame mode complemented with playable characters that differ in ability.

2.7 Physical Characteristics

To be able to link the physical characteristics of a user to an in-game avatar, the physical characteristics of the user need to be found. There are several possibilities to find the physical characteristics of a user.

2.7.1 Body Mass Index

One way to measure some physical characteristics of a user is with the help of the Body Mass Index (BMI). A person's BMI is measured from a person's weight (Mass) and their height and is reported as $kg/m^2 =$ BMI (World Health Organization et al., 2005).

BMI is a very simple index that is often used as a way to categorize if a person is underweight, normal weight, overweight or obese. Unfortunately, due to the simplicity of the index, the BMI has some limitations. BMI is deemed less accurate and may overestimate the body fat in people who are younger and/or more athletically built. On the other hand, older people, who tend to lose weight due to muscle loss could also be measured incorrectly (Kok et al., 2004).

Therefore, BMI is a simple yet ineffective tool to be able to be used as the measurement tool for this research, as it is too unreliable to measure someone's physical characteristics.

2.7.2 Muscle Strength Tests

A muscle strength test is often used to determine muscle functions in sports and exercises but is also occasionally used in other movement-related sciences. It is defined as the maximum force or torque that occurs when a maximal voluntary contraction is given from a muscle or muscle group (Sale & Norman, 1991).

Unfortunately, two problems occur when muscle strength tests are performed. The first problem that occurs is that most studies that have presented data based on these tests were not normalized or used different normalization techniques for datasets, which makes it hard to compare different results. The second limitation is the, quite important, influence that body size has on the measurement. A bigger person will generally score higher in the muscle strength tests than a smaller person, even if that person has worse physical characteristics. This oversight has been neglected when the functional movement had to be determined by measuring someone's muscle strength (Jaric, 2002).

Due to these limitations, especially the second one, it is quite hard to justify using this technique to find the physical characteristics of the users and is therefore not used in this research.

2.7.3 Standardized Physical Measurement Tests

As stated in the previous section, one of the limitations of muscle strength tests is the absence of standardization of the used tests. One of the options that circumnavigates this limitation is the International Physical Activity Questionnaire (IPAQ). The IPAQ is a widely used questionnaire that has been extensively reviewed. The IPAQ consists of 4 subgroups of questions to provide a questionnaire that enables the user to obtain data on a participant's physical activity (Craig et al., 2003).

A special version of the IPAQ, The International Physical Activity Questionnaire - Short Form (IPAQ-SF) has been created as a cost and time-effective tool to measure someone's physical activity. Unfortunately, P. H. Lee et al. (2011) tested the validity of the IPAQ-SF and found that the correlation between the measure of activity was lower than an acceptable standard, they also found that the questionnaire generally overestimated physical activity. These findings led to the conclusion that the IPAQ-SF, while being cost and time-efficient, is not the correct tool for this research.

The IPAQ also has a long version, the IPAQ-LF. This questionnaire consists of 5 subgroups of questions that ask about different activity domains specifically (Craig et al., 2003). In this research, the IPAQ-LF will be used.

3 Method

The research questions will be answered with the help of a group recommender system, the game Mario Tennis Aces, heart rate monitors and several different questionnaires. Research subquestion 1 is used to determine if certain circumstances can be found that influence the data, the main focus is on whether the collected data is influenced by the order in which the recommendation is given or the duration of the games. Research subquestion 2 will be answered with the help of the data from the heart rate monitors, which can measure the physical activity of each participant. Research subquestions 3 and 4 are answered with the help of a questionnaire based on the questionnaires by Knijnenburg et al. (2012) and Masthoff (2015). Research questions 5, 6, 7 and 8 will also be answered with a questionnaire, these questions will be created specifically for this research. This research is structured in such a way that the heart rate measurements of the groups of participants will result in a dependent test, this way the groups are measured by themselves and compared with themselves. The data resulting from research subquestions 3, 5 and 7 will be analysed with a correlation test and the data related to subquestions 4, 6 and 8 will be answered using their descriptive statistics.

3.1 Design Science

The Design Cycle framework created by Wieringa (2014) will be used in this research as a guide. The framework consists of several steps, which are shown in Figure 1. The framework can be divided into three main steps: problem investigation, treatment design, and treatment validation.

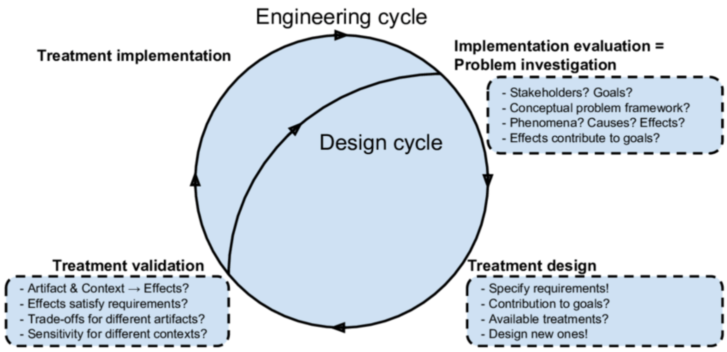

Figure 1: Wieringa's Enginereering Cycle

3.1.1 Problem Investigation

In the problem investigation step, the focus is on the exploration of the problem. This step allows for a better understanding of the things that need to be taken into account during the research and the different underlying challenges.

To explore the possible solutions for the proposed problem a literature review is conducted. The problem that this research will try to address is the fact that in general, there is a decrease in physical activity of people and that the optimization of engagement in exergames can be a good way to ensure an increase in physical activity of the users.

3.1.2 Treatment Design

In the treatment design step, the artefact, or multiple artefacts, are created to resolve the earlier identified problem in the problem investigation phase.

In this research, the main focus of the treatment design phase will be on the creation of the different questionnaires and the creation of the group recommender system.

3.1.3 Treatment Validation

The created artefact or artefacts are tested in the treatment validation phase. They are evaluated to determine their effectiveness in solving the problem. This process oftentimes creates new insights that can result in a new problem evaluation, and therefore, a new iteration of the Design Cycle.

This research will start the treatment validation phase with a focus group, which will result in newfound insights into the problem, the validation of the created artefact will be done with the help of data analysis.

3.2 Participants

To be able to test the created artefacts, participants are necessary. So the first step in the treatment design is creating the criteria for these participants. There are not many criteria for the participants in this research. The participants will need to have some experience with video games in general as that will simplify explaining the game. The participants are not allowed to have any prior experience with the game, Mario Tennis Aces, as that would result in an unfair advantage for them and their teammate. Furthermore, participants that have a significant physical hindrance will also not be selected to partake in this research as they will not be able to play the game in the so-called "Swing Mode". The participants will participate in this study in groups of four and play the game in a two-versus-two configuration. The groups cannot be bigger as the chosen game does not facilitate more than 4 players at a time. The number of participants in this research is planned to be 32 but as it was planned to have an equal balance between the groups of 4 it has led to 24 participants, which in turn will result in six groups of four participants. The participants will be selected randomly but the participants in this research will likely be selected with convenience sampling.

3.3 The Game

The next part of the treatment design is the decision of which game and, consequently, which avatars will be used to test the created group recommender system. The game used in this research is Mario Tennis Aces. Mario Tennis Aces is a tennis-like game created for the Nintendo Switch game console. The game won the prize of "Sports Game of the Year" in 2019 (McWhertor, 2019). The game consists of playing tennis games against other players or the game's AI. In this game, a character is chosen by the player which could result in different experiences, as the characters in this game differ in ability. The characters are grouped in several different classes of play style: *All-Around, Powerful, Defensive, Speedy, Technical and Tricky.* The most general characters of each class will be selected for this research, with the help of the physical characteristics questionnaire and a group recommender system both explained in one of the next paragraphs, the players will be linked to these characters.

3.3.1 Avatars

As stated before, the Mario Tennis Aces game distinguishes between different classes of avatars, in this section the different classes and the avatar(s) from that class that are selected for this research will be explained. Some classes will have 2 possible characters selected, this is because there is a male and female option for that class. Whether a participant gets recommended either one of these characters will not rely on the gender of the participant, but purely on their preference, this will be explained further in the questionnaire and group recommender system sections below.

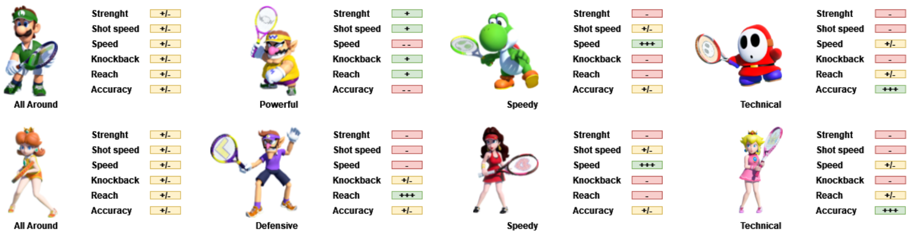

Figure 2: Selected avatars for this study with their characteristics.

All-Around

The first class is also the most general class. The *All-Around* class consists of avatars that have very balanced character statistics. The *All-Around* class is made up of four characters; Mario, Luigi, Daisy and Birdo. Mario is not chosen for this research as Mario is too popular. Mario is the most popular video game character ever created (Bhasin, 2012) and is recognised by more than half of the world's population (WatchData, 2020). This popularity could influence the results of this research and therefore Mario is not selected for this research. Birdo is a pink dinosaur and will also not be selected for this research. Therefore, two other characters remain, Luigi and Daisy. As Luigi is a male character and Daisy a female character, both of these avatars are selected for this research.

Powerful

The *Powerful* class is the second class. As the name suggests, these characters are more powerful. This increased power statistic results in strong and fast shots. To add to the increase in power, these characters are also harder to knock back and generally have greater reach due to their physical size. To balance these characters, they also have some disadvantages. The main disadvantage of *Powerful* characters is that they generally lack a lot of speed. The character chosen for this category is Wario, as the Wario character is the most human-like. No female characters are chosen for this category as there are no *Powerful* avatars that are female.

Defensive

The third class is the *Defensive* class. The characters of this class have a longer reach than other characters from different classes. Due to this longer reach, they can easily get to balls and are therefore harder to score points against. The characters of the *Defensive* class generally lack speed and power, to counterweigh this advantage of a longer reach. Waluigi is chosen to represent this category as he is the most human-like. There are no female characters in this class.

Speedy

The fourth class is the *Speedy* class, and as the name suggests, the characters in this class have more speed than other characters. This helps the characters move faster on the field. This advantage gets balanced out by the fact that they are easier to knock back, and do not have the same reach or power as many other characters. The characters chosen in this category are Yoshi (for the male option) and Pauline (for the female option).

Technical

The fifth class is the *Technical* class. The *Technical* class consists of characters that generally have an increased accuracy which results in higher ball control and better aimed shots. To counterbalance these advantages, they generally lack strength and are therefore easier to knock back. The characters used for this research are Shy Guy for the male option and Peach for the female option.

Tricky

The last character class is called *Tricky*. The *Tricky* class, as the name suggests, is a bit harder to master. Every shot of a *Tricky* player will result in a shot with a curve. This is generally their only advantage and can only really be mastered if the game is played for a longer time. Since this research uses participants who have no experience with this game, this class is omitted from the research.

3.4 Physical Characteristics Questionnaire

An imperative part of the treatment design phase is the artefact that helps link the physical characteristics of the participants to, with the help of the group recommender system, the characteristics of the avatars. To measure the physical characteristics of the participants, a questionnaire will be used. The International Physical Activity Questionnaire (IPAQ) measures the physical activity of the participant performed in the last 7 days. The IPAQ is a widely used questionnaire and has been extensively reviewed. In this research the Long-Form specification of the IPAQ will be used (IPAQ-LF), this questionnaire consists of 31 questions divided into 5 subgroups (International Consensus Group, 1998). This questionnaire can be found in Appendix A: Questionnaire 1.

3.4.1 Avatar Preference

At the end of the first questionnaire, all possible characters are shown to the participant with the help of images and the question is asked to order these characters based on their preference from top to bottom. The used images can be found in Figure 2.

3.5 Group Recommender System

When the physical characteristics of the participants are found, a group recommender system will be used to link the physical characteristics of the users to the avatars from the game. The system will try to recommend the best representative avatars for the group based on the characteristics of these avatars. The group recommender system in this research will be built for this specific purpose. This artefact is the most important part of the treatment design phase as this artefact will be the main focus of this research.

3.5.1 System Design

The design of the group recommender system is based on the literature and customised towards the results of the physical characteristics questionnaire and the Mario Tennis Aces game. The recommender system follows several steps to, eventually, get an avatar recommendation for each participant. The compact summary of all the different steps can be found in Appendix C.

No design choices were made in the first 5 steps, as these steps consist of loading and cleaning the data stemming from the Physical Characteristics Questionnaire.

In steps 6 and 7 the participants are ordered based on each activity intensity per week and their relative positions are logged. The relative positions are chosen because the absolute differences between the participants do not matter in the rest of the recommender system. An example of this would be that, even if there are many "strong" participants the "strongest" participant would get recommended the powerful character as each avatar class can only be chosen once (see step 10 for an explanation).

The recommender system uses 5 classes of the Mario Tennis Aces game in step 8; All-Around, Powerful, Defensive, Speedy and Technical. The recommender function (*recommender_system*) takes a participant at random and checks if their positions fall within the restrictions of a specific class. The restrictions of each class are based on their respective strengths and weaknesses.
Example: if somebody has a high place in *Vigorous_Activity* and *Sitting*, but a lower place in *Moderate_Activity*, they are likely to be recommended a powerful character. See Figure 3 for a specification for each

class.

All-Around:	Strenghts:	None
	Weaknesses:	None
Powerful:	Strenghts:	Power -> Vigorous high
	Weaknesses:	Speed -> Sitting high and/or Moderate low
Defensive:	Strenghts:	Reach -> Walk high
	Weaknesses:	Power & Speed -> Vigorous low and/or Sitting high and/or Moderate low
Speedy:	Strenghts:	Speed -> Sitting low and/or Moderate high
	Weaknesses:	Power -> Vigorous low
Technical:	Strenghts:	Accuracy -> Walk high
	Weaknesses:	Power -> Vigorous low

Figure 3: Specification of each class

The design choices made to link these activity intensities to the different abilities were done as follows:
The first link was made between power and vigorous activity as Norton et al. (2010) stated that vigorous-intensity activity can be defined as physical activity which makes the user "breathe harder or puff and pant" with examples that include heavy lifting, which can be directly linked to someone's strength and thus, power.
The second link was made between moderate activity and speed. Statton et al. (2015) stated that speed can be linked to the maximal oxygen uptake. The general consensus in the research field is that the maximal oxygen uptake can be linked to both vigorous activity and moderate activity (Arboleda-Serna et al., 2019; Tjønna et al., 2013). The reason for linking the speed characteristic to moderate activity is that both referenced studies concluded that moderate activity had a greater effect on the maximal oxygen uptake versus vigorous activity, but no significant differences were found. But, as vigorous activity was already linked to power, and moderate activity generated slightly better results, moderate activity was linked to the speed characteristic.
The next link was made between sitting and a weakness in speed. Many studies can be found that link sedentary time with someone's body weight (Caballero, 2004; Elgar et al., 2005; Must & Tybor, 2005). A higher body weight will negatively influence someone's speed, for this reason, the link was made between sitting and a weakness in speed.

The last link was made between walking activity and the accuracy and reach characteristics. The main reason for these links was the fact that these characteristics needed to be linked to at least one activity, and this was the activity that was still available

In step 9, if the system does not find a matching class for the participant's data, the All-Around class gets recommended and the participant is put into the *recommender_system* function again, this time with more lenient restrictions. This recursiveness in the function can only happen twice, so a maximum of 3 iterations takes place. The iteration is logged to make sure that the strongest recommendations (the recommendations with a low amount of iterations) are found. If after 3 iterations no specific class is found, the All-Around class gets recommended. An example of the resulting DataFrame is shown in Figure 4.

recommender_1	recommender_2	recommender_3
All-Around	Defensive	Powerful,Defensive,Technical
All-Around	All-Around	Powerful,Defensive
All-Around	Speedy	Speedy
All-Around	Powerful,Defensive	Powerful,Defensive

Figure 4: DataFrame after *recommender_system* function

The choice of which class eventually gets recommended in step 10, has undergone a specific design choice, as only one of each class can be chosen in this scenario. Studies suggest that avatar abilities can affect users that interact in a group, A lack of uniqueness in these avatars can counter-intuitively lead to users striving for individuality instead of working with the group towards a goal (Kim, 2010, 2011; Kim & Park, 2011). For these reasons, the choice to only use one avatar of each class for the research was made to make sure that the differences in the ability

of each avatar were as explicit as possible. How these choices are made can best be described with an example based on Figure 4:
Start by looking at the first iteration (all participants are recommended the All-Around class, this is not very remarkable since the restrictions to get recommended a specific class are very strict in the first iteration of the system. When looking at the second iteration, it is clear that three out of the four participants can be recommended a specific class, so those are logged:
- 1 -> Defensive
- 2 -> Nothing (yet)
- 3 -> Speedy
- 4 -> Powerful (as Defensive is already linked to the first participant)
Then the third iteration is taken into account, if the participant gets recommended a class that is still available then this class is chosen, this is not the case, and since the Powerful and Defensive recommendations are stronger in the second iteration for the other participants, the second participant will be recommended a character from the All-Around class.

In step 11, the selected class for each participant is inputted into the DataFrame.

Then, for each participant, their class is checked in step 12. If their recommended class has two possible characters, the system checks which of those characters got the highest preference from that specific participant and is selected. If the recommended class has one possible character, that character is selected. The possible characters for this research were chosen with three things in mind. Firstly, to make sure that the participants identify as much as possible with the character, only the characters that best-represented people were chosen for each class (Lucas et al., 2016). The second decision was that if a class had a possible male and female character, both characters would be possible to play with (based on someone's personal preference). Lastly, Mario is not chosen for this research as Mario is too popular. Mario is recognised by more than half the world population (WatchData, 2020) and is the most popular video game character ever (Bhasin, 2012). This could influence the results of this research and therefore Mario is not chosen for this research. In the last step, the system prints the participants' ID and the name of the recommended character, see Figure 5.

```
1: Waluigi
2: Luigi
3: Yoshi
4: Wario
```

Figure 5: Group Recommender System output

3.6 Focus Group

A focus group was used to gain insights into potential flaws of the research, this is a perfect example of treatment validation, as the insights gained have helped with some additional findings for the problem investigation and some changes in the treatment design. The participants started by signing an informed consent. Then, they filled in the physical characteristics questionnaire that is based on the research by Craig et al., 2003. The group recommender system was used to link the physical characteristics of the users to the avatar characteristics in the game. Once the recommendation was shown to the participants, their reaction was recorded. The participants were further asked about what they would expect from their physical activity and experience whilst playing with this character. Then with the help of an image (Figure 2), the differences in the ability of each of the possible characters were shown. This image helped the participants understand the differences in strengths and weaknesses of each avatar. Once some background information was given, the participants were asked about the perceived social fairness, and if they expected a change in their physical activity due to this avatar recommendation. During the focus group, no active gameplay was involved. The focus group did not yield any important changes to the system. The only request from the focus group was a bit more information about the game and the swing mode beforehand. This was executed by showing two videos to each of the participant groups before playing the game. The first video was the original trailer for the Mario Tennis Aces game (Nintendo UK, 2018). The second video explained the Swing Mode (GameXplain, 2018).

3.7 Playing the Game

After the group recommender system has given all the players their recommended avatar, the participants will put on heart rate monitors on their non-dominant hand, the aforementioned videos are watched, and the game is played 2 times: once with avatars chosen by the participants themselves, and once with the recommended avatars. This way two measurements will be taken; a base and a recommender measurement. The order of these two measurements is random. This is part of the treatment validation phase in the design cycle framework as this part of the method is an element of the test of the created artefacts, just like the three questionnaires about the perceived experience, social fairness, and difference in ability below.

3.8 Experience Questionnaire

As stated above, to be able to answer research subquestions number 3 and number 4, a combination of two surveys will be used. The first questionnaire that is used is the questionnaire by Knijnenburg et al. (2012). This questionnaire has a subgroup of questions concerning the experienced recommendation quality and satisfaction of the users. The other survey that is used is the survey created by Masthoff (2015). This survey measures the satisfaction of users after watching video clips. The questions used in this research will be adjusted slightly to be able to be in line with the experience of the used system. This questionnaire can be found in the first part of Appendix B: Questionnaire 2.

3.9 Social Fairness Questionnaire

As the perceived social fairness has not (yet) been measured in gaming contexts, there is no questionnaire available. That is why the questions to measure the perceived social fairness for research subquestions 5 and 6 will be created specifically for this purpose. The concept of fairness alone, however, has been researched before (Chiu et al., 2009; Konietzny & Caruana, 2019), these questionnaires are useful for this research and the questions for this research will be based on this. This questionnaire can be found in the second part of Appendix B: Questionnaire 2.

3.10 Difference in Ability Questionnaire

To measure the difference in ability, and answer research questions number 7 and 8 about the difference in ability, another set of questions will be used. These questions will ask the participants directly if they had the feeling that the different abilities of the avatars influenced their physical activity and experience. This questionnaire can be found in the last part of Appendix B: Questionnaire 2.

3.11 Data Analysis

At the end of the treatment validation phase, the measurements and data from the research will be analysed using statistical tests in Python. Dependent Samples T-tests will be performed to find if the recommender system creates a difference in measurements before and after the recommender system is used. Pearson correlation coefficients will be used to test the potential correlation between heart rate and perceived experience, fairness and difference in ability. Descriptive statistics will be used to measure the perceived experience, fairness and difference in ability directly.

4 Results

4.1 Subquestion 1: What kind of circumstances could have an impact on the data?

H_0: There is no circumstance that has an impact on the data.

H_1: There is a circumstance that has an impact on the data.

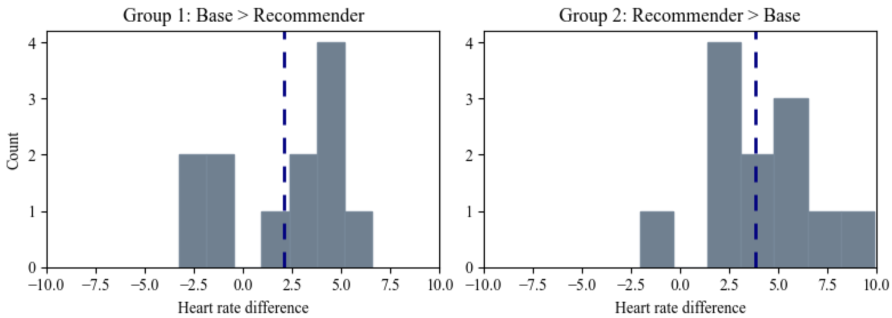

Figure 6: Heart rate distribution between participant groups

To test whether starting or ending with the recommended avatar has an impact on the data, the two groups were compared. Before a T-test could be performed, the groups needed to be tested for normality. A Shapiro–Wilk test resulted in p-values of both 0.21 and 0.97. As $p > 0.05$ for both measurements, a parametric test can be used. The two-tailed T-test between the groups resulted in a p-value of 0.16685. The p-value is not significant as $p > 0.05$.

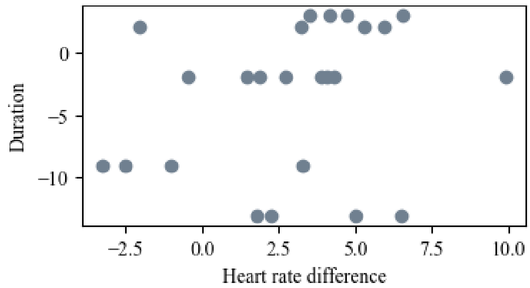

Figure 7: Heart rate distribution compared to duration

To test if the duration of the game has an impact on the data, a Pearson correlation coefficient was computed between the heart rate difference of a participant and the duration difference. The correlation coefficient was 0.25 with a p value of 0.25. A weak, positive correlation is found, but $p > 0.05$ so this correlation is not significant. As both circumstances result in p values that are larger than 0.05, the ***H_0*:** is accepted: There is no circumstance that has an impact on the data.

4.2 Subquestion 2: How will a group recommender system for avatars influence the participants' heart rate?

***H_0*:** The recommender system for avatars does not influence participants' heart rate.

***H_1*:** The recommender system for avatars influences participants' heart rate.

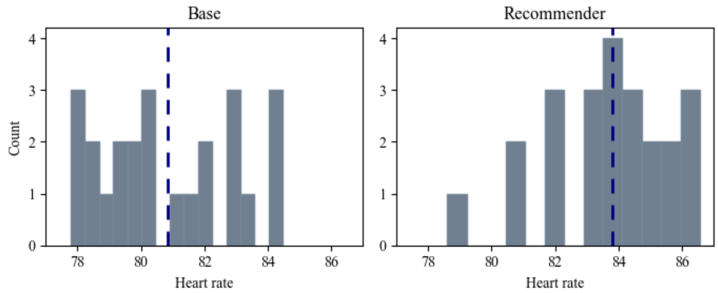

Figure 8: Heart rates for base measurement and recommender measurement

The measurements revealed a mean heart rate difference of 2.96 between the base and recommender groups, with a standard deviation of 3.15. To determine the significance of this variation, a T-test was employed.

Normality for both the base and the recommender groups was tested with the help of a Shapito-Wilk test, which yielded p-values of 0.15 for the base group and 0.76 for the recommender group. Since both p-values are larger than the significance level of 0.05, which indicates that the data in both groups is distributed normally, a parametric test is used. A two-tailed T-test was conducted, which resulted in a p-value of 0.00001.

As the computed p-value is less than the significance level ($p < 0.05$), the results are considered to be significant. Therefore, for subquestion 2, the $\boldsymbol{H_0}$**:** is rejected and the $\boldsymbol{H_1}$**:** is accepted: The recommender system for avatars influences participants' heart rate.

4.3 Subquestion 3: Does a correlation exist between a participant's heart rate difference and their perceived experience?

$\boldsymbol{H_0}$**:** There is no correlation between a participant's heart rate difference and their perceived experience.

H_1: There is a correlation between a participant's heart rate difference and their perceived experience.

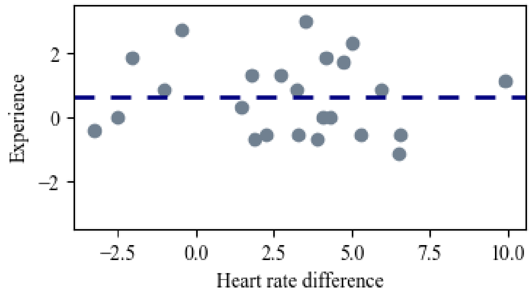

Figure 9: Heart rate distribution for perceived experience

To test if there is a correlation between the heart rate difference and the perceived experience, a Pearson correlation coefficient was used. The correlation coefficient was -0.07 with a p value of 0.73. A very weak, negative correlation is found, but $p > 0.05$ so the correlation is not significant. For subquestion 3, the ***H_0:*** is accepted: There is no correlation between a participant's heart rate difference and their perceived experience.

4.4 Subquestion 4: How is a participant's perceived experience of the recommender system?

This is a descriptive question so no hypotheses were formed.

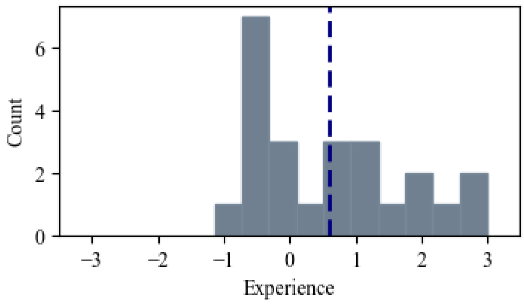

Figure 10: Distributon of perceived experience

In order to investigate a potential effect of the recommender system on a participant's perceived experience, the participants were asked directly about their experience with the recommended avatars, with the help of statements that could be answered on Likert scales ranging from -3 to 3. It is expected that the mean would be 0 if the perceived experience was the same in the base state as the recommended state. A mean of 0.61 is found, with a standard deviation of 1.2.

4.5 Subquestion 5: Does a correlation exist between a participant's heart rate difference and their perceived fairness?

***H_0*:** There is no correlation between a participant's heart rate difference and their perceived fairness.
***H_1*:** There is a correlation between a participant's heart rate difference and their perceived fairness.

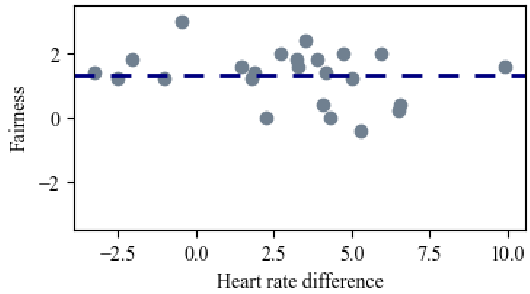

Figure 11: Heart rate distribution for perceived fairness

To test if there is a correlation between the heart rate difference and the perceived fairness, a Pearson correlation coefficient was used. The correlation coefficient was -0.25 with a p value of 0.24. A weak, negative correlation is found, but $p > 0.05$ so the correlation is not significant. For subquestion 5, the $\boldsymbol{H_0}$**:** is accepted: There is no correlation between a participant's heart rate difference and their perceived fairness.

4.6 Subquestion 6: How is a participant's perceived fairness of the recommender system?

This is a descriptive question so no hypotheses were formed.

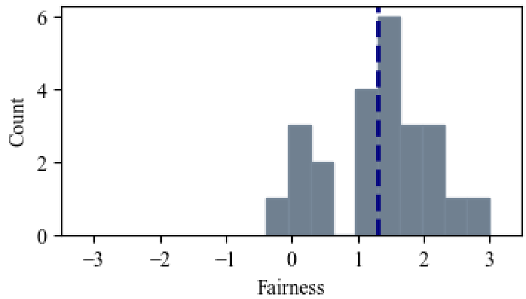

Figure 12: Distributon of perceived fairness

In order to investigate a potential effect of the recommender system on participants' perceived fairness, the participants were asked directly about their experience with the recommended avatars, with the help of statements that could be answered on Likert scales ranging from -3 to 3. It is expected that the mean would be 0 if the perceived fairness was the same in the base state as the recommended state. A mean of 1.3 is found, with a standard deviation of 0.83.

4.7 Subquestion 7: Does a correlation exist between a participant's heart rate difference and their perceived difference in the ability of the avatars?

H_0: There is no correlation between a participant's heart rate difference and their perceived difference in the ability of the avatars.
H_1: There is a correlation between a participant's heart rate difference and their perceived difference in the ability of the avatars.

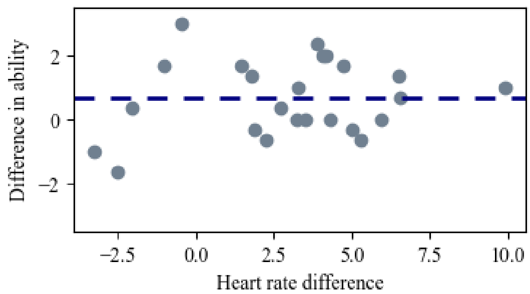

Figure 13: Heart rate distribution for perceived ability difference of the avatars

To test if there is a correlation between the heart rate difference and the perceived difference in ability of the avatars, a Pearson correlation coefficient was used. The correlation coefficient was 0.18 with a p value of 0.40. A very weak, positive correlation is found, but $p > 0.05$ so the correlation is not significant. For subquestion 7, the ***H_0:*** is accepted:

There is no correlation between a participant's heart rate difference and their perceived difference in the ability of the avatars.

4.8 Subquestion 8: How is a participant's perceived difference in the ability of the avatars of the recommender system?

This is a descriptive question so no hypotheses were formed.

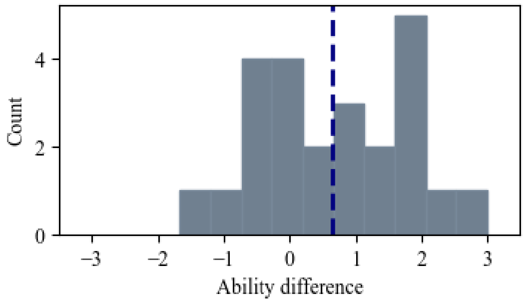

Figure 14: Distributon of perceived difference in ability

In order to investigate a potential effect of the recommender system on a participants' perceived difference in ability, the participants were asked directly about their experience with the recommended avatars, with the help of statements that could be answered on Likert scales ranging from -3 to 3. It is expected that the mean would be 0 if the perceived difference in ability was the same in the base state as the recommended state. A mean of 0.65 is found, with a standard deviation of 1.17.

5 Discussion

This study first explored the potential influence of two circumstances on the data. Notably, no statistical differences were found. The order in which a participant was exposed to a recommended avatar (starting or ending) and the duration of the game appeared to have no impact on the measurements, and thus, the outcomes.

This was in line with the expectations, considering that the participants were given a period of rest between the two measurements. This period of rest was designed to let the participants' heart rate return to a baseline rhythm before initiating the second measurement (Buddies, 2014). By standardizing the starting conditions of both measurements, the potential effects that could have arisen from this were minimized.

Furthermore, the result that the duration of the game does not affect the heart rate can be explained by an interesting physical phenomenon. During the game, participants' heart rates would rise quickly towards a stable range (Fletcher et al., 2001). So the duration of the game did not affect the heart rate of the participants once the participants' heart rates reached their average range.

5.1 Heart Rate

The main focus of this study was to investigate the potential impact of a recommender system on a participant's heart rate, specifically compared to a state in which the participant had the freedom to choose an avatar themselves. Comparing the two measurements revealed a significant result, indicating that the heart rate was indeed higher when participants used the recommended avatar.

This higher heart rate can possibly be explained due to the fact that, when participants use an avatar that closely represents their physical ability or characteristics, their engagement increases (Birk et al., 2016; Birk & Mandryk, 2019; Dechant et al., 2021). With this increase in engagement, they are likely to have an increase in their physical activity compared to a normal state (S. Lee et al., 2017; Lyons, 2015).

5.2 Counterintuitive Effects

This study also aimed to explain whether a difference in heart rate due to the use of recommended avatars would therefore lead to a decrease in participants' perceived experience during gameplay. Additionally, the participants' perception of fairness concerning their recommended avatars and the recommended avatars of the other participants would be measured. On top of that, as the different avatars have different abilities, it was investigated if this would impact the participants.

Measuring these three themes allows for a better interpretation of the results, as it would be counterintuitive if the recommended avatar had the desired effect on the participants' heart rate but would decrease any of these three themes. This could therefore result in a decrease in engagement. This would mean that, in the long run, the recommender system would only decrease physical activity.

5.2.1 Experience

The measured, negative, Pearson correlation between heart rate difference and perceived experience suggests a very weak, but non-significant correlation between these variables. This indicates that the differences in heart rate due to the recommender system do not significantly correlate to the differences in the participants' perceived experience. Furthermore, the slight positive mean in the perceived experience value indicates that participants had the same or even a slightly better experience in the recommended state. This is demonstrated by Clay et al.

(2022) who state that someone's experience will only be influenced by the amount of effort when they are rewarded for it. This is not the case in this research as the only reward will come from winning the game, which will always be either one of the teams. On top of that, with the other team getting negative feedback due to losing, their experience will likely be influenced a bit more negatively (Zhang et al., 2023), which will counteract the potential positive experience of the winning team.

5.2.2 Fairness

The measured, Pearson correlation between heart rate difference and perceived fairness suggests a weak, but non-significant correlation between these variables. This indicates that the differences in heart rate due to the recommender system do not significantly correlate to the differences in the participants' perceived fairness. Furthermore, the positive mean in the perceived fairness value indicates that participants had slightly better perceived fairness in the recommended state. This can be explained by the fact that Mario games, or essentially all games, are considered fair at their core, which in turn, would then not influence the effort of a user.

5.2.3 Ability difference

The measured, Pearson correlation between heart rate difference and perceived difference in ability suggests a very weak, but non-significant correlation between these variables. This indicates that the differences in heart rate due to the recommender system do not significantly correlate to the differences in the participants' perceived difference in ability. Furthermore, the slight positive mean in the perceived difference in ability value indicates that participants had the same or even a slightly better perceived difference in ability in the recommended state. The better perceived difference in ability in the recommended state could be explained by the idea that the user would identify better with an avatar that had abilities that were in line with the user's ability (Peña & Kim, 2014). On the other hand, Peña et al. (2016) found that users compare their avatars to the opponent's character and found a decrease in physical activity when the avatars differed from each other. These two counteracting principles could explain why this research did not find any significant correlation between perceived ability difference and heart rate.

5.3 Limitations

This study contains certain limitations, the primary limitation comes from the relatively small participant group. The participant group consisted of 24 individuals, split into 6 groups of 4. In an ideal scenario, a larger sample size would have been desirable. However, the unique scenario, in which the data could only be collected if 4 participants, who were familiar with each other, were present concurrently, prevented the formation of additional groups.

The second limitation refers to the imbalance in the number of male participants (20) versus female participants (4). The main consequence of this skewed distribution is that it is hard to draw a definitive conclusion about the general population. The calculated effects may only be significant for the male population, given their large representation in this research.

The third limitation of this research relates to the heart rate monitors that were used during the data collection. When the research was conducted a couple of times, a pattern was observed wherein participants' heart rates were never lower than 65 and were oftentimes reaching 99 beats per minute, but never surpassed this. This could indicate potential inaccuracies in the heart rate monitors. Furthermore, in some cases, the heart rate monitors took a bit longer to finish a measurement, which resulted in the fact that not all measurements were taken exactly at the same time.

The next limitation of this study comes from the data collection. The measurements for perceived experience, fairness, and difference in ability were only collected for the recommended state, and not for the base state. Due to this limitation, no statistical tests could be performed, and only demographic numbers could be shown.

The last limitation arises from the selection of the game, Mario Tennis Aces, for this study. Even though the game creates an optimal testing space for the recommender system, it is hard to draw definitive conclusions about the perceived fairness and perceived differences in ability. One of the main characteristics of Mario games is the balance between characters, so the users will always have the feeling that the game is fair and that the abilities between the characters are balanced.

5.4 Future Work

This study enables researchers to explore several possibilities by addressing certain limitations. This section outlines the potential directions researchers can take to increase the knowledge of the effect of group recommender systems for avatars in exergames.

The main thing to focus on in a future study is increasing the sample size, while concurrently making it more diverse. Efforts should be made to be able to measure more participants and include a balanced representation between male and female participants. However, it is important to note that acquiring such a sample could present some challenges, especially when trying to find participant groups that represent a friend group, as those groups tend to be the groups that play games together.

Another aspect for future work is the improvement and use of more accurate heart rate monitors. It would be an option to use medical-grade wearable devices or even chest straps to get a more accurate heart rate measurement. Additionally, exploring other possibilities for the measurement of physical activity, beyond heart rate, would be a valuable extension of the data collection.

In order to gain better insights into participants' perceived experience, perceived fairness, and perceived difference in ability it is important to start with a base value for these three variables. This will strengthen the data collection, and therefore, the results of the study.

Future studies should explore the possibility of other exergames beyond Mario Tennis Aces. Testing a variety of games, with different gameplay mechanics and physical activity levels, can provide a better understanding of the effectiveness of the group recommender system used in this study. This, in turn, would also enable the researchers to focus on another aspect, namely character customisation.

Another possible future research could focus on the group recommender system itself. The main aim of the current system is to find a group of avatars for a group of users that represent their physical characteristics the best. On the other hand, research could be initiated to study the effect of links between the users and the avatars where the avatar does not represent their physical characteristics. This would probably result in less identification between the user and the avatar, which could affect engagement, but it could also increase motivation as users might try to compensate for the difference in their ability.

The current study used a subgroup of human-like characters from the game Mario Tennis Aces. The next step would be to explore the possibility of avatars that are based on a user's physical characteristics.

Future research could also investigate the effect of allowing users to customise their avatars themselves and if this would lead to different levels of engagement and physical activity.

6 Conclusion

In summary, the main focus of this research was to investigate the impact of a group recommender system for avatars in exergames on participants' physical activity, with an additional focus on the impact of the perceived experience, fairness, and difference in ability. The results provided important insights into these distinctive aspects.

This study first focused on the influence of specific circumstances on the recorded data, such as the order in which the tests were performed or the duration of the test. No significant influence was found, which suggests that either beginning or ending with the recommended avatar and the duration did not influence the participants' heart rate.

The primary focus of this study was to determine whether the recommender system influenced participants' heart rate. The measurements and results showed a significant increase in the participants' heart rate when they used the recommended avatar. The most important conclusion for this research can be drawn from this, a group recommender system for avatars increases heart rates, and therefore the physical activity of users.

Furthermore, this study showed that the increase in heart rate did not lead to a decrease in participants' perceived experience, fairness, or difference in ability. This suggests that the participants kept the same level or even slightly improved their perceived experience, fairness, or difference in ability in the recommended state.

Even though the research revealed some good insights into group recommender systems for avatars in exergames, it was also subject to some limitations. The primary constraints originated from a (relatively) small and skewed participant group, potential inaccuracies in the used heart rate monitors, a limit in base values for a couple of variables and the specific game choice.

Future work should prioritize a more extensive and diverse participant pool, with a better gender representation, but still consider that the most representative participants will be in friend groups. On top of that, the use of medical-grade devices should be considered. A pre-test should be used to get some base values to get a better understanding of the perceived experience, fairness, and difference in ability. The last focus for future research should be on the game and the characters themselves, as this study solely focuses on the Mario Tennis Aces game and its respective characters. An extra focus could be on the characters themselves by offering a more customizable product for the user, as a way to increase identification with the character.

All these possible options for future work aim to decrease the effect of the limitations in this research and help create a better understanding of the group recommender systems' impact on exergame experiences.

References

Abt, C. (1970). Serious games. *The Viking Press.*

Adams-Prassl, A., Boneva, T., Golin, M., & Rauh, C. (2020). The impact of the coronavirus lockdown on mental health: Evidence from the us. *Economic Policy.*

Alfano, V., & Ercolano, S. (2020). The efficacy of lockdown against covid-19: A cross-country panel analysis. *Applied health economics and health policy*, *18*(4), 509–517.

Anderson, C. (2006). *The long tail: Why the future of business is selling less of more*. Hachette Books.

Arboleda-Serna, V. H., Feito, Y., Patiño-Villada, F. A., Vargas-Romero, A. V., & Arango-Vélez, E. F. (2019). Effects of high-intensity interval training compared to moderate-intensity continuous training on maximal oxygen consumption and blood pressure in healthy men: A randomized controlled trial. *Biomedica*, *39*(3), 524–536.

Asahi, K., Undurraga, E. A., Valdés, R., & Wagner, R. (2021). The effect of covid-19 on the economy: Evidence from an early adopter of localized lockdowns. *Journal of global health*, *11*.

Banks, J., & Xu, X. (2020). The mental health effects of the first two months of lockdown during the covid-19 pandemic in the uk. *Fiscal Studies*, *41*(3), 685–708.

Berkovsky, S., Bhandari, D., Kimani, S., Colineau, N., & Paris, C. (2009). Designing games to motivate physical activity. *Proceedings of the 4th international conference on persuasive technology*, 1–4.

Berkovsky, S., Coombe, M., Freyne, J., Bhandari, D., & Baghaei, N. (2010). Physical activity motivating games: Virtual rewards for real activity. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 243–252.

Bhasin, K. (2012). *Why mario is the most popular video game character ever* [Visited on 14-04-2022]. Business Insider. https://www.businessinsider.com/why-mario-is-the-most-popular-video-game-character-ever-2012-3?international=true&r=US&IR=T

Birk, M. V., Atkins, C., Bowey, J. T., & Mandryk, R. L. (2016). Fostering intrinsic motivation through avatar identification in digital games. *Proceedings of the 2016 CHI conference on human factors in computing systems*, 2982–2995.

Birk, M. V., & Mandryk, R. L. (2019). Improving the efficacy of cognitive training for digital mental health interventions through avatar customization: Crowdsourced quasi-experimental study. *Journal of medical Internet research*, *21*(1), e10133.

Blackwood, K. (2006). *Casino gambling for dummies*. John Wiley & Sons.

Buddies, S. (2014). *Cardiovascular system science: Investigate heart-rate recovery time*. Scientific American. https://www.scientificamerican.com/article/cardiovascular-system-science-investigate-heart-rate-recovery-time1/

Caballero, B. (2004). Obesity prevention in children: Opportunities and challenges. *International journal of obesity*, *28*(3), S90–S95.

Chiu, C.-M., Lin, H.-Y., Sun, S.-Y., & Hsu, M.-H. (2009). Understanding customers' loyalty intentions towards online shopping: An integration of technology acceptance model and fairness theory. *Behaviour & Information Technology*, *28*(4), 347–360.

Clay, G., Mlynski, C., Korb, F. M., Goschke, T., & Job, V. (2022). Rewarding cognitive effort increases the intrinsic value of mental labor. *Proceedings of the National Academy of Sciences*, *119*(5), e2111785119.

Craig, C. L., Marshall, A. L., Sjöström, M., Bauman, A. E., Booth, M. L., Ainsworth, B. E., Pratt, M., Ekelund, U., Yngve, A., Sallis, J. F., et al. (2003). International physical activity questionnaire: 12-country reliability and validity. *Medicine & science in sports & exercise*, *35*(8), 1381–1395.

Dechant, M. J., Birk, M. V., Shiban, Y., Schnell, K., & Mandryk, R. L. (2021). How avatar customization affects fear in a game-based digital exposure task for social anxiety. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CHI PLAY), 1–27.

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

Douris, P. C., McDonald, B., Vespi, F., Kelley, N. C., & Herman, L. (2012). Comparison between nintendo wii fit aerobics and traditional aerobic exercise in sedentary young adults. *The Journal of Strength & Conditioning Research*, *26*(4), 1052–1057.

Elgar, F., Roberts, C., Moore, L., & Tudor-Smith, C. (2005). Sedentary behaviour, physical activity and weight problems in adolescents in wales. *Public health*, *119*(6), 518–524.

Ellenberg, J. (2008). This psychologist might outsmart the math brains competing for the netflix prize. *Wired Magazine*.

Elmer, T., Mepham, K., & Stadtfeld, C. (2020). Students under lockdown: Comparisons of students' social networks and mental health before and during the covid-19 crisis in switzerland. *Plos one*, *15*(7), e0236337.

Fink, J. (1999). *Cyberseduction: Reality in the age of psychotechnology* (Vol. 1). Prometheus Books.

Fletcher, G. F., Balady, G. J., Amsterdam, E. A., Chaitman, B., Eckel, R., Fleg, J., Froelicher, V. F., Leon, A. S., Piña, I. L., Rodney, R., et al. (2001). Exercise standards for testing and training: A statement for healthcare professionals from the american heart association. *Circulation*, *104*(14), 1694–1740.

Foley, L., & Maddison, R. (2010). Use of active video games to increase physical activity in children: A (virtual) reality? *Pediatric exercise science*, *22*(1), 7–20.

Fujiki, Y., Kazakos, K., Puri, C., Buddharaju, P., Pavlidis, I., & Levine, J. (2008). Neat-o-games: Blending physical activity and fun in the daily routine. *Computers in Entertainment (CIE)*, *6*(2), 1–22.

GameXplain. (2018). *Mario tennis aces - swing mode explained w/ demonstration*. Youtube. https://www.youtube.com/watch?v=uGX6xycQRlM

Georgiadou, E., Koopmann, A., Müller, A., Leménager, T., Hillemacher, T., & Kiefer, F. (2021). Who was shopping more during the spring lockdown 2020 in germany? *Frontiers in Psychiatry*, *12*, 215.

Hamilton, M. T., Healy, G. N., Dunstan, D. W., Zderic, T. W., & Owen, N. (2008). Too little exercise and too much sitting: Inactivity physiology and the need for new recommendations on sedentary behavior. *Current cardiovascular risk reports*, *2*(4), 292–298.

Hanani, U., Shapira, B., & Shoval, P. (2001). Information filtering: Overview of issues, research and systems. *User modeling and user-adapted interaction*, *11*(3), 203–259.

Hills, A. P., Okely, A. D., & Baur, L. A. (2010). Addressing childhood obesity through increased physical activity. *Nature Reviews Endocrinology*, *6*(10), 543–549.

Huang, Z., Chen, H., & Zeng, D. (2004). Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering. *ACM Transactions on Information Systems (TOIS)*, *22*(1), 116–142.

International Consensus Group. (1998). *International physical activity questionnaire long last 7 days self-administered format* [Visited on 01-04-2022]. https://www.researchgate.net/profile/Malgorzata-Wasilewska/post/How_to_measurement_physical_activity_and_food_frequency_questionnaire_among_children_and_adolescents/attachment/59d642a779197b807799e501/AS%5C%3A439636968382467%5C%401481829029197/download/IPAQ_English_self-admin_long.pdf

Jaric, S. (2002). Muscle strength testing. *Sports Medicine*, *32*(10), 615–631.

Jin, C.-H., Liu, J.-G., Zhang, Y.-C., & Zhou, T. (2009). Adaptive information filtering for dynamic recommender systems. *arXiv preprint:0911.4910*.

Joshi, A., Bhaskar, P., & Gupta, P. K. (2020). Indian economy amid covid-19 lockdown: A prespective. *J. Pure Appl. Microbiol*, *14*, 957–961.

Kaos, M. D., Beauchamp, M. R., Bursick, S., Latimer-Cheung, A. E., Hernandez, H., Warburton, D. E., Yao, C., Ye, Z., Graham, T. N., & Rhodes, R. E. (2018). Efficacy of online multi-player versus single-player exergames on adherence behaviors among children: A nonrandomized control trial. *Annals of Behavioral Medicine*, *52*(10), 878–889.

Kaye, L. K., & Bryce, J. (2012). Putting the fun factor into gaming: The influence of social contexts on the experiences of playing videogames. *International Journal of Internet Science*, *7*(1), 24–38.

Kilpatrick, M., Hebert, E., & Bartholomew, J. (2005). College students' motivation for physical activity: Differentiating men's and women's motives for sport participation and exercise. *Journal of American college health*, *54*(2), 87–94.

Kim, J. (2010). Balancing uniqueness and assimilation in computer-mediated groups. *Computers in Human Behavior*, *26*(4), 778–784.

Kim, J. (2011). Two routes leading to conformity intention in computer-mediated groups: Matching versus mismatching virtual representations. *Journal of Computer-Mediated Communication*, *16*(2), 271–287.

Kim, J., & Park, H. S. (2011). The effect of uniform virtual appearance on conformity intention: Social identity model of deindividuation effects and optimal distinctiveness theory. *Computers in Human Behavior*, *27*(3), 1223–1230.

King, D. L., Delfabbro, P. H., Billieux, J., & Potenza, M. N. (2020). Problematic online gaming and the covid-19 pandemic. *Journal of Behavioral Addictions*, *9*(2), 184–186.

Knijnenburg, B. P., Willemsen, M. C., Gantner, Z., Soncu, H., & Newell, C. (2012). Explaining the user experience of recommender systems. *User modeling and user-adapted interaction*, *22*(4), 441–504.

Kok, P., Seidell, J., & Meinders, A. (2004). The value and limitations of the body mass index (bmi) in the assessment of the health risks of overweight and obesity. *Nederlands tijdschrift voor geneeskunde*, *148*(48), 2379–2382.

Konietzny, J., & Caruana, A. (2019). Fair and easy: The effect of perceived fairness, effort expectancy and user experience on online slot machine gambling intention. *International Gambling Studies*, *19*(2), 183–199.

Kooiman, B. J., & Sheehan, D. P. (2015). The efficacy of exergames for social relatedness in online physical education. *Cogent Education*, *2*(1), 1045808.

Lee, P. H., Macfarlane, D. J., Lam, T. H., & Stewart, S. M. (2011). Validity of the international physical activity questionnaire short form (ipaq-sf): A systematic review. *International journal of behavioral nutrition and physical activity*, *8*(1), 1–11.

Lee, S., Kim, W., Park, T., & Peng, W. (2017). The psychological effects of playing exergames: A systematic review. *Cyberpsychology, Behavior, and Social Networking*, *20*(9), 513–532.

Lessig, L. (2009). *Code: And other laws of cyberspace*. ReadHowYouWant. com.

Lü, L., Medo, M., Yeung, C. H., Zhang, Y.-C., Zhang, Z.-K., & Zhou, T. (2012). Recommender systems. *Physics reports*, *519*(1), 1–49.

Lucas, G. M., Szablowski, E., Gratch, J., Feng, A., Huang, T., Boberg, J., & Shapiro, A. (2016). Do avatars that look like their users improve performance in a simulation? *International Conference on Intelligent Virtual Agents*, 351–354.

Lyons, E. J. (2015). Cultivating engagement and enjoyment in exergames using feedback, challenge, and rewards. *Games for health journal*, *4*(1), 12–18.

Masthoff, J. (2011). Group recommender systems: Combining individual models. *Recommender systems handbook* (pp. 677–702). Springer.

Masthoff, J. (2015). Group recommender systems: Aggregation, satisfaction and group attributes. *Recommender systems handbook* (pp. 743–776). Springer.

McWhertor, M. (2019). *God of war wins big at dice awards 2019* [Visited on 20-04-2022]. Polygon. https://www.polygon.com/2019/2/14/18224190/dice-awards-2019-winners-god-of-war-ps4

Michaud, L., Alvarez, J., Alvarez, V., & Djaouti, D. (2010). *Serious games: Training & teaching - healthcare - defence & security - information & communication (2e ed.)*

Miller, K. J., Adair, B. S., Pearce, A. J., Said, C. M., Ozanne, E., & Morris, M. M. (2014). Effectiveness and feasibility of virtual reality and gaming system use at home by older adults for enabling physical activity to improve health-related domains: A systematic review. *Age and ageing*, *43*(2), 188–195.

Morelli, M., Cattelino, E., Baiocco, R., Trumello, C., Babore, A., Candelori, C., & Chirumbolo, A. (2020). Parents and children during the covid-19 lockdown: The influence of parenting distress and parenting self-efficacy on children's emotional well-being. *Frontiers in Psychology*, *11*, 2584.

Muntean, C. I. (2011). Raising engagement in e-learning through gamification. *Proc. 6th international conference on virtual learning ICVL*, *1*, 323–329.

Must, A., & Tybor, D. (2005). Physical activity and sedentary behavior: A review of longitudinal studies of weight and adiposity in youth. *International journal of obesity*, *29*(2), S84–S96.

Nafees, M., & Khan, F. (2020). Pakistan's response to covid-19 pandemic and efficacy of quarantine and partial lockdown: A review. electron j gen med. 2020; 17 (2): Emxxx.

Nguyen, T. T., Nguyen, M. H., Pham, T. T., Le, V.-T. T., Nguyen, T. T., Luong, T. C., Do, B. N., Dao, H. K., Nguyen, H. C., Ha, T. H., et al. (2021). Negative impacts of covid-19 induced lockdown on changes in eating behavior, physical activity, and mental health as modified by digital healthy diet literacy and ehealth literacy. *Frontiers in nutrition*, *8*.

Nintendo. (2022). *Mario tennis aces* [Visited on 16-03-2022]. https://www.nintendo.nl/Games/Nintendo-Switch/Mario-Tennis-Aces-1325791.html

Nintendo UK. (2018). *Mario tennis aces - launch trailer (nintendo switch)*. Youtube. https://www.youtube.com/watch?v=jxrHwK-e1vk

Norton, K., Norton, L., & Sadgrove, D. (2010). Position statement on physical activity and exercise intensity terminology. *Journal of science and medicine in sport*, *13*(5), 496–502.

Pathak, G., & Warpade, S. (2020). Impact of lockdown due to covid-19 on consumer behaviour while selecting retailer for essential goods. *Pathak, GP, & Warpade, S.(2020, July 31). Impact of Lockdown due to COVID*, *19*, 282–289.

Peña, J., Khan, S., & Alexopoulos, C. (2016). I am what i see: How avatar and opponent agent body size affects physical activity among men playing exergames. *Journal of Computer-Mediated Communication*, *21*(3), 195–209.

Peña, J., & Kim, E. (2014). Increasing exergame physical activity through self and opponent avatar appearance. *Computers in Human Behavior*, *41*, 262–267.

Resnick, P., & Varian, H. R. (1997). Recommender systems. *Communications of the ACM*, *40*(3), 56–58.

Ricci, F., Rokach, L., Shapira, B., & Kantor, P. B. (2011). *Recommender systems handbook*. Springer.

Sale, D., & Norman, R. (1991). Testing strength and power. *Physiological testing of the high-performance athlete*, 21–106.

Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2002). Incremental singular value decomposition algorithms for highly scalable recommender systems. *Fifth international conference on computer and information science*, *1*(012002), 27–8.

Schafer, J. B., Konstan, J. A., & Riedl, J. (2001). E-commerce recommendation applications. *Data mining and knowledge discovery*, *5*(1), 115–153.

Smallwood, J., Davies, J. B., Heim, D., Finnigan, F., Sudberry, M., O'Connor, R., & Obonsawin, M. (2004). Subjective experience and the attentional lapse: Task engagement and disengagement during sustained attention. *Consciousness and cognition*, *13*(4), 657–690.

Staiano, A. E., & Calvert, S. L. (2011). Exergames for physical education courses: Physical, social, and cognitive benefits. *Child development perspectives*, *5*(2), 93–98.

Statton, M. A., Encarnacion, M., Celnik, P., & Bastian, A. J. (2015). A single bout of moderate aerobic exercise improves motor skill acquisition. *PloS one*, *10*(10), e0141393.

Steinberger, F., Schroeter, R., & Watling, C. N. (2017). From road distraction to safe driving: Evaluating the effects of boredom and gamification on driving behaviour, physiological arousal, and subjective experience. *Computers in Human Behavior*, *75*, 714–726.

Susi, T., Johannesson, M., & Backlund, P. (2007). Serious games: An overview.

Thakur, K., Kumar, N., & Sharma, N. (2020). Effect of the pandemic and lockdown on mental health of children. *The Indian Journal of Pediatrics*, *87*(7), 552–552.

Tjønna, A. E., Leinan, I. M., Bartnes, A. T., Jenssen, B. M., Gibala, M. J., Winett, R. A., & Wisløff, U. (2013). Low-and high-volume of intensive endurance training significantly improves maximal oxygen uptake after 10-weeks of training in healthy men. *PloS one*, *8*(5), e65382.

University of Würzburg. (2019). *Do video games drive obesity?* [Visited on 08-01-2022]. www.sciencedaily.com/releases/2019/06/190617110522.html

Van Eck, R. (2006). Digital game-based learning: It's not just the digital natives who are restless. *EDUCAUSE review*, *41*(2), 16.

WatchData. (2020). *Comparison: Most famous characters* [Visited on 14-04-2022]. Youtube. https://www.youtube.com/watch?v=ZNwRK20C7JM&t=140s

Wieringa, R. J. (2014). *Design science methodology for information systems and software engineering*. Springer.

Wilmot, E. G., Edwardson, C. L., Achana, F. A., Davies, M. J., Gorely, T., Gray, L. J., Khunti, K., Yates, T., & Biddle, S. J. (2012). Sedentary time in adults and the association with diabetes, cardiovascular disease and death: Systematic review and meta-analysis. *Diabetologia*, *55*(11), 2895–2905.

World Health Organization et al. (2005). The surf report 2: Surveillance of chronic disease risk factors. *Geneva: WhO*.

Yadav, M., Goel, M., & Sharma, B. (2020). Consumer perception towards online shopping during lockdown. *International Journal of Management*, *11*(07), 1578–1584.

Zagal, J. P., Rick, J., & Hsi, I. (2006). Collaborative games: Lessons learned from board games. *Simulation & gaming*, *37*(1), 24–40.

Zagal, J. P., Nussbaum, M., & Rosas, R. (2000). A model to support the design of multiplayer games. *Presence: Teleoperators & Virtual Environments*, *9*(5), 448–462.

Zamboni, L., Carli, S., Belleri, M., Giordano, R., Saretta, G., & Lugoboni, F. (2021). Covid-19 lockdown: Impact on online gambling, online shopping, web navigation and online pornography. *Journal of public health research*, *10*(1).

Zhang, Y., Liang, Q., & Wang, F. (2023). Influence of competition-outcome feedback in video games on players' flow experience. *Current Psychology*, *42*(21), 17583–17594.

7 Appendices

Appendix A: Questionnaire 1

Demographic questions

1. What is your age?
2. What is your gender?
3. Identification number

IPAQ Questionnaire

Part 1: Job-Related Physical Activity

1. Do you currently have a job or do any unpaid work outside your home?

2. During the last 7 days, on how many days did you do vigorous physical activities like heavy lifting, digging, heavy construction, or climbing up stairs as part of your work? Think about only those physical activities that you did for at least 10 minutes at a time.

3. How much time did you usually spend on one of those days doing vigorous physical activities as part of your work?

4. Again, think about only those physical activities that you did for at least 10 minutes at a time. During the last 7 days, on how many days did you do moderate physical activities like carrying light loads as part of your work? Please do not include walking.

5. How much time did you usually spend on one of those days doing moderate physical activities as part of your work?

6. During the last 7 days, on how many days did you walk for at least 10 minutes at a time as part of your work? Please do not count any walking you did to travel to or from work.

7. How much time did you usually spend on one of those days walking as part of your work?

Part 2: Transportation physical activity

8. During the last 7 days, on how many days did you travel in a motor vehicle like a train, bus, car, or tram?

9. How much time did you usually spend on one of those days travelling in a train, bus, car, tram, or other kind of motor vehicle?

10. During the last 7 days, on how many days did you bicycle for at least 10 minutes at a time to go from place to place?

11. How much time did you usually spend on one of those days to bicycle from place to place?

12. During the last 7 days, on how many days did you walk for at least 10 minutes at a time to go from place to place?

13. How much time did you usually spend on one of those days walking from place to place?

Part 3: Housework, House Maintenance, and Caring for Family

14. Think about only those physical activities that you did for at least

10 minutes at a time. During the last 7 days, on how many days did you do vigorous physical activities like heavy lifting, chopping wood, shovelling snow, or digging in the garden or yard?

15. How much time did you usually spend on one of those days doing vigorous physical activities in the garden or yard?

16. Again, think about only those physical activities that you did for at least 10 minutes at a time. During the last 7 days, on how many days did you do moderate activities like carrying light loads, sweeping, washing windows, and raking in the garden or yard?

17. How much time did you usually spend on one of those days doing moderate physical activities in the garden or yard?

18. Once again, think about only those physical activities that you did for at least 10 minutes at a time. During the last 7 days, on how many days did you do moderate activities like carrying light loads, washing windows, scrubbing floors and sweeping inside your home?

19. How much time did you usually spend on one of those days doing moderate physical activities inside your home?

Part 4: Recreation, Sport, and Leisure-Time Physical Activity

20. Not counting any walking you have already mentioned, during the last 7 days, on how many days did you walk for at least 10 minutes at a time in your leisure time?

21. How much time did you usually spend on one of those days walking in your leisure time?

22. Think about only those physical activities that you did for at least 10 minutes at a time. During the last 7 days, on how many days did you do vigorous physical activities like aerobics, running, fast bicycling, or fast swimming in your leisure time?

23. How much time did you usually spend on one of those days doing vigorous physical activities in your leisure time?

24. Again, think about only those physical activities that you did for at least 10 minutes at a time. During the last 7 days, on how many days did you do moderate physical activities like bicycling at a regular pace, swimming at a regular pace, and doubles tennis in your leisure time?

25. How much time did you usually spend on one of those days doing moderate physical activities in your leisure time?

Part 5: Time Spent Sitting

26. During the last 7 days, how much time did you usually spend sitting on a weekday?

27. During the last 7 days, how much time did you usually spend sitting on a weekend day?

Avatar Preference

1. Please order the characters based on your preference.

Appendix B: Questionnaire 2

Experience Questionnaire

1. The recommended avatar was relevant.

2. I liked the recommendation provided by the system.

3. The recommended item fitted my preference

4. The recommender system is providing good recommendations

5. I didn't like any of the recommended avatars.

6. The recommender system is not predicting my ratings accurately.

7. The recommendation did not include my favourite avatars.

Social Fainess Questionnaire

8. I think the avatar I got recommended is fair compared to the answers I gave in the previous questionnaire.

9. I think the avatar that I got recommended is considered to be a good avatar.

10. I think the procedures used by the system for recommending the avatar in the recommendation process are fair.

11. I think the policies used by the system for recommending the avatar in the recommendation process are applied consistently across all users.

12. I think the recommender system allows the users to state their preferences.

Difference in Ability Questionnaire

13. I feel that there was a difference in the ability of the different avatars during the game.

14. The different abilities of the avatars influenced my physical activity in a positive way (my physical activity increased).

15. The different abilities of the avatars influenced my experience in a positive way.

Appendix C: Pseudocode

Group Recommender System

The system follows the following steps:

1. The participants fill in Questionnaire 1 and rank the 8 possible characters based on their personal preference.

2. The questionnaire data is downloaded in .csv format.

3. The data is loaded into the Python file as a Pandas DataFrame and stripped of redundant data. 3. All columns are renamed to have more relevant titles.

4. The total activity for each kind of activity per week is computed. Example: if a participant filled in that they have vigorous activity during their work for 3 days a week, and that takes half an hour, then the *Work_Vigorous* gets computed by multiplying the 30 minutes by 3, so the final computed result will be 90 (minutes).

5. The different activity intensities are grouped to get a total number of minutes per week for each activity intensity, which results in 4 data points:
- Total number of **vigorous** activity in a week.
- Total number of **moderate** activity in a week.
- Total number of **walking** activity in a week.
- Total number of **sitting** activity in a week.

6. The participants are ordered for each activity intensity and the positions are logged.
Example: If the participants have the following total moderate activities

during the week, the following positions are logged:
- Participant 1: 100 minutes -> Place 3
- Participant 2: 90 minutes -> Place 4
- Participant 3: 135 minutes -> Place 2
- Participant 4: 630 minutes -> Place 1

7. These positions are used in the recommender system, this is the way to compare each participant with other participants in the group.

8. The recommender system uses 5 classes out of the Mario Tennis Aces game; All-Around, Powerful, Defensive, Speedy and Technical. The recommender function (*recommender_system*) takes a participant at random and checks if their positions fall within the restrictions of a specific class. The restrictions of each class are based on their respective strengths and weaknesses.
Example: if somebody has a high place in *Vigorous_Activity* and *Sitting*, but a lower place in *Moderate_Activity*, they are likely to be recommended a powerful character. See Figure 15 for a specification for each class.

All-Around:	Strenghts:	None
	Weaknesses:	None
Powerful:	Strenghts:	Power -> Vigorous high
	Weaknesses:	Speed -> Sitting high and/or Moderate low
Defensive:	Strenghts:	Reach -> Walk high
	Weaknesses:	Power & Speed -> Vigorous low and/or Sitting high and/or Moderate low
Speedy:	Strenghts:	Speed -> Sitting low and/or Moderate high
	Weaknesses:	Power -> Vigorous low
Technical:	Strenghts:	Accuracy -> Walk high
	Weaknesses:	Power -> Vigorous low

Figure 15: Specification of each class

9. If the system does not find a matching class for the participant's data, the All-Around class gets recommended and the participant is put into the function again, this time with more lenient restrictions. The iteration is logged to make sure that the strongest recommendations (the recommendations with a low amount of iterations) are found. An

example of the resulting DataFrame is shown in Figure 16.

recommender_1	recommender_2	recommender_3
All-Around	Defensive	Powerful,Defensive,Technical
All-Around	All-Around	Powerful,Defensive
All-Around	Speedy	Speedy
All-Around	Powerful,Defensive	Powerful,Defensive

Figure 16: DataFrame after *recommender_system* function

10. The creation of the system to make the choice of which class eventually gets recommended to the participants was started, as only one of each class can be chosen in this scenario, but the choice was made to do this by hand. This decision was made due to personal circumstances preventing me from writing the code. The choices made can best be described with an example based on Figure 16:
Start by looking at the first iteration (all participants are recommended the All-Around class, this is not very remarkable since the restriction to get recommended a specific class is very high in the first iteration of the system. When looking at the second iteration, it is clear that three out of the four participants can be recommended a specific class, so those are logged:
- 1 -> Defensive
- 2 -> Nothing (yet)
- 3 -> Speedy
- 4 -> Powerful (as Defensive is already linked to the first participant)

Then the third iteration is looked at, if the participant gets recommended a class that is still available then this class is chosen, this is not the case, and since the Powerful and Defensive recommendations are stronger in the second iteration for the other participants, the second participant will be recommended a character from the All-Around class.

11. The selected class for each participant is inputted into the DataFrame.

12. Then, for each participant, their class is checked. If their recommended class has two possible characters, the system checks which of those characters got the highest preference from that specific participant and is selected. If the recommended class has one possible character, that character is selected. 13. The system prints the participants' ID and the name of the recommended character, see Figure 17.

```
1: Waluigi
2: Luigi
3: Yoshi
4: Wario
```

Figure 17: Group Recommender System output