What makes people accept recommendations?

A comparison of different recommender types on music streaming platforms

Student Name: Ina Helene Weber
Student Number: 481707iw

Supervisor: Dr. Joao Ferreira Goncalves

Master Media Studies - Media, Culture & Society
Erasmus School of History, Culture and Communication
Erasmus University Rotterdam

Master's Thesis
June 2019

What makes people accept recommendations?

ABSTRACT

This research aims to contribute to the development of a better understanding of how users of online services react to recommendations. To do so, it examines how the acceptance of a recommendation on a music streaming platform differs depending on the source of the recommendation, thereby drawing a comparison between algorithms, experts and peers as recommenders. Data for this research was gathered by conducting a between-subjects online survey experiment with three experimental conditions according to the three recommenders under study. The research used a vignette and a fabricated music recommendation to elicit the reactions of users to recommendations on streaming platforms. Based on the concept of algorithm appreciation by Logg, Minson and Moore (2019), it was hypothesised that algorithms will be perceived more positively as recommenders than peers, but less positive than experts. Changes in the outcome variables attitudes and intended behaviours were expected to be caused by the type of recommender, and – in consideration of the concept of source credibility – also by the degree to which participants perceived the recommender as trustworthy and knowledgeable. This relation in turn was expected to be influenced by eWOM scepticism. To analyse the considered effects, the research used moderated mediation. As the results show, both recommender type and source credibility had an impact on the acceptance of a recommendation and algorithm appreciation was confirmed as the algorithm recommender was perceived significantly better than the expert and the peer recommender. However, the effects under study occurred independently of each other, thus the level of source credibility attributed to a recommender does not serve as an explanation for why the recommender type affected the acceptance of a recommendation. eWOM scepticism was not confirmed as a moderator but did have a negative impact on source credibility. Perceived personalisation is discussed as a potential alternative explanation and it is suggested to future research to further explore the conceptual connections of perceived personalisation and algorithm appreciation. This research provides insights into how practices of music discovery are evolving on streaming platforms under the prevalent influence of technologies such as machine learning. In a broader sense, this thesis is a contribution to the larger theoretical framework of understanding the implications of artificial intelligence for society and culture.

<u>KEYWORDS:</u> *algorithms, music streaming, recommendations, source credibility, eWOM*

ACKNOWLEDGEMENT

Writing this thesis was a substantial amount of work, which would probably have been never ending if it weren't for a number of very important people who have accompanied me through this process and contributed to this project.

First and foremost, I would like to thank my supervisor Joao for all the support, guidance and feedback, for answers to questions that seemed unanswerable and most importantly for all the motivation and time invested in this piece of work. A big thank you goes to my highly skilled team of designers Camila, Lubna, Nada and Shannon who should really consider making album covers for real bands because their work was amazing. I furthermore would like to thank everybody who participated in several rounds of pre-testing and provided feedback that was truly important for my experiment and highly appreciated, as well as the anonymous music lovers from all around the world who spared the time to take part in my research.

Clearly, this thesis would not be finished without the support of the regular cast at Polak. Katja, Andrea, Dilek: many thanks for enduring endless days and nights of study sessions together and for bearing with my aggressive typing. Finally, I want to thank my mother for making it possible for me to pursue this master and for supporting me throughout all stages of it.

And for authenticity, obviously this thesis was written while listening to countless hours of algorithmically selected, expert curated and peer recommended music.

Table of contents

Abstract and keywords

Acknowledgement

1 Introduction 1

2 Theoretical Framework 4

2.1 On the context of music streaming 4

2.1.1 The music industry and its relation to technology.. 4

2.1.2 Perspectives on music.. 5

2.1.3 Implications for music listeners 6

2.2 Comparing recommender types 7

2.2.1 Algorithm appreciation 8

2.2.2 Algorithm aversion.. 9

2.2.3 Expertise as intervening factor... 10

2.3 Perception of online messages 12

2.3.1 Source credibility 12

2.3.2 Attitudes and intended behaviours.. 13

2.4 eWOM scepticism 15

2.5 Summary 17

3 Method 19

3.1 Choice of Method... 19

3.2 Sampling.. 21

3.2.1 Sampling Method.. 21

3.2.2 Description of sample 23

3.2.3 Random Assignment 24

3.3 Operationalisation 25

3.3.1 Independent variable recommender type.. 25

3.3.2 Dependent variable attitude... 26

3.3.3 Dependent variable intended behaviour 27

3.3.4 Mediating variable source credibility 28

3.3.5 Moderating variable eWOM scepticism... 29

3.3.6 Additional variables 29

3.4 Pre-test 31

3.4.1 Pre-test of potential stimuli 32

3.4.2 Pre-test survey experiment 34

3.5 Research design and data collection 35

3.6 Validity and reliability 38

3.6.1 Validity 38

3.6.2 Reliability 39

3.7 Method of data analysis 42

4 Results 44

4.1 Direct effects 44

4.2 Indirect and conditional effects 46

4.3 The impact of expertise and age 48

5 Conclusion 50

5.1 Discussion 50

5.1.1 Summary of results 50

5.1.2 Perceived personalisation as possible explanation 52

5.2 Conclusion and implications 54

5.3 Limitations 56

5.4 Future research 57

References 59

Appendix A 65

Appendix B 70

Appendix C 72

Appendix D 89

List of figures

Figure 2.1 Research model and overview over hypotheses.. 18

Figure 4.1 Moderated mediation of accepting a recommendation 47

Figure A1. Album Cover *Between Planets* by fictional artist *Blair* 65

Figure A2. Album Cover *Dogma Beware* by fictional artist *Boots & Laces* 65

Figure A3. Album Cover *Cherry Vagabond* by fictional artist *Curb Appeal* 66

Figure A4. Album Cover *Ab*sent Shores by fictional artist *Dream of Abyss* 66

Figure A5. Album Cover *Decisions, Decisions* by fictional artist *Inescapable Immortality* .. 67

Figure A6. Album Cover *Hell & Hope* by fictional artist *Pawn Panic* 67

Figure A7. Album Cover *Night in the Ship* by fictional artist *Place Stamp Here* 68

Figure A8. Album Cover *Sp*ace by fictional artist *The Sockets* 68

Figure A9. Album Cover Academy & the Child by fictional artist Tipsy Jones 69

Figure A10. Album Cover *Blissful Ignorance* by fictional artist *Victor & the Eight* 69

List of tables

Table 3.1. Factor and reliability analysis for source credibility.. 41
Table 3.2. Factor and reliability analysis for eWOM scepticism 42
Appendix B1. Outcome for *Dogma Beware* by fictional artist *Boots & Laces* 70
Appendix B2. Outcome for *Space* by fictional artist *The Sockets* 70
Appendix B3. Outcome for *Blissful Ignorance* by fictional artist *Victor & the Eigh*t 71
Appendix D1. Relative conditional indirect effect *recommender type* on *attitude* 89
Appendix D2. Relative conditional indirect effect r*ecommender type* on *intended behavior* 89

1 Introduction

In today's highly digitised world, one does not necessarily have to be a computer scientist to come across artificial intelligence or algorithms, as these are recurring terms in the popular discourse and have become deeply integrated into everyday life: many people rely on adaptive algorithms to retrieve information by using Google, to find a movie to watch on Netflix or to find the most convenient way to a place by asking Siri. Beyond that, the implementation of artificial intelligence is not limited to these kind of applications, but is also becoming a significant aspect in many parts of the public sector such as hospitals or courts (Crawford & Calo, 2016).

The increasing prevalence of artificial intelligence has made it a topic of interest for researchers beyond the field of computer sciences as it is increasingly acknowledged that "[a]rtificial intelligence presents a cultural shift as much as a technical one" (Crawford & Calo, 2016, p. 313). This has led to different disciplines within the social sciences being called upon to critically investigate the social and cultural impact of artificial intelligence. An understanding of this impact can only be achieved by conducting research with regards to various contexts and the nuanced ways in which artificial intelligence plays a role in each of them (Thurman, Moeller, Helberger, & Trilling, 2018). By building upon the findings of these distinct studies and combining them, it will subsequently be possible to obtain a more holistic picture of the relation between artificial intelligence and society (Crawford & Calo, 2016).

However, what complicates the research on the social impact of artificial intelligence is the level of abstractness that is inherent to the topic. Although algorithms are integrated into many common applications, they are to a large extent 'black boxes', meaning that users most often know very little about their functionality. In addition to this lack of understanding, the algorithms themselves are often non-transparent and companies such as Google or Facebook are reluctant to reveal detailed information about them (Pasquale, 2015). The latter especially complicates the grasping of the actual impact that algorithms might have. Therefore, it can be more useful to approach the topic from the user's perspective. Bucher (2017) has summarised the rationale for this kind of research by stating:

> "If we want to understand the social power of algorithms, it is important to understand how users encounter and make sense of algorithms, and how these experiences, in turn, not only shape the expectations users have towards computational systems, but also help shape the algorithms themselves." (p. 33)

A particularly relevant context to study these kinds of encounters between people and algorithms are situations in which people rely on advice or recommendations provided by algorithms (Thurman et al., 2018). Gaining insights on whether, to what extent and under what conditions people tend to follow the suggestions from algorithmic advisors enhances the knowledge on how people perceive algorithms. These insights moreover allow researchers to draw conclusions on how algorithms may have an impact on the specific context in which the advice is given (Thurman et al., 2018). Despite the apparent possibilities that the suggested approach offers, Logg (2017) has identified a lack of research on this matter. Thus, this thesis aims to contribute to the development of a broader understanding of the social impact of artificial intelligence by researching the way in which algorithms are perceived and what reactions people display to recommendations provided by algorithms. To do so, it draws upon a specific case that shall be outlined in the following.

As mentioned previously, there is an immense number of applications and online services through which users encounter algorithms, encompassing almost every aspect of daily life. Naturally, entertainment makes up a vital part of that and nowadays, cultural products such as movies, series or music are commonly consumed through online platforms, or on-demand streaming services to be precise. Since a medium partly shapes the listening or viewing experience, the increased usage of streaming contributes to the emergence of new norms and practices of consumption of cultural goods, which are yet to be determined (Hagen, 2015). In that regard, not only the act of consuming is of significance, but also how the content is presented on these platforms and how users discover new movies or music.

In the face of today's online media environment, characterised by constant stimulation and a seemingly infinite offer of information and entertainment, it is comprehensible that many users prefer to rely on some form of guidance in order to allocate the scarce resources of time and attention to the content that is most relevant to them (O'Reilly, Larsen, & Kubacki, 2013). Online recommendations therefore play an important role, as they can be a "solution to the issue of information overload online" (Hagen, 2015, p. 628).

On streaming platforms, recommendations are a central feature not only for the users but also for the company behind each platform. The market for music streaming is highly dynamic, with both aspiring services expanding their offer as well as established tech companies launching streaming services as additions to their portfolio (Deahl, 2018). To stand out from competitors, streaming providers can therefore not just rely on a broad catalogue, but also offer functionalities such as recommendation systems (Morris & Powers, 2015). Furthermore, the more a platform provides its users with options to discover new songs and facilitates the process of finding music that matches different situations and activities, the higher is the chance to keep users subscribed (Seaver, 2018).

According to Senecal and Nantel (2004), there are generally three types of sources that provide online recommendations, namely peers, experts from the respective field, and recommender systems, i.e. algorithms. All of these types of recommenders can be found within the context of music streaming: some services like Apple Music for example collaborate with experts who curate themed or genre-specific playlists for the platform (Gibbs, 2017). By connecting with each other, creating public playlists and sharing music with others via social media, users have the chance to both make and receive recommendations. These options are complemented by the usage of machine learning, as in the case of Spotify, in order to evaluate and compare the user behaviour while also processing different information about the songs (Ciocca, 2017). Streaming providers put extensive amounts of work into the development of these elaborate recommender systems which are, as Seaver (2018) argues, not just a feature or an additional element to streaming, but are deeply embedded into the structures of the platforms and inseparable from the streaming service itself. This is not to imply that the practice of content curation is a new phenomenon only emerging with the growing popularity of streaming or that 'traditional' recommenders are being replaced by algorithmic ones (Morris, 2015) – au contraire, different recommenders are often used within one service. However, it is noticeable how technology in the form of machine learning has fairly recently become deeply involved in the process of curation (Morris, 2015).

Inevitably, this leads to questions about how the curation of cultural content is impacted by the emergence of machine learning and how that in turn impacts the users' interactions with cultural content online (Morris, 2015). By using music streaming platforms as a case/example to compare reactions to recommendations made by different types of recommenders, it is possible to critically examine this matter and to gain some insights into the curation of content by algorithmic intermediaries (Morris, 2015). As indicated earlier, this provides the potential to contribute to the broader field of research on the social implications of artificial intelligence by considering the perspective and the behaviour of users with this type of technology. In order to specifically fill the gap of experimental research on the reactions to algorithmic recommendations, as pointed out by Logg (2017), this thesis is conducted through an experiment in order to answer the following research question:

RQ: How do different recommender types affect the users' acceptance of music recommendations on streaming platforms?

2 Theoretical Framework

2.1 On the context of music streaming

Before shedding a light on the processes that potentially affect people in their acceptance of recommendations, it is important to sufficiently contextualise the environment in which these recommendations are being made. In other words, the following section shall provide some necessary information about music streaming and how the prevalence of the streaming technology has affected the way how music is distributed, conceptualised and consumed. Ultimately, the possibility to conduct this research and to draw comparisons between human and non-human recommender types is in itself a result of the emergence of music streaming and the development of related technologies. Thus, it is crucial to comprehend the larger changes in music consumption and the music market that can be observed throughout the last years.

2.1.1 The music industry and its relation to technology

When looking at developments in the music industry that have taken place over the recent decades, a strong connection between music and technology can be noticed. Advancement in technology is a significant driver for change of the production, distribution and consumption of music (Wikström, 2013). For example, the emergence of each new carrier medium over the past decades has highly impacted these aspects (O'Reilly et al., 2013), introducing challenges to the industry but also providing new opportunities of generating revenue (Arditi, 2017). As such, digitisation undoubtedly has caused one of the biggest disruptions the music industry yet had to face, leading to a major decline in physical album sales and consequently an immense decline of profits (O'Reilly et al., 2013). Wikström (2013) describes the resulting current situation as the 'new music economy', characterised by "high connectivity and little control, music provided as a service, and increased amateur creativity" (p. 86).

As a result of conglomeration and globalisation, the commercial music market today is for the most part dominated by the three biggest record labels Universal Music Group, Sony Music Entertainment and Warner Music Group (O'Reilly et al., 2013). In the past, these record labels were in a very powerful, almost monopolistic position which gave them control over production and distribution resources and also market prices. However, with the possibilities provided by digital technologies, emerging actors such as streaming providers are entering the field, leading to the development of a new digital value chain which expectedly will further change the market in the long term (Bockstedt, Kauffman, & Riggins, 2006). While services like Spotify were initially facing a lot of scepticism as to whether streaming will establish itself as a commonly used and thus profitable way of music consumption (Wikström, 2013), recent numbers prove sceptics to be wrong: according to the International Federation of the Phonographic Industry (IFPI), on-demand streaming has become the most popular way of consuming music, with 86% of consumers worldwide using streaming platforms (IFPI, 2018). In the US alone, streaming moreover accounted for 75% of the music industry's total revenue of 2018 (RIAA, 2018). Such significant changes have far reaching consequences, not only for the industry but also for its customers - the music listeners - and for the way how music itself is being thought of and understood. Both of these aspects shall be elaborated on in the following.

2.1.2 Perspectives on music

Music can be defined in different ways, depending on the perspective from which it is conceived because different viewpoints prompt particular ways of understanding and interacting with music. On the one hand, from a cultural point of view music can be approached as a sociocultural artefact, fulfilling different functions such as providing social experience on a collective level or being an expression of self-identity and mood on an individual level (O'Reilly et al., 2013). Furthermore, it is an inherently nonmaterial form of artistic and aesthetic expression (Wikström, 2013). On the other hand, music today is also commonly understood as a commodity – something that is marketed, monetized and consumed. The intangible and ephemeral artefact of music, when recorded onto a medium allowing it to be stored, reproduced, marketed and sold, can thus be thought of as a product (O'Reilly et al., 2013).

However, the accuracy of this notion is questionable because of a number of differences between classical consumer goods and cultural goods such as music (O'Reilly et al., 2013; Wikström, 2013). For example, music can be listened to an infinite number of times: the consumption of a song neither destroys nor alters it, which differs significantly from the consumption of many other consumer products (Arditi, 2017; O'Reilly et al., 2013). Moreover, the purchase of a good usually gives an individual the right of ownership over it. But by purchasing a song, one only gains the right to consume it, while the full ownership still remains with the copyright holder (Wikström, 2013). The consumption of music furthermore does not necessarily require purchase, as it can for example also take place by listening to the radio. In addition, it is difficult to define the value of music in comparison to defining the value of a regular consumer product. One has to consider not only the revenue that a song generates, but also its artistic and creative value, which do not always reflect each other (O'Reilly et al., 2013).

For a long time, music and its carrier medium could not be separated, so the music industry focused on the sale of physical goods and despite the aforementioned arguments, music was widely understood as a consumer product. Through digitisation however, music is mostly not purchased and consumed in a physical format anymore. Thus, the perception of music as an information good might be more accurate (Wikström, 2013). Digital information goods are characterised as “easily reproduced, easily transferred, easily searched, and easily stored.“ (p. 16), processes which can be repeated multiple times without high costs (Bockstedt et al., 2006).

Going a step further, one could argue that in today’s digitised environments, music might overall be best described as a service. According to Wikström (2013), the shift from selling tangible goods to providing immaterial services is one of the aspects that define the ‘new music economy’, and it is a development that affects the entire music industry. Music as a service can be understood by two main characteristics, the first being the simultaneity of distribution and consumption processes. This signifies that the provision of the service, i.e. the streaming of a song, takes place at the same as the user receives the service, i.e. listens to the song. The user is thereby integrated into the service (Dörr, Wagner, Benlian, & Hess, 2013). The streamed song is usually not stored on the consumer’s device but is only available through an internet connection and no rights of ownership are transferred (Dörr, Benlian, Vetter, & Hess, 2010; Dörr et al., 2013).

Secondly, the revenue model of music as a service is based on constantly paid subscription fees instead of a payment per purchase. This is often complemented by free subscriptions which are financed by the placement of advertisements (Dörr et al., 2013). These two characteristics of music as a service capture the process of music consumption via streaming platforms as well as the respective business model behind it. Thus, it is concluded that understanding music as a service might be more accurate and timelier than the notion of music as a product.

2.1.3 Implications for music listeners

For the music industry, the disruption caused by digitisation and the rising popularity of music streaming entails that exposure to and sale of music, two formerly distinct concepts, are now being merged. Consequently, revenues depend not on a single moment of discovery and purchase, but on the listeners’ subscription to music streaming services (Kjus, 2016). Thus, the shift of the music industry towards a business model based on service provision redefines music listeners as users and the generation of profit is subject to their continuing subscription to services (Arditi, 2017). Accordingly, this also affects the listener’s self-perception and their behaviours of music consumption. Instead of collecting physical records, music enthusiasts nowadays often build music collections digitally, for example in the form of playlists without actually owning ‘their’ music (Arditi, 2017; Hagen, 2015).

The role of the service providers in this relation “is not the sale or lending of music, but the service of making all the music available all the time.“ (Dörr et al., 2010, p. 16).

Admittedly, this provides users with the possibility to discover and listen to an abundance of music, but at the same time leaves them with the challenge to find the songs they like. Naturally, this is perceived as an overload of information (O'Reilly et al., 2013). The motivation to pay for streaming services thus not only stems from having access to music but more and more importantly to have some form of guidance that helps to navigate through the vast databases (Wikström, 2013). Consequently, the high importance of music recommendations and the development of increasingly sophisticated recommendation systems on music streaming platforms can be seen as closely connected to the idea of music as service (Dörr et al., 2013).

For the streaming platforms, providing options for musical discovery has become an essential feature by which they define themselves as distinct brands and differentiate their services from those of competitors. While some platforms rely on curation of music either by experts or by their users, others seek to accomplish the task of successful curation by using artificial intelligence (Morris & Powers, 2015). However, the workings of the embedded algorithms are generally not comprehensible to most users which bears the risk that the recommendation systems can be used as promotional tools with potentially negative consequences for less popular artists (Kjus, 2016; Morris & Powers, 2015). Furthermore, since algorithms base their recommendations on previous listening behaviour, there is a chance that users are exposed to less diverse content (Morris, 2015). This in turn leads back to the initial question of this research, namely how are different recommenders perceived and what consequences does this have on the user's reactions to the recommended content.

2.2 Comparing recommender types

AS outlined in the introduction, artificial intelligence is implemented into a growing number of applications and devices today and is thereby becoming increasingly integrated into the users' daily lives. To fill the surfacing research gaps on the social implications of algorithms and their impact on culture, an appeal is being made especially to the social sciences (Crawford & Calo, 2016). Hence, the ways in which people perceive and interact with algorithmic advisors and whether they tend to follow the obtained advice when making a decision has become a topic of interest for researchers in the social sciences. Studies have examined different contexts in which decision-making processes take place and considered various factors that potentially influence them (Logg, 2017). Still, the question whether people are overall willing to follow suggestion made by algorithms cannot be answered definitely because previous studies have arrived at diverging conclusions in this matter (Thurman et al., 2018). The following sub-sections shall therefore provide an overview over the current state of research and summarise findings that led to the formulation of two opposing concepts, whose validity shall be addressed in the subsequent hypotheses of this thesis.

2.2.1 Algorithm appreciation

What studies have relatively unanimously shown is that algorithms are able to outperform humans as recommenders in many contexts, as they are much more accurate in their ability to make predictions based on data (Logg, 2017). Interestingly, algorithms display a superior performance not only in matters of logical reasoning, but also in highly subjective areas where one would initially think of humans to have an advantage. As an example, Yeomans and colleagues have found out that algorithms were more accurate in predicting the sense of humour of the participants in their experiment than the participants' own friends (Yeomans, Shah, Mullainathan, & Kleinberg, 2019). However, in attempts to determine how much these predictions are accepted by people, prior research found evidence for both appreciation and aversion of algorithmic advice (Thurman et al., 2018). Both of these concepts will be discussed in more detail in the following.

In order to understand how laypeople perceive algorithms, Logg, Minson and Moore (2019) conducted a series of experiments in which they compared how much people in different contexts relied on advice provided by algorithms. In their experiments, participants were performing different estimation and ranking tasks but also tasks of having to predict the chart position of a song or the likelihood of romantic attraction between two subjects. The experiments were conducted in a similar manner, in that participants were asked to answer concrete questions, for example on what position a song will be on the Billboard Charts in the upcoming week. They were then offered some advice that could improve the accuracy of their forecast or estimation. Naturally, all participants in these experiments received the same advice, but it was labelled as coming either from an algorithm or from other people (Logg et al., 2019). The results from these experiments showed a clear preference among participants to follow the advice of algorithmic advisors, both when the task was based on evaluating factual information as well as when it required a more subjective reasoning. This phenomenon is subsequently referred to as algorithm appreciation (Logg et al., 2019).

Recent studies have confirmed these findings: In a secondary analysis of data on people's news consumption behaviour in 26 different countries, Thurman et al. (2018) for instance have compared the opinions of respondents on news selection by different sources. Their results not only overall verify the findings of Logg (2017) and Logg et al. (2019), but also contribute to a more nuanced understanding of algorithm appreciation. For instance, the study showed that people were more favourable of an algorithm which used their own previous news consumption as a data source in comparison to if it evaluated the previous consumption of their peers. Although the recommender systems on popular music streaming platforms such as Spotify are using a number of different filtering techniques, it is a common functionality of them to take the previous behaviour of the respective user into account (Möller, Trilling, Helberger, & van Es, 2018; Morris, 2015). Thus, transferring the findings of both Logg et al. (2019) and Thurman et al. (2018) to the context of this study, it can be assumed that users of music streaming platforms are generally more favourable of recommendations by algorithms than recommendations by other people.

In addition, Thurman et al. (2018) included a number of contextual factors in their analysis, which brings to mind that people's preference for algorithms as recommenders has to be understood in relation to the particular context. As such, the way how people consume news and the degree to which they are invested in it affects their general preference for news curation and also their perception of algorithms as curators. For example, an increased interest in news was found to positively impact the preference for algorithmic news selection. Hence, in the context of music streaming, the way how people use streaming platforms, their listening behaviours as well as their interest in discovering new music could affect their perception of different recommenders and should be paid attention to.

2.2.2 Algorithm aversion

As mentioned before, a number of previous studies has arrived at opposing conclusions as the ones outlines above. For example, Yeomans et al. (2019) have identified a tendency towards algorithm aversion. In their study, participants were given the task of evaluating jokes recommended to them either by algorithms or by their friends and other participants. Overall, participants were rather reluctant to follow the recommendations which an algorithm suggested to them. One possible reason for this observation is that algorithmic recommenders are perceived as less understandable (Yeomans et al., 2019). Interestingly, although people's inherent ability to understand humour and humorous nuances would presumably make them perform better at this task than artificial intelligence systems, this assumption was not supported by the findings of Yeomans et al. (2019): even a simple algorithm was able to outperform the participants in regard to the accuracy of predicting other people's preferences for humour. This study was not the first one to point towards algorithm aversion, in fact other researchers such as Dietvorst, Simmons and Massey (2015) have also found people to favour the suggestions of a person over those of an algorithm, in this particular case however after observing the algorithm being inaccurate.

By critically by critically examining the research design of some of these studies, Logg (2017) highlights some constraints to the universality of algorithm aversion. For instance, previous studies concluded that their participants favoured suggestions made by human advisors by letting the participants choose between a suggestion from an algorithm and their own assessment in order to solve a task (Logg, 2017). Yet, this conclusion ignores a confounding factor: people usually exhibit a general predisposition to favour their own judgement and value their own assessment more than that of others – a phenomenon which is also referred to as overconfidence. The related concept of over-precision moreover describes the tendency to over-estimate the quality and accuracy of one's own judgement (Logg, 2017).

In their experiments, Logg et al. (2019) have taken the influence of overconfidence into account and examined the moderating role of the concept. To do so, they evaluated the influence of the way in which advice was presented on participants tendency to accept it. Their assumption was that it makes a difference to choose between an algorithm's and a random person's assessment in comparison to choosing between an algorithm's and one's own assessment. Indeed, when participants had the choice between the advice of a random person and an algorithmic advisor, they tended to favour the algorithm. In the second case however, participants displayed a preference for their own estimation (Logg et al., 2019).

This is consistent with other previous studies on the phenomenon of overconfidence, in which participants - when being offered the choice between the suggestion of an algorithm and their own judgement - tended to discard the advice of the algorithm (Soll & Larrick, 2009). As a possible explanation for this, Logg (2017) refers to the fact that people cannot exactly know or reproduce other people's way of thinking: "[t]he literature on advice-taking shows a robust effect of discounting advice from peers because people have access to their own reasoning and not to others'. Thus, people often disregard advice when comparing it to their own knowledge." (p. 7). This emphasises that previous studies supporting algorithm aversion have to be reflected critically and people may not be inherently averse towards suggestions by algorithms. While there is a general tendency to prefer one's own judgement over that of other sources, this does not generally shed a light on people's reactions to algorithms as advisors or recommenders.

Therefore, following the model of algorithm appreciation as introduced by Logg et al. (2019), it is expected that participants in this research will display a favourable attitude towards algorithms as recommenders for music on streaming platforms. Furthermore, considering the tendency to discard suggestions from other people, it is expected that peers will be perceived less favourably as recommenders. Thus, the following hypothesis is stated:

H1: Participants show a more positive attitude towards algorithmic recommendations than towards recommendations made by peers.

2.2.3 Expertise as intervening factor

Logg et al. (2019) have furthermore indicated that people's perception of expertise might be an important aspect in understanding how and why they tend to accept advice or follow certain recommendations respectively. While expertise will also be discussed as one of the dimensions of source credibility, at this point it is necessary to assess the role that expert recommenders might play in comparison to algorithms and peers. Experts as recommenders have not been included in the experiments by Logg et al. (2019), yet Logg and colleagues assume that expertise is a factor that influences how people perceive advice and suggest that expert sources might be valued more than algorithms (Logg et al., 2019).

A possible reason for this assumption is that people have a predisposition to positively perceive any advisor which is presented to be an expert. Hence, by labelling a source to be an expert, that source is by default expected to give superior advice compared to any non-expert (Logg, 2017). Additionally, believing in the skills and competencies of the advisor increases the degree to which people trust it. Naturally, experts are expected to have extensive knowledge about their specialist field and are therefore trusted more (White, 2005).

As research shows, the idea that an expert is more valued as advisor than an algorithm has partly been confirmed, but also been contested. According to Thurman et al. (2018), respondents when given the choice between news articles selected for them by an expert or an algorithm preferred the expert only when the algorithm based its selection on the news consumption behaviour of their peers. If the algorithm processed data about the respondents themselves, they preferred it even over the expert. Only in a few countries, people overall preferred news selection from an expert over any kind of algorithm (Thurman et al., 2018).

Again, this emphasizes the importance of examining different kinds of algorithms and also different contexts before any universal statements about the acceptance of recommendations by algorithms can be made. The quoted study by Thurman et al. (2018) did not involve a data collection specifically designed to study social implications of algorithms but in fact used secondary data from the Reuters Digital News Report. Thereby, their analysis is limited to the specific topic of news article selection.

It should be noted that the characteristics of the respondents' news consumption and contextual factors such as their interest in news have an influence on their favourability towards curation of news articles. Furthermore, the amount of trust that people generally have in the media and how they judge the quality of their respective national news outlets impacts their perception of journalists and thus probably has a significant influence on their opinions on the experts in the study by Thurman et al. (2018). In countries where generally a positive impression of the national news media prevails, people might also value the opinion of journalists and editors higher, thus expressing a preference for expert news selectors in the respective survey. Conversely, in countries where people have a lower opinion of the national news media, they might be less favourable of journalists and editors and hence prefer news selection by algorithms (Thurman et al., 2018).

Noticeably, the topic of music recommendations differs substantially from the context of news. The expert recommenders in the presented research are music curators, a profession which participants most likely will not have encountered directly and thus will not have many preconceived ideas about. This allows them to perceive the expert recommenders in this study according to their own definition of expertise instead of being influenced by culturally and socially shaped notions which are attached to the profession of a journalist. Thus, under consideration of various contextual factors, it is reasonable to assume that expertise can be a highly influential factor in accepting advice and forming behavioural intentions. Following the initial assumption about expertise made by Logg et al. (2019), the second hypothesis is stated as:

H2: Participants show a more positive attitude towards expert recommendations than towards recommendations made by algorithms.

2.3 Perception of online messages

After providing a substantial insight into different recommenders and previously researched reactions to these recommenders, it is crucial to subsequently explore mechanisms that could potentially provide explanations for why different recommendations are perceived positively or negatively. To do so, the following sub-section introduces the topic of source credibility and outlines how the perception of it might be a significant influence on an individual's tendency to accept a recommendation. Because in order to understand how people evaluate content they come across online, it is necessary to not only look at the content itself, but to also consider other factors such as their perception of the respective source. Transferring this to the context of the presented research means that the way how users of music streaming platforms perceive the different recommender types algorithms, music curators or peers potentially has an effect on whether they tend to accept the respective recommendations. A later sub-section shall reflect on the construct acceptance of recommendations and illustrate how it can be understood and theoretically conceptualised.

2.3.1 Source credibility

The degree to which people perceive a source of information positively can substantially influence their attitudes and reactions in regard to the information (Cheung, Lee, & Rabjohn, 2008). This can help to explain why people trust certain sources of information and accept certain messages more easily than others (Mahapatra & Mishra, 2017). Thus, in order to gain an understanding of people's acceptance of music recommendations, it is necessary to assess the role of the recommender. Especially the degree to which a recommender is seen to be trustworthy and knowledgeable in the field of music could influence people in their acceptance of the recommendation (Mahapatra & Mishra, 2017). This effect can be further examined by considering the concept of source credibility.

Source credibility helps to analyse the perception of messages by assessing the extent to which people rate the source of information as credible (Attaran, Notarantonio, & Quigley, 2015). In essence, it entails the notion that any sender of a message that is perceived as possessing a high degree of credibility is more effective in achieving its communicative goal and evoking a positive attitude towards its message (Sternthal, Dholakia, & Leavitt, 1978). Source credibility is thus also a relevant concept in the study of persuasion (Attaran et al., 2015; Cheung et al., 2008). Hovland, Janis and Kelley (1953) have described expertise and trustworthiness as the two most important dimensions that constitute credibility. Expertise can be defined as the extent to which a source is "qualified to discuss a subject and has ability to perform subject-related tasks" (Hansen, Lee, & Lee, 2014, p. 255). Trustworthy sources in comparison are described as honest and integer (McGinnies & Ward, 1980), meaning sources whose messages can be accepted and approved of without a need for further verification (Hansen et al., 2014).

Previous studies on source credibility subsequently have looked into the importance of expertise and trustworthiness for gaining people's acceptance of communicated messages and achieving persuasive effects (Ohanian, 1990), whereby the factor of trustworthiness has been found to be a stronger predictor for high source credibility (McGinnies & Ward, 1980; Sternthal et al., 1978). Low source credibility in turn cannot only result from lower levels of trustworthiness and expertise, but also if the source is perceived as being biased or as recognizably attempting to persuade or influence the receiver of a message, for example to buy something (Attaran et al., 2015). On that score, commercial sources of information are often perceived as less credible and trusted less (Purnawirawan, De Pelsmacker, & Dens, 2012). Previous research has furthermore shown that it is necessary for people to have sufficient information about a source in order to determine its credibility. This can be difficult in online environments, where sources are often anonymous (Cheung et al., 2008).

Source credibility has been found to be especially useful for researchers in the fields of advertising and marketing. A common scale with which trustworthiness and expertise can be measured was developed by Ohanian (1990) and originates from the same context. It furthermore includes the dimension of celebrity attractiveness which can be neglected in the case of this research.

2.3.2 Attitudes and intended behaviours

Taking up some insights from the previous discussion, the perceived credibility of a source has been found to have a significant influence on a person's attitude and intended behaviour (Attaran et al., 2015). Both of these concepts are associated with each other as attitudes generally play an important role in determining intended behaviours (Gursoy, Spangenberg, & Rutherford, 2006). Concrete examples for this apparent impact can be found in studies on brand attitude and purchase behaviour that have detected an increase in the intention to buy a product when attitudes are improved (Schivinski & Dabrowski, 2016). Thus, attitudes towards music recommendations are closely connected to a person's intention to listen to a song, add it to a playlist or search for other works by the same artist. In order to comprehend people's acceptance of different music recommendations, it is therefore necessary to assess both their attitudes and their intended behaviours in relation to the recommendation.

As previous research has moreover shown, perceived credibility of a source positively affects the reaction to eWOM messages such as reviews or recommendations. Noticeably, the positive effect of source credibility in this context was even stronger than the effect of the perceived credibility of the message itself (Mahapatra & Mishra, 2017). Following these findings from previous research on online reviews, it is expected that similar effects can be found in the context of music recommendations on streaming platforms: recommendations vary in the degree to which they are being perceived positively and this might not only be connected to the recommended content but also to the users' perception of the recommender type. Users potentially assign different levels of credibility to recommenders such as algorithms, experts and peers, which impacts their acceptance of the respective recommendation. Recommender types which are seen to be trustworthy and to possess a high degree of expertise about music could be more likely to elicit the users' acceptance of the recommended songs or artists (Mahapatra & Mishra, 2017). Thus, the following two hypotheses are stated as:

H3: Recommenders with a higher perceived credibility positively affect the attitude towards the recommendation.

H4: Recommenders with a higher perceived credibility positively affect the intended behaviour in relation to the recommendation.

Attitudes can be defined as representations of "a person's general feeling of favorableness or unfavorableness toward the target object" (Purnawirawan et al., 2012, p. 247). A possibility to assess the extent to which people are favourable for a music recommendation is to look at how useful they consider it to be. The perceived usefulness of information is an important aspect of its perceived quality and therefore impacts people's tendency for information adoption (Cheung et al., 2008). Mahapatra and Mishra (2017) for example have made a connection between accepting a recommendation and perceiving it as useful and valuable. Information usefulness is thus not only an influential key-factor in the process of decision-making, it is also closely associated with people's attitudes and intended behaviours in the context of online reviews (Purnawirawan et al., 2012).

Therefore, it is assumed that when a recommendation is seen as useful people have a more positive attitude towards it. A positive attitude in turn increases the likelihood for interaction with the recommendation - for example listening to the recommended song - as attitudes have been found to determine intended behaviour (Gursoy et al., 2006; Purnawirawan et al., 2012). To quote Cheung et al. (2008), "if others think that a comment within an online community is useful, they will have greater intention of adopting the comment. The perceptions of usefulness of opinions would predict intentions towards adopting that idea" (p. 233).

Noticeably, this approach of treating attitudes and intended behaviours as outcomes of information usefulness represents only one possibility of assessing people's acceptance of music recommendations. There are numerous factors that influence how a person adopts information that have been left out of this research. Precisely, influential theories that address persuasion and information processing such as the Elaboration Likelihood Model by Petty and Cacioppo (1984) or the theory of reasoned action which focuses on relations between attitudes and behaviours by Fishbein and Ajzen (Madden, Ellen, & Ajzen, 1992) have not been included in the previous reflection. However, it has to be acknowledged that within the scope of this research, not all factors can be taken into account, thus making it necessary to use a simplified but effective research model.

2.4 eWOM scepticism

Previous sections have repeatedly referred to recommendations as a type of communication in particular online settings, which can also be described as eWOM contexts. In the following, this context shall be explained in more detail. The term eWOM – short for electronic word of mouth – derives from word of mouth communication, i.e. interactions between people in which they share information about products or services (Lee & Youn, 2009). Cheung et al. (2008) have described eWOM as "any positive or negative statement made by potential, actual, or former customers about a product or company, which is made available to a multitude of people and institutions via the Internet" (p. 230 f.). Hence, eWOM can be seen as an extension of word of mouth that makes each statement visible to a great amount of people all over the world (Cheung et al., 2008). Thereby, eWOM significantly differs from the interpersonal setting and familiarity between the communicators associated with word of mouth communication where people know each other and therefore are able to establish relatively high levels of trust (Lee & Youn, 2009). In contrast, eWOM is passed on in a context “characterized by uncertainty, anonymity, and lack of users' control“ (Zhang, Ko, & Carpenter, 2016, p. 199).

Over the past two decades, online communication channels and the opportunities they offer in terms of consumer communication and user-generated product information such as reviews and recommendations have become a focus of attention for both the corporate and the academic field. Especially researchers in the fields of advertising and marketing have conducted numerous studies on eWOM communication. eWOM messages have generally been determined as being able to strongly influence customers in their decision to buy a product or not. The impact can potentially be even bigger than that of conventional advertising measures, making it an important aspect for companies to include in their marketing strategy (Cheung et al., 2008).

However, as Lee and Youn (2009) have stated, there are no constraints on who can make what kind of claims about a product online. Hence, this bears the possibility that companies use eWOM settings as a way of positively influencing the reputation of their products through biased or fake product reviews (Lee & Youn, 2009). The experience of being manipulated or deceived by untruthful eWOM messages in turn might lead to the development of a general suspicion among users towards product recommendations in online settings and the respective sources. This factor, also referred to as eWOM scepticism, can play a crucial role for explaining the behaviour of customers and users of online services and should therefore also be considered when analysing the perception of source credibility (Zhang et al., 2016).

People with higher levels of scepticism towards online product information might overall be less favourable about online content recommendations – regardless whether the source is another person or an algorithm – and consequently assign lower levels of credibility to any type of online source. Additionally, as pointed out by Mahapatra and Mishra (2017), a substantial amount of previous research has reported influences of eWOM not only on people’s purchase decisions, but also on how they evaluate products and adopt information in online settings. This suggests that eWOM messages could have an impact on a person’s general attitudes and behavioural intentions. Despite the fact that is has been recognised as an influential factor, eWOM scepticism was not taken into account and included into the research design of numerous previous studies on source credibility, which potentially could have skewed their results (Zhang et al., 2016).

Building on these insights about the influential role of eWOM scepticism, it can thus be assumed that the degree to which people are affirming or sceptical towards online recommendations could influence their perception of the credibility of different recommender types. To account for this potential influence, eWOM scepticism is included into the presented research model as a moderating variable and the following moderation hypothesis is stated:

H5: eWOM scepticism will moderate the relationship between recommender type and the perceived credibility of the source of the recommendation.

Although scepticism towards advertising has been analysed by prior research, Zhang et al. (2016) emphasise that this context differs considerably from eWOM. When confronted with a conventional advertisement, people are usually aware of both the originator as well as the persuasive aim of the message. This however is not the case with eWOM: people have more or less no possibility of knowing who is making claims about a product with what motives and whether those claims are truthful or not. Based on established scales to measure mistrust towards advertising, Zhang et al. (2016) have therefore developed a scale with which these three dimensions of scepticism in the context of eWOM communication can be measured.

2.5 Summary

In today's complex online environment, source credibility remains a highly useful concept to examine which factors make users perceive certain sources of information as trustworthy and knowledgeable and how this perception affects their interaction with online content. This is particularly relevant in the study of eWOM messages like reviews or recommendations and when dealing with anonymous, non-transparent or even non-human sources of information, such as algorithms.

As outlined, gaining an understanding of how and why recommendations are accepted requires to carefully look into each concept that is involved in the process and to take various influential factors into account. Drawing from the conceptually close connection between attitudes and intended behaviours, acceptance of a recommendation in this case is approached as a construct consisting of positive attitudes and intentions to interact with the recommended song. Based on the previous discussion, it is furthermore inferred that the perceived credibility of different recommender types could possibly explain why users of music streaming platforms accept certain recommendations and discard others. In other words, source credibility can be described as a mediator in the relation between the recommender type and people's acceptance of the respective recommendation. This implies that source credibility can be expected to account for this relation and to provide insights on how exactly the recommender type affects people's acceptance of music recommendations (Geuens & De Pelsmacker, 2017). To assess the validity of this assumption, the following mediation hypothesis is stated as:

H6: Perceived source credibility mediates the effect of the recommender type on the acceptance of a recommendation.

It should be noted that the mediator is not to be confused with the moderator – in this case eWOM scepticism - which influences how strongly and in which way two variables are related (Geuens & De Pelsmacker, 2017). In regard to all of the discussed variables and concepts, the presented research model can be described as a model of moderated mediation. In summary, figure 2.1 shows how the different concepts are assumed to be related and which hypothesis addresses which expected effect. It is assumed that there will be a direct effect of the recommender type on attitudes and intended behaviours with expert recommenders causing more positive reactions than algorithms, which are in turn expected to have a more positive influence than peers. Furthermore, an indirect effect of the recommender type on attitudes and behaviours via its perceived credibility is expected, with higher credibility leading to both more positive attitudes and a higher likeliness for positive interaction with the recommendation. The relation between the recommender type and source credibility is expected to depend on participants' levels of scepticism towards eWOM sources, as expressed by the moderation hypothesis.

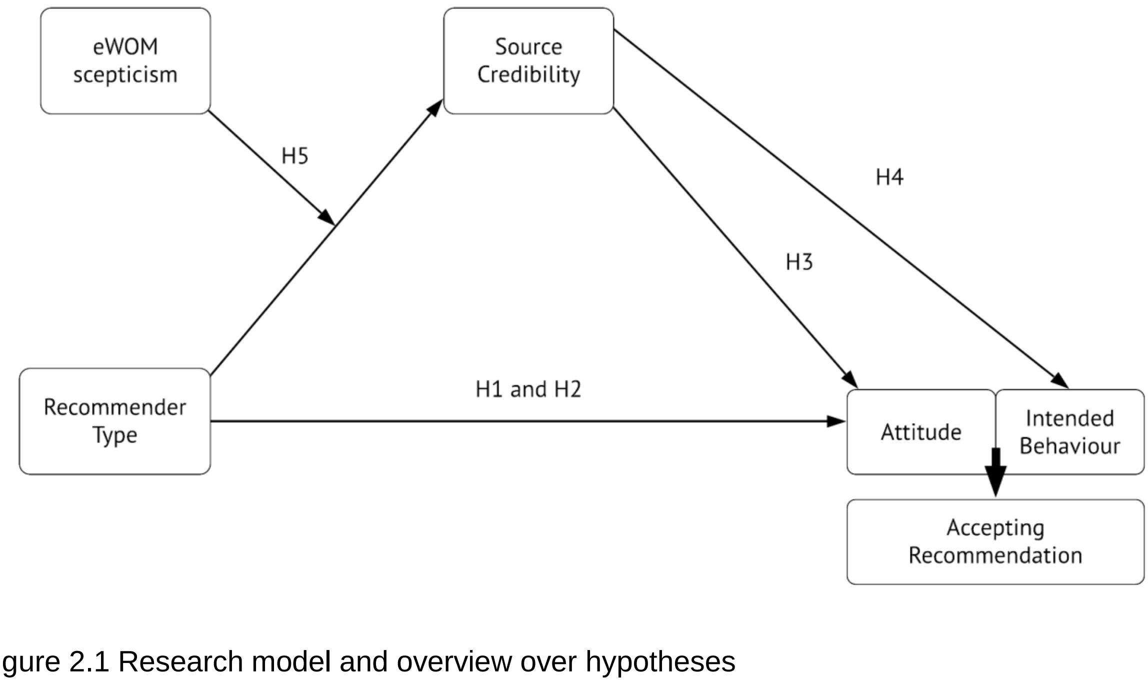

Figure 2.1 Research model and overview over hypotheses

3 Method

3.1 Choice of Method

Overall, the aim of the presented research is to explain in what ways different recommender types have an effect on streaming users' attitudes towards and intentions to interact with music recommendations on streaming platforms. As quantitative research is generally concerned with detecting influences between different concepts and finding explanations for existing relations (Wrench, 2017), quantitative methods of data collection and analysis are the most suitable approach for this research aim. The main assumption, that there is a relation between the recommender and the acceptance of a recommendation, was deductively derived from previous research on concepts such as algorithm appreciation and source credibility, as explained in the previous chapter. A deductive approach is commonly used in quantitative research to formulate hypotheses which are then tested by using standardised measures to gather numerical data for statistical analysis (Neuman, 2014).

Quantitative research methods furthermore appear to be suitable for this thesis as it is concerned with detecting a broader explanation for the ways in which streaming platform users' acceptance of recommendations are affected by the recommender type. Thereby, the aim is not to gain a complete in-depth understanding of a few specific cases but to understand the effects and relations of interest with the possibility to project findings to the population of music streaming users (Neuman, 2014).

Adapting explanations from research on algorithm appreciation, source credibility and eWOM and testing their applicability in the context of music recommendations gives this research an explanatory character, which in turn prompts experiments as a suitable research method (Neuman, 2014). More important for the decision for this method however is the fact that experimental research offers the possibility to analyse causal relations between two or more variables (Neuman, 2014): it is assumed here that the recommender type and the degree of source credibility individuals assign to it are a cause of their acceptance of a recommendation.

Due to the way how an experiment is designed, it allows to limit the influence of confounding factors which are not included in the research model and focus only on the relations between the variables under study (Neuman, 2014). In experiments, it is furthermore ensured that the exposure to a stimulus takes place before the measurement of reactions. Consequently, an experiment is the only scientific method that fulfils all conditions of causality and provides the researcher with a high degree of control (Bellman, 2017). This aspect is highly important for this research since there are a number of factors, such as how a music recommendation is presented to the user, that could possibly influence the way how it is perceived.

To make sure that changes in the measured variables *attitude* and *intended behaviour* can be ascribed solely to the respective recommender, the experiment is conceptualised as a between-subjects experiment with three experimental conditions and no control group. Depending on the experimental condition, participants are exposed to a recommendation either suggested by an algorithm, a music curator or a peer. The choice for experiments is additionally supported by previous literature: studies on the reactions to recommendations, for example by Senecal and Nantel (2004), as well as research on both the concept of algorithm appreciation (cf. Logg et al., 2019) and source credibiity (cf. Attaran et al., 2015) have used experiments in the process of data collection.

Furthermore, following the examples of research on reviews in eWOM settings (Hansen et al., 2014; Purnawirawan et al., 2012) online survey experiments are seen as particularly appropriate in this case. This method of data collection uses online surveys and implements a number of crucial features of experiments into the research design, such as random assignment and the manipulation of the independent variable. The dependent variables can subsequently be measured by using scales. For this thesis, the online service Qualtrics was used to realise the online survey experiment, since neither the exposure to the different recommenders nor the measurement of attitudes and intended behaviours in reaction to the recommendation required participants to be physically present.

Online survey experiments additionally bear the advantage that data is not collected in the rather artificial environment of a laboratory, thus making it easier to create a more realistic setting. To address the issue of realism, this experiment uses a vignette, an artificial copy of the situation that is being researched (Geuens & De Pelsmacker, 2017). Vignettes can easily be implemented into combinations of experimental and survey research. They are a useful technique to approach complex situations by re-creating relevant aspects in the form of descriptions and visuals. In this way, researchers can gain insights into how participants would effectively react in the respective situation (Engelmann, 2017), like receiving a music recommendation on a streaming platform for example. The elements that are used to create the vignette are described in further detail in the section on research design.

Another strength of online survey experiments is that a more diverse sample can be obtained by including people from different national backgrounds and with different occupations or education levels (Bellman, 2017). Letting people take part in the experiment remotely from their own devices additionally lowers the effort for them to participate, which might increase the response rate and facilitate the recruitment of a sufficient number of participants.

3.2 Sampling

The units of analysis of this research are users of music streaming platforms. As this research focuses on music recommendations in general and not on the way how music is recommended on one specific platform, it is reasonable to include users from all different kinds of streaming services in the sample. Since many platforms have embedded the practice of recommending music into their service, it can be expected that most streaming users are to a certain degree familiar with music recommendations - independent of the platform they use - and have encountered at least one of the different recommenders. On popular platforms such as Spotify or YouTube, recommendations are furthermore available to both users with a paid and with a free subscription, thus the type of subscription is not considered as a selection criterion. In addition, findings from the reviewed literature do not support the inclusion of specific demographics such as gender, age or nationality as sampling criteria. Hence, the sampling criterion is limited to being a user of at least one music streaming platform.

The sample of this study is divided into three sub-samples corresponding to the three experimental conditions. Generally, these subsamples should have a sufficiently big size to allow for meaningful comparisons of the measured attitudes and intended behaviours and to have a sufficient level of statistical power. This is important to avoid type II errors and to register relevant effects as statistically significant (Geuens & De Pelsmacker, 2017). Furthermore, it is recommended to increase the sample size if the research model includes one or multiple moderator variables (Geuens & De Pelsmacker, 2017). However, Geuens and De Pelsmacker (2017) emphasise that very large subsamples can also be problematic by increasing the likelihood for type I errors to occur - meaning that effects which are de facto not existent appear to be statistically significant. Hence, their suggestion is to achieve a subsample size of 30 to 40 participants, with respectively higher numbers if moderation of effects is analysed in the research as well. Thus, the ideal sampling size for this research is determined between 50 to 60 participants per condition, leading to a total sample size of 150 to 180 participants.

3.2.1 Sampling Method

As is common in quantitative research, the sample of participants should be determined by probabilistic sampling procedures and drawn randomly from a sampling frame, because this provides an optimal basis for making statistical inferences from findings in the sample to the population (Neuman, 2014). However, it is barely possible to assemble a sampling frame containing all users of music streaming services from which a representative sample can be drawn within the time and cost constraints of a master thesis. Although the use of non-probabilistic sampling methods is less preferable from a quantitative perspective, the disadvantages tend to be connived in cases like this when the realisation of random sampling lies outside the scope of the research project or when the population is rather abstract and its characteristics are difficult to determine (Saldaña, 2017). Additionally, it is considered more important to retrieve a sample that is relevant in regard to the study's theoretical background and to find participants who are able to relate to the stimuli and questions in the survey experiment than to establish representativeness (Geuens & De Pelsmacker, 2017). Therefore, a combination of convenience and purposive sampling was used.

Convenience sampling refers to the practice of including individuals in the sample who are most easy to reach (Neuman, 2014), while purposive or expert sampling requires a certain degree of knowledge about the population in order to find participants who are particularly relevant in regard to the aim of the research (Saldaña, 2017). For this thesis, it was deemed useful to include individuals in the sample who are interested in music streaming and who are reflective about music recommendations. This was achieved by obtaining the sample through social media channels and administrating the survey experiment through Facebook groups for audiophiles and streaming enthusiasts. Among these were groups like 'StreamingMusicMatters & Qobuz fan page' and 'TIDAL High Fidelity Music Streaming'. It was estimated that members of these groups might be more receptive to the questions in the questionnaire in comparison to less frequent streaming users

By following this strategy, the sampling relies on the voluntary decisions of individuals from these groups to participate and be included in the sample. This is referred to as self-selection sampling (Sterba & Foster, 2011). While it ensures that the people included in the sample fulfil the criteria and, due to their interest in the topic are likely to give relevant answers, self-selection sampling can also lead to self-selection bias (Sterba & Foster, 2011), meaning that the motivation for people to participate in the experiment stems from their interest in using streaming platforms. Thus, it is closely related to their attitudes and listening behaviours, which in turn are concepts under study. Self-selection bias additionally has a negative impact on the generalisability of the findings (Olsen, 2011).

Despite the weaknesses that the chosen sampling methods have in regard to the representativeness of the sample, they also offer some positive aspects. For instance, response rates are usually higher than among random samples (Saldaña, 2017) and in experimental social science research common practice of using non-representative student samples was avoided (Bellman, 2017). Besides these crucial advantages, the usage of selective sampling methods can be justified under certain circumstances, as indicated previously. Waterfield (2018) recommends providing compensation for potential biases for example by giving a detailed description of demographics in the sample and if possible drawing comparisons to the respective population, which shall be done in the following sub-section.

3.2.2 Description of sample

In total, a number of 242 responses were recorded. In the process of data cleaning, 69 partial, unusable responses and 2 responses from participants who did not fulfil the sampling criterion of using a music streaming platform had to be excluded. One response was furthermore deleted because of apparent straight lining, i.e. most items throughout the different questions were answered identically (Yan, 2008). This leaves a sample of N = 170 valid responses for analysis, with a discernibly bigger share of 103 male participants (60.6 %) and a smaller share of 63 female participants (37.1 %). The remaining 1.8 % preferred not to state their gender and one participant (.6 %) skipped the question. The average age of the sample is 35 years (N = 167, M = 34.83, SD = 13.26), with the largest share of 71 participants having obtained a bachelor's degree (42 %).

The sample (N = 168) was obtained from a broad variety of 22 countries with the majority coming from Germany (44 %), followed by the Netherlands (18.5 %) and the US (8.3 %). Although a large share of participants are from in the same country, the sample is still quite diverse with regards to nationality, or places of residency to be precise. Hence, this emphasises how conducting the data collection online and distributing the survey experiment via international specific-interest groups on Facebook offers the advantage of including participants from different countries in the sample.

In total, 116 responses come from participants with a paid subscription for the streaming platform they use the most (68.2 %), while 54 participants have a non-paid subscription (31.8 %). The most popular streaming platform is YouTube, with 110 participants (64.7 %) stating that they use it for streaming music, closely followed by Spotify, which is used by 107 participants (62.9%) and the services SoundCloud and Tidal, both used by 31 participants (18.2 %). The corresponding question in the survey experiment allowed the selection of multiple answers, accounting for the fact that participants might use more than one platform.

These numbers reflect some patterns which can also be found when looking at how market shares are distributed among different streaming services: YouTube – although strictly speaking not a music streaming platform – is frequently used by people for music consumption (IFPI, 2018) and claims that each month, 1.5 billion registered users watch videos on its platform (Wojcicki, 2017). Among the providers of music streaming, Spotify is the distinct market leader with an alleged market share of at least 36 percent (Music Industry Blog, 2018). Particularly striking however is the relatively high share of participants using the platform Tidal. In comparison to Spotify, which as of April 2019 reports to have reached 100

Million subscribers (Spotify, 2019), Tidal is assumed to have an approximate amount of 3 million subscribers, as reported in 2016 (Singleton, 2018). Therefore, the share of Tidal users in the sample is unlikely to be representative of the population. However, this bias in the sample can be explained by the fact that one of the groups in which the survey was distributed is dedicated to users of Tidal.

On average, participants spend almost 16 hours per week listening to music (N = 170, M = 15.79, SD = 13.47), which is slightly below the global average of 17.8 hours of music consumption per week stated in the recent Music Consumer Insight Report (IFPI, 2018). The three most-selected favourite genres among the participants are Rock (48.8 %), Indie/Alternative (34.7 %) and Pop (32.4 %), which deviates slightly from the listing of top genres by IFPI (2018), according to whom the most popular genres worldwide are Pop (64 %), Rock (57 %) and Dance/Electronic/House (32 %).

3.2.3 Random Assignment

An important element of experimental designs is to randomly assign participants to the experimental groups, or conditions in this case. Random assignment means that every participant has an equal chance to get assigned to each group and it is important for two reasons (Neuman, 2014). First of all, it ensures that the groups are composed as similar as possible, which is the basis for comparing them to each other. Secondly, it helps to rule out confounding factors: if one experimental group significantly differs from another in regard to a certain characteristic, changes in the measured outcome variable cannot solely be ascribed to the experimental manipulation, but could also be related to this difference (Neuman, 2014).

Therefore, random assignment is highly important to the internal validity of the measurement (Harvey & Harvey, 2018). For this research, the question randomization function in Qualtrics was used, which randomly displayed one of the three recommenders to each respondent, while making sure that each recommender was presented equally often. The researcher had no influence on the assigning procedure. Random assignment should not be confused with random sampling, meaning it cannot establish representativeness or eliminate biases in the sample that result from the sampling method. However, by assuring similarity of the experimental groups, any bias in the sample should occur to an equal degree in all groups making comparisons between them valid nonetheless (Harvey & Harvey, 2018).

As mentioned previously, it is useful to compare the sample with the population based on important characteristics such as age, gender and education levels in order to assess the sample's representativeness. But the population of all streaming users is a very broad population and rather vaguely defined, thus the exact distribution of demographic variables cannot be determined with certainty. The usage of random assignment is assumed to account for the potential discrepancies between the sample and the population, as the different characteristics will be equally distributed across the conditions. Yet, it is recommended to assess if the random assignment was successful (Bellman, 2017). Hence, to test whether any substantial differences between the three conditions of this research occurred, a one-way ANOVA was conducted to compare the three conditions in terms of age of the participants and Chi-square test of independence was used for a comparison of gender and education levels across conditions.

The results of these tests are stated as follows: The one-way ANOVA, $F(2,164) = .04$, $MSE = 177.84$, $p = .963$, partial $\eta^2 < .001$ demonstrated no statistically significant differences between the three experimental conditions in terms of age of participants. For gender, the Chi-square test of independence χ^2 ($N = 166$, 2) = 4.16, $p = .125$ demonstrated no statistically significant differences between the three experimental conditions. For education levels, the Chi-square test of independence χ^2 ($N = 169$, 10) = 4.94, $p = .895$ similarly showed no statistically significant differences between the three experimental conditions. These results indicate that the proportions of demographic variables in each experimental condition do not differ significantly from another. Thus, random assignment was successful, providing a basis for valid comparisons between participants that received a recommendation by an algorithm, a music curator or a peer. This is important for the further analysis, especially in regard to H1 and H2.

3.3 Operationalisation

A crucial part of the quantitative research process is to bring the theoretical and the empirical level together by establishing a way in which the theory under study can be made measurable. This process of transforming mostly abstract concepts into variables and determining concrete survey questions which measure these variables is referred to as operationalisation (Neuman, 2014). This section explains how theoretical concepts were operationalised for this survey experiment, and how the five main variables *recommender type*, *attitude*, *intended behaviour*, *source credibility* and *eWOM scepticism* were measured. Furthermore, an overview shall be given over the most important additional variables.

3.3.1 Independent variable recommender type

Recommender type, i.e. the source of the music recommendation, serves as the independent variable in this research, also referred to as predictor variable. Based on the reviewed literature, it was deduced that the recommender could play a crucial role for the way people react to a recommendation. Studies on algorithm appreciation in particular prompted the assumption that different types of recommenders are perceived differently and lead to different reactions to the recommendation (Logg et al., 2019). Thus, it was hypothesised that *recommender type* predicts the acceptance of a recommendation, as well as the degree to which the recommender itself is perceived as credible.

Recommender type was operationalised as a categorical variable, with the three categories algorithm, music curator and peer representing the different recommenders which are compared in this thesis. In the survey experiment, the variable *recommender type* was not directly measured but it served as the stimulus that participants were exposed to. Therefore, the recommenders are not referred to as categories, but as experimental conditions. The stimulus consisted of a fabricated music recommendation that was consistent across all conditions, and a description of the recommender which varied according to the condition participants were randomly assigned to. In the survey flow, the exposure to the stimulus took place after the introductory questions and before the measurement of the dependent variables.

The stimulus of this survey experiment had to fulfil a number of specific requirements. It was particularly important to create a realistic yet not confounding stimulus and to balance the need to provide participants with sufficient information about the recommender with the aim to not lead them on in their perception of it (Geuens & De Pelsmacker, 2017). Following the methodological approach of Logg et al. (2019), it was therefore decided not to narrowly define what an algorithm or a music curator is, but to leave some room for the participants' own sensemaking of their respective recommender. A more detailed elaboration on the creation and presentation of the stimulus will be given in the sections on pre-testing and survey design.

3.3.2 Dependent variable attitude

One of the two dependent variables measured in the survey experiment is the *attitude* a person has towards a music recommendation. In this research, attitudes are described by how positive or negative someone is attuned to a recommendation by assessing how useful as well as how accurate and satisfying it is perceived to be. These approaches are referred to as information usefulness and review impression (Cheung et al., 2008; Purnawirawan et al., 2012). Studies from the field of advertising and marketing have shown that attitudes towards a message are to a certain degree dependent on the respective source of the message. Thus, *attitude* was hypothesised to be influenced by the recommender type - as stated in H1, H2, H3 and H6.

As it is common in the social sciences to measure abstract notions such as the thoughts and feelings of individuals by using scales, *attitude* is a continuous variable (Neuman, 2014). Multiple-item scales are generally preferred, because they provide the researcher with richer, more detailed insights into the measured concept. Moreover, if available, it is recommended to make use of scales that have been tested and applied in previous studies. This practice ensures that the respective survey questions effectively measure the concept under study (Geuens & De Pelsmacker, 2017). A further advantage of these scales is that they ensure a higher level of reliability and validity (Netemeyer, Haws, & Bearden, 2011).

Thus, to gain an insight into participants' attitudes in this survey experiment, a scale for information usefulness by Bailey & Pearson (1983) (average reliability coefficient of .93), as found in Cheung et al. (2008), was adapted to the context of music recommendations. This measure was complemented with the adaptation of a scale for review impression used by Purnawirawan et al. (2012) (Cronbach's α = .96). Both scale questions contained three statements and measured participants' agreement on a seven-point Likert scale from strongly disagree to strongly agree. The first question about information usefulness assessed how valuable, informative and helpful participants found the recommendation to be. The second question about review impression targeted participants' satisfaction with and appreciation of the recommendation.

3.3.3 Dependent variable intended behaviour

The second dependent variable in this research is *intended behaviour*. It refers to the way how an individual intends to interact with a recommended song, which is treated as a proxy of their actual behaviour. Both dependent variables of this research are closely connected, as studies from different fields have shown that attitudes determine intentions for certain behaviours, for example purchase intentions (Gursoy et al., 2006; Schivinski & Dabrowski, 2016). Because of this conceptual interrelatedness, it was expected that the effects occurring between *recommender type*, *source credibility* and *intended behaviour* will be similar to the respective effects on *attitude*, as it was stated in H4 and H6. Furthermore, it was inferred that when a recommendation is perceived as useful and satisfying, participants are for example more inclined to listen to it, thus adverting to a relation between the attitude towards a recommendation and the intention to interact with it (Cheung et al., 2008). If that is the case, *attitude* and *intended behaviour* can be incorporated within the construct of accepting a recommendation.

Similar to *attitude*, *intended behaviour* is a continuous variable, measuring the participants' likeliness to perform specific interactions with the suggested recommendation on a seven-point Likert scale from extremely unlikely to extremely likely. Although intended behaviour was on the theoretical level compared with purchase intentions, the latter was not found to be a suitable measurement, because music consumption on streaming platforms is rather conceptualised as the usage of a service than the purchase of goods (Dörr et al., 2013). Hence, despite the advantages of using validated scales, it was decided to create a measurement for this survey experiment which specifically relates to practices common among streaming users. Six statements were drafted, drawing inspiration from qualitative studies by Hagen (2015) and Kjus (2016) which have explored the user behaviour on music streaming platforms and for instance have highlighted adding songs to playlists and searching for songs by similar artists as common practices on streaming platforms. Furthermore, the functionality of common streaming services such as Spotify was considered, i.e. reflecting on the specific ways how users are able to interact with songs on the platform. One item was phrased negatively (‘I would skip/ignore this song’) and will thus have to be reverse coded for the analysis. Reverse items can be a useful measure to avoid response sets and to identify responses from inattentive participants (Neuman, 2014).

Since participants’ *attitude* and *intended behaviour* are an instant reaction to the presented stimulus, measuring these variables directly after the presentation of the stimulus creates a natural flow of the survey experiment for participants (Elson, 2017). Furthermore, Elson (2017) points out the possibility of order effects, i.e. an influence that a response to one item can have on the response to a subsequent one. Considering that attitudes are determinants of intended behaviour, it is possible that an order effect from *attitude* on *intended behaviour* could occur. To avoid this, the measurement of *attitude* was placed after that of *intended behaviour*.

3.3.4 Mediating variable source credibility

A substantial number of studies have shown that attitudes towards a message and consequently also behaviours are affected by how credible the sender of the message is evaluated to be (Attaran et al., 2015). The assumptions of H3 and H4 seized this idea. Additionally, the degree to which a recommender is perceived as being knowledgeable and trustworthy (Mahapatra & Mishra, 2017) was estimated to serve as a mechanism through which *recommender type* influences *attitude* and *intended behaviour*. In other words, *source credibility* possibly provides a more thorough understanding of how this influence comes into effect and is treated as a mediating variable in this research (Hayes, 2018). A corresponding mediation hypothesis was stated as H6.

Source credibility is a continuous variable, measured by using the Expertise, Trustworthiness and Attractiveness of Celebrity Endorsers Scale (Cronbach’s $\alpha > .80$) by Ohanian (1990). The scale was adapted to this research by selecting items from the two relevant dimensions expertise and trustworthiness. The items were used to formulate positive statements such as ‘I think an algorithm is reliable when it comes to music’. Furthermore, one reversed item based on the semantic pair ‘biased-unbiased’, originally not part of the scale, was added (Cronbach’s $\alpha = .91$), as found in Mahapatra and Mishra (2017). In total, the scale question consisted of eight items and participants’ agreement to each statement was measured on a seven-point Likert scale from strongly disagree to strongly agree. To avoid order effects with the dependent variables, *source credibility* was measured after *attitude* and *intended behaviour*.

3.3.5 Moderating variable eWOM scepticism

In online settings, users often have limited possibilities to identify the sources of messages or reviews. This uncertainty can lead to the development of a certain level of mistrust towards online sources which is described by the term eWOM scepticism (Zhang et al., 2016). Considering that recommendations on music streaming platforms are passed on in an eWOM setting, it was assumed that a person's level of scepticism in this context might influence how much credibility they assign to a recommender. Therefore, as stated in H5, *eWOM scepticism* is treated as a potential moderator: a variable that affects how positive or negative and how strong the effect between the independent variable *recommender type* and the mediating variable *source credibility* is (Geuens & De Pelsmacker, 2017).

eWOM scepticism is a continuous variable, measured by using an eponymous scale developed by Zhang et al. (2016). The scale consists of the three dimensions truthfulness of messages (Cronbach's α = .77) as well as motivation (Cronbach's α = .71) and identity of the message sender (Cronbach's α = .79) (Zhang et al., 2016). As the scale was originally developed to assess levels of scepticism towards online reviews and the people writing them, some rephrasing was necessary to adjust the items to this research. Thereby, two items of the identity dimension were combined, because their literal sense did not suit the context of music recommendations. This resulted in the item 'The source of an online recommendation is not necessarily who they appear to be'.

Finally, the corresponding scale question consisted of eight statements and the level of agreement to these statements was measured on a seven-point Likert scale with the same labels as used for information usefulness, review impression and *source credibility*. Again, to avoid order effects with the other variables involved in the moderation hypothesis, *eWOM scepticism* was measured after *source credibility* as the last of the main variables.

3.3.6 Additional variables

Apart from these main variables, which were included in the hypotheses of this research and thus crucial components of the data analysis, ten additional variables were measured in the survey experiment. Although these are not in the focus of interest, they can nonetheless provide information to increase the understanding of the effects between independent and dependent variables and explore potential relations that were not hypothesised initially. Furthermore, by measuring variables which are not part of the research model but could potentially influence it, confounding factors can be identified later and it may help to explain the validity or lack thereof within the results (Webb, 2018).

In the first part of the questionnaire, a number of introductory questions were asked. The main purpose of these questions was to conceal the actual aim of the survey experiment and to create a realistic context for the measurement. These questions addressed the participants' usage of music streaming platforms as well as their musical preferences. This provided more detailed information about the participants, which could also be used to enhance the analysis of their attitudes and intended behaviours.

Music streaming platform is a categorical variable that measured which platforms participants use to stream music. It was measured by a multiple-choice question that allowed the selection of more than one answer, since participants might use several platforms. The twelve answer categories consisted of ten widely used and well-known music streaming platforms - such as Spotify, Apple Music and SoundCloud - and an open text field to give participants the option to state platforms they use which were not part of the list. The question also included YouTube as an answer category, although this is strictly speaking not a music streaming platform. However, according to IFPI (2018), 47% of the time spent listening to music via on demand streaming happens on YouTube, making it reasonable to include the platform here. Furthermore, the answer option 'I do not use any music streaming platforms' was included. Participants selecting this answer option were not allowed to proceed with the experiment because they did not fulfil the sampling criterion.

Average hours is a continuous variable and measured the average amount of time spent listening to music on streaming platforms per week. An open text field was used so that participants could enter their own estimation of their weekly music listening time.

Type of subscription is a categorical, dichotomous variable measured by the question whether participants have a paid or free subscription for the streaming platform they use the most.

Listening behaviours is a continuous variable describing the role that listening to music plays in participants' daily life. The corresponding question in the survey experiment asked participants in which situations they listen to music. It consisted of nine items measured on a seven-point Likert scale from strongly disagree to strongly agree. The statements describing different contexts of music consumption were adapted from the Music Consumer Insight Report, which for example lists 'commuting to work', 'studying' or 'going to sleep' as the most common situations in which people consume music (IFPI, 2018).

Usage purposes is a continuous variable expressing the ways in which participants use music streaming platforms. The corresponding question in the survey experiment consisted of eight items measured on a seven-point Likert scale from strongly disagree to strongly agree. The statements described different possibilities of user behaviour on streaming platforms, such as creating playlists or seeing what friends listen to. The items were based on the findings of different qualitative studies on the user behaviour of streaming users (Hagen, 2015; Kjus, 2016), as well as the different functionalities that the particular platforms offer.

Favourite genres is a categorical variable that measured participants' music preferences. From a list of fourteen categories, participants were asked to choose their three most preferred music genres. In order to provide a list of genres that was exhaustive yet not overwhelmingly detailed, it was decided to use a combination of genre classifications used on different streaming platforms. This procedure made the list comprehensible and ensured that the participants were familiar with the answer categories. Additionally, the option 'other' with an open text field was included, to account for specific genres that were not listed.

The last section of the survey experiment recorded some important demographic variables, which is a standard procedure in quantitative research (Neuman, 2014). Demographic data was also used to compare the experimental conditions and to test whether the procedure of random assignment was successful. The demographic variables used in this research are the following: *Age* was measured as a continuous variable by asking participants to state the year they were born in an open text field. The categorical variable *gender* was measured with a multiple-choice question, that included the additional answer categories 'other' and 'prefer not to say'. *Education* was recorded as the highest level of education participants have obtained by asking a question with six answer categories. Finally, instead of nationality or country of origin, it was decided to use the variable *country of residence*, since the question 'In which country do you currently reside?' bore the least ambiguities. To answer the question, participants could choose the respective country from an exhaustive list.

3.4 Pre-test

Before initiating the actual data collection, two different pre-tests were conducted. This section gives an overview over this procedure, the rationale behind each test as well as the outcome and how it affected the design of the final questionnaire that was used as a tool of measurement in this survey experiment. Generally, pre-testing is a phase in between completing the design of a survey and data collection, in which a survey is tested and critically reviewed. It is highly recommended to include this step into the research process and failing to do so can result in issues during the analysis and lead to an invalidation of the results of the study. In addition, pre-tests are a way of increasing both reliability and validity of the measurement (Ruel, Wagner, & Gillespie, 2019).

There are various issues that could potentially occur in the design of a survey and remain unnoticed by the researcher without pre-testing. For example, questions might not be phrased understandably or unambiguously, the order of the questions could lead to biased responses or – if part of the study – a cover story could be not credible and thus not properly conceal the aim of the study (Neuman, 2014; Ruel et al., 2019). Through pre-testing it is furthermore possible to get an accurate estimation of the time it takes to complete a survey. Ideally, pre-tests should be conducted among a representative and diverse sub-sample of the population under study (Ruel et al., 2019).

While pre-testing, it is crucial to not only test the tool for data collection, but also to interview respondents afterwards to gain an in-depth understanding of their experience with the survey (Neuman, 2014). Furthermore, the usage of behaviour coding, i.e. observing respondents while they fill in the survey, can be helpful to identify potential issues. In the case of this survey experiment, it is highly important to insure that participants are not confused by any of the recommender descriptions and that they read them attentively. Behaviour coding can provide insights on that and thus help to optimise the design of the survey experiment (Ruel et al., 2019).

3.4.1 Pre-test of potential stimuli

As suggested by Geuens and De Pelsmacker (2017), a pre-test was conducted in the preparation of the data collection to determine the presentation of the stimulus. The stimulus of this survey experiment should display a fabricated music recommendation in form of a made-up song. In order to create a realistic and more concrete representation of that song, it was decided to not only describe it but also to present it visually with an album cover, thereby imitating the way how songs are commonly displayed on the interface of music streaming platforms. This is also referred to as context realism, which is particularly important in experimental advertising research (Geuens & De Pelsmacker, 2017).

Following further practices from that field of research, the number of confounding factors that a stimulus adds to the research design should be limited effectively. Thus, the stimulus should be “well made and realistic but at the same time as simple as possible” (Geuens & De Pelsmacker, 2017, p. 85). Since familiarity with existing songs and the participants’ personal tastes might influence their answers, the displayed song had to be fictional without strong resemblances to any real equivalents (De Keyzer, Dens, & De Pelsmacker, 2015; Geuens & De Pelsmacker, 2017). To enhance simplicity, audio was not included since it would introduce confounds and people would most likely be influenced in their perception of the recommender by how much the audio appeals to them. In that case, participants’ reactions would be less driven by their perception of the recommender and more influenced by their preference for the song. An image of an album cover by contrast was estimated to make the representation more realistic while adding a negligible confounding influence.

All things considered, the ideal image to be used in the stimulus of the final survey experiment should look realistic enough to be perceived as an actual album cover, while at the same time not resembling any existing cover too much and not evoking extreme emotional reactions among the participants (Geuens & De Pelsmacker, 2017). Finding the most suitable album cover was achieved by measuring the reactions of a sub-sample of streaming users to ten different made-up album covers by different fictional artists. The ten covers, as well as the names of the artists and titles of the songs, were created by the researcher. An overview over all the covers that were tested can be found in Appendix A.

To conduct the pre-test, the survey software Qualtrics was used. In a set of introductory questions, respondents were asked about their musical preferences and usage of streaming platforms. This allows to check for possible relations between these parameters and the preference for a certain album cover in the analysis. Subsequently, the ten covers were presented in randomised order, because it was assumed that respondents would be less attentive when evaluating the later covers (Elson, 2017). For each cover, respondents were asked to indicate how familiar the artist seems to them, how likely they are to interact with a song from the album and how much they find the cover to be appealing and matching their musical taste. All questions were measured on a five-point Likert scale. After the individual evaluation of each cover, respondents had to select their two most and their two least favourite covers. The last part of the pre-test questionnaire recorded demographics - i.e. age, gender, education - and included a manipulation check by asking respondents whether they had searched for additional information about the artists or albums. The pre-test questionnaire was administered among students and personal contacts fulfilling the sampling criterion, which is a practice recommended by Ruel et al. (2019). In total, 26 responses were collected. The respondents' average age was 30 years (N = 24, M = 30.38, SD = 11.59) with 61.5 % female and 38.5 % male respondents.

The outcome of the pre-test was analysed in a two-stage process. First, covers that respondents had less strong opinions about were separated by looking at the ranking question and selecting those that were chosen less often as the most or least favourite cover. From the three remaining covers, the one with the least variability in terms of familiarity, appeal and likeliness to interact with was chosen by comparing the variances for each question. The variance indicates how much the answers of respondents are centred around the mean of the Likert scale, i.e. not showing an extreme reaction to the cover (Privitera, 2015). The variance of the chosen cover for familiarity is s^2 = .49 (M = 4.58, SD = .70), the variance for appeal is s^2 = 1.15 (M = 3.23, SD = 1.07) and the variance for consistency with musical taste is s^2 = .83 (M = 3.12, SD = .91). For an overview of the variances for each of the three considered album covers, see Appendix B. Based on the given values, the album cover corresponding to the song *Dogma Beware* by the fictional artist *Boots & Laces* was chosen as the most appropriate visual to be presented in the stimulus of the survey experiment. The image of this cover can be found in Appendix A (figure A2).

3.4.2 Pre-test survey experiment

The second pre-test was conducted to test the survey experiment with the main goal to assess whether the presentation of the recommenders was salient enough. Moreover, the credibility of the cover story was reviewed, meaning whether participants believed the suggested song to be an actual recommendation by one of the three recommenders based on their musical preferences and listening habits which they stated in the beginning of the survey. In addition to that, the pre-test served the general purpose of determining if the phrasing throughout the survey was clear and estimating how much time it would take to answer the survey. Again, Qualtrics was used to conduct the pre-test. Nine participants who use music streaming platforms and who did not participate in the first pre-test were recruited from personal contacts, so each condition could be tested by an equal number of people.

When possible, the technique of behaviour coding was used in combination with short interviews after completing the survey experiment, otherwise probing questions to obtain feedback were asked via e-mail. The questions were focused on whether the survey was understandable and clear, especially in regard to the presentation of the recommenders and the subsequent questions measuring the dependent and mediating variables. The participants' answers from the interviews, as well as some observations that were made while they filled in the survey, were analysed and compared after the test. Statistical analyses of the results were not possible due to the very small size of the sub-sample.

The results of the pre-test showed that most participants were not aware of the recommender behind the music recommendation they saw and subsequently had problems answering the related questions. Much confusion in that regard was caused by the expression 'source of recommendation'. Based on this feedback, a number of changes had to be made to the survey, including more detailed descriptions of the recommenders and a change in the overall structure: to avoid the word source throughout the questions and to use the label of the respective recommender instead, three different survey branches had to be created. For instance, all participants who were assigned the recommender music curator followed a specific branch of the survey and statements were phrased as 'I think music curators are experienced with music' instead of 'I think the source of this recommendation is experienced with music'.

Furthermore, the initial cover story - which was intended to make participants believe that they received an actual music recommendation from a database of songs compiled by the researcher - led to confusion, as participants frequently thought that it was the researcher who had selected the recommendation. This issue was fixed in the final experiment by creating a scenario in which participants were asked to imagine that they were setting up an account on a new music streaming platform.

In conclusion, the outcomes of both pre-tests illustrate the high importance of the pre-testing stage in the process of this research. A number of significant issues were brought to light that could have severely interfered with a successful data collection and analysis when undetected. Furthermore, pre-testing was helpful to determine a realistic yet not confounding stimulus for the survey experiment. After evaluating the outcome of both pre-tests, the design of the survey experiment could be finalised as described in the following section.

3.5 Research design and data collection

The final survey experiment was created under consideration of a number of specifics that were implemented in order to create a tool of measurement suitable to the aim of this research. The survey experiment consisted of six parts: an introduction, the set-up of a cover story, the presentation of the stimulus, measurement of the dependent, mediating and moderating variables, a manipulation check and demographics. The full questionnaire that was used to conduct the survey experiment is included in Appendix C.

In the beginning, the participants were given some basic information about the purpose of the survey experiment. This information was kept very general to not reveal crucial details. Before they could start the experiment, participants were asked to give their consent by clicking a corresponding option. They were then presented a list of streaming platforms and asked to indicate which ones they use. Furthermore, it was asked how many hours they spend on average listening to music via streaming per week and whether they had a paid or a free subscription for the platform they use the most. The purpose of this first section was to gain a more detailed picture of the sample and to separate out participants who did not fulfil the sampling criterion. Hence, participants selecting the answer option 'I do not use any music streaming platforms' in the first question were re-directed to an exit page.

The following part was used to set up a cover story. Participants were asked to imagine that they had just opened an account on a new music streaming platform. In order to receive adequate music recommendations on that imaginary platform, they were asked to select their three favourites from a list of music genres. Following that, two more questions asked about the participants' listening behaviours, i.e. in which situations they listen to music, and for what purposes they use music streaming. This not only provided potentially interesting insights for further analyses, but more importantly helped to conceal the real purpose of the study.

Using some form of deception is a common practice in experimental research, since participants might adjust their responses when they know the exact question the researcher is aiming to answer (Neuman, 2014). It was assumed that participants would be more likely to state their genuine opinions about the recommenders if they did not know that this is the focus of the research. Furthermore, the implementation of narrative elements contributed to creating a vignette and increased the level of realism in the survey experiment. Hereafter, participants were told that the data about their musical preferences and listening habits was now being processed and that they would receive a recommendation based on their answers shortly. An animated graphic was used to indicate that the selection is in progress. Additionally, a timer function was implemented on the page, so the survey experiment automatically advanced after six seconds.

In the next section, all participants were randomly assigned to one of the three experimental conditions, meaning they were all presented a song which was described as being recommended either by an algorithm, a music curator or another user of the fictional music streaming platform. Following common practices in experimental research, changes were only applied where it was necessary for the manipulation, while the remaining part of the recommendation was kept consistent (Geuens & De Pelsmacker, 2017). Thus, the song was held constant across all conditions and represented by the album cover that was chosen in the first pre-test, while the descriptions of the recommenders were slightly different.

In order to be able to measure *source credibility*, it was necessary to give some information about the respective source (Cheung et al., 2008). To provide participants with sufficient insights about their recommender and to increase its salience, the functionality of the algorithm, the work of a music curator or the practice of social sharing of music recommendations was briefly explained. The descriptions were phrased as neutral and comparable as possible to not influence participants in their perception of the recommender. The full descriptions are included in the attached questionnaire in Appendix C.

Depending on which condition participants were assigned to, they followed one of three different paths in the survey flow, as the label of the recommender was repeated in the subsequent questions. These were presented in the format of matrix tables. First, the participants' *intended behaviour* was measured by asking them how they would most likely interact with the recommended song based on their first impression. After that, *attitude* was measured by two questions, one of which addressed participants directly by asking how valuable and helpful recommendations from the respective recommender seemed to them, while the other question was phrased in the third person, i.e. participants were asked to state how much they thought other users would appreciate the kind of recommendation they just saw. This wording was adopted from Purnawirawan et al. (2012) and accounts for a potential tendency of people to think of themselves as less affected by mediated messages than others (White & Andsager, 2017).

In order to avoid unwanted influences between the measurements of the different concepts, mediating, moderating and controlling variables should be measured after the dependent variable (Geuens & De Pelsmacker, 2017). Hence, the next question addressed the degree of credibility that participants assigned to the recommender. After this question, the survey followed one joint path again, because in the following it was not necessary to further repeat the label of the recommender. The last matrix table focused on participants' general opinion on online recommendations and reviews by asking how much they agreed or disagreed with different statements adapted from the context of eWOM scepticism.

As recommended by Geuens and De Pelsmacker (2017), a manipulation check was added after the measurement of the dependent, mediating and moderating variables. Manipulation checks are a way of assessing whether participants in an experiment have perceived and understood the manipulation in the stimulus as it was intended by the researcher (Hoewe, 2017). In this case, participants were asked if they remembered the recommended song or artist, who made the recommendation and – if they did – to state the respective information in an open text field.

The manipulation check intended to measure how attentive participants were when looking at the stimulus and if they were aware of their assigned recommender. Recording this was important because if participants were not aware of the source of their recommendation, their attitudes and intended behaviours cannot be interpreted as being caused by it. In other words, outcomes of a manipulation check can support or reject claims about causal relations. In an analysis, they can furthermore help to identify potential confounds when compared with other variables (Hoewe, 2017). If participants of this survey experiment did not remember their recommender, it either might not have been emphasised clearly enough or they might not have given enough attention to the stimulus. This can be cross-checked with the time measurement that was included on the respective page. Thereby, potentially untruthful answers from inattentive participants can be detected and excluded from the analysis (Geuens & De Pelsmacker, 2017).

As online experiments provide the researcher with very little control over the circumstances under which participation takes place (Geuens & De Pelsmacker, 2017), a further manipulation check was included at the very end of the survey experiment with the intention to find out if participants have searched for additional information on the artist or the song while answering the questions. Doing so could affect the degree to which they believe the cover story and thus also affect their answers.

The last part of the survey measured the most important demographics, such as age (stated as year of birth), gender, educational background and current country of residence. Since it was necessary for the researcher to use deception throughout this survey experiment, it was required to include a de-briefing in the end (Neuman, 2014). In a short message, participants were informed that the recommended song was fictional and not recommended based on musical tastes and listening preferences. Additionally, the true purpose of the study was briefly explained with a note that in order to not influence the answers given in the experiment, this information could not have been disclosed earlier.

3.6 Validity and reliability

Throughout the previous sections, various steps in the construction of the survey experiment were mentioned to improve the validity and reliability of the measurement. This section shall summarise those procedures and provide a concise insight into how it was ensured that the survey experiment is consistent and effectively measures relations between the different concepts as intended. Furthermore, the results of the factor analyses are reported.

3.6.1 Validity

In experimental research, a division is made between the internal and external validity of a measurement. The former refers to the exclusion of confounding factors to strengthen the capacity of measuring causal effects, while the latter describes the degree to which effects observed in an artificially created experimental setting can be generalised to real situations (Neuman, 2014). In the conceptualisation of this survey experiment, attention was paid to both of these notions.

A common practice throughout the social sciences is to make use of scales that have been developed and tested in previous studies, since these scales provide a high degree of validity (Geuens & De Pelsmacker, 2017). For this research, it was seen as very useful to draw on advertising and marketing research, as numerous studies from these fields have explored the concepts under study, hence established and widely-used scales to measure them are available. *Source credibility* and *eWOM scepticism* for instance could be measured by the respective scales by Ohanian (1990) and Zhang et al. (2016). Furthermore, research on advertising and marketing is often concerned with brand attitudes or purchase intentions, thus related insights and methodological considerations could be adapted to this research (De Keyzer et al., 2015; Li, 2016).

To establish a high degree of internal validity, pre-testing of the potential stimuli and of the survey experiment were useful steps (Ruel et al., 2019). The first pre-test gave important insights into how a subsample of streaming users perceived different album covers. Based on their responses, it was possible to choose an album cover that evoked the least polarised opinions and emotional reactions – in other words, the album cover with the least amount of confounding influence was determined. The second pre-test helped to assess the effectiveness of the manipulation. Without this test, it could not have been assured that participants notice the recommender they were assigned to as planned and consequently, changes in the two dependent variables could not be ascribed with full certainty to *recommender type* and *source credibility*.

Another way of verifying internal validity is to include a manipulation check in the experimental design (Neuman, 2014). In this survey experiment, one of these checks was used to assess how many of the participants correctly recalled the recommender, which provides some knowledge about how well they perceived it. As the results show, out of *N* = 169 responses, 37 participants (21.9 %) correctly recalled the recommender they were assigned to. One participant did not recall the recommender correctly and the remaining part of the sample stated to not recall it at all. However, it can be assumed that the actual number of participants being aware of the recommender is higher, as the respective labels were repeated in the questions throughout the survey experiment. Another manipulation check was used to determine the number of participants that have searched for additional information on the recommended artist or song while filling out the survey. Finding out that the recommended artist and song are fictional could naturally influence them in their responses and thus entail a confounding influence. The majority of 132 participants (80.5 %, *N* = 169) however did not search for additional information.

A practice to increase the external validity of the measurement is random assignment to compensate for the usage of purposive sampling methods and the resulting lower degree of representativeness of the obtained sample. If participants are assigned randomly to the experimental conditions, the validity of comparisons between these conditions is ensured. In addition, as Engelmann (2017) points out, vignettes can positively impact external validity by adding more realism to an experimental design which might otherwise lack resemblance with the way how the researched situation occurs in reality.

3.6.2 Reliability

A tool for data collection is seen as reliable when repetition of the measurement process produces very similar results (Neuman, 2014). Even though most of the scales used in this survey experiment have been tested in previous studies, it was seen as useful to verify the reliability of the measurement by conducting factor analyses and reliability checks for the continuous variables. The insights from these analyses provide important information for the data analysis about possible underlying dimensions in the data. In addition, it is shown whether the scales, based on which variables for the analysis are created, were reliable in this research (Pallant, 2005). However, before conducting factor analysis, it should be ensured that the data fulfils the requirements, i.e. being continuous, normally distributed and each scale should consist of at least three items (Pallant, 2005). All of these conditions are met, with the exception of one item on the source credibility scale, for which the value of kurtosis indicated that it slightly differs from the normal distribution. However, the deviation turned out as rather small and the sample was sufficiently big to account for this. The results of the factor analyses are reported as follows.

Intended behaviour: In preparation of the factor analysis, one item of this scale had to be reverse coded. Subsequently, the six items which were based on a seven-point Likert scale were entered into factor analysis using Principal Components extraction with Varimax rotation based on Eigenvalues (> 1.00), $KMO = .79$, $\chi^2 (N = 170, 15) = 392.53$, $p < .001$. All items loaded onto one factor and the resultant model explained 54.92% of variance in *intended behaviour*. Subsequently, the reliability for all items of the unidimensional scale was tested. It revealed a Cronbach's α of .83, which indicates that the scale has good reliability.

Attitude: For the variable *attitude*, two factor analyses had to be conducted for the distinct scales of information usefulness and review impression. Each scale consisted of three items based on a seven-point Likert scale. For information usefulness, the factor analysis using Principal Components extraction with Varimax rotation based on Eigenvalues (> 1.00), $KMO = .72$, $\chi^2 (N = 170, 3) = 193.82$, $p < .001$ revealed a model explaining 74.83% of variance. All items loaded onto one factor and the scale is reliable with a Cronbach's α of .83. For review impression, the factor analysis using Principal Components extraction with Varimax rotation based on Eigenvalues (> 1.00), $KMO = .70$, $\chi^2 (N = 170, 3) = 176.39$, $p < .001$ revealed a model explaining 72.95% of variance. All items loaded onto one factor and the scale is reliable with a Cronbach's α of .81. For the analysis, it was necessary to combine both scales into the variable *attitude*. Hence, a reliability check with all six items of both scales was done, revealing a Cronbach's α of .86. Based on this result, it was reliable to combine both scales.

Source credibility: All eight items, based on a seven-point Likert scale, were entered into factor analysis using Principal Components extraction with Varimax rotation based on Eigenvalues (> 1.00), KMO = .83, χ^2 (N = 170, 28) = 636.52, p < .001. The resultant model explained 65.41% of variance in *source credibility*. As expected, items loaded onto two factors, which could be identified accordingly to the literature on source credibility as the following:

The first factor expertise included three items (experienced, knowledgeable, qualified) and is reliable with a Cronbach's α of .87. The factor trustworthiness included four items (honest, reliable, sincere, trustworthy) with a Cronbach's α of .85, also indicating reliability. The item biased, which was reverse coded before the factor analysis, was not included in the original scale by Ohanian (1990) and did not load onto any factor. It was thus omitted from the creation of an overall scale for *source credibility*, which was necessary for the analysis and reliable with a Cronbach's α of .87. All factor loadings are presented in table 3.1.

Table 3.1. Factor and reliability analysis for source credibility

(N = 170)

Item	Expertise	Trustworthiness
I think an algorithm / music curator / other user is ...		
knowledgeable	.91	-
experienced	.89	-
qualified	.78	-
reliable	-	.82
sincere	-	.82
honest	-	.82
trustworthy	.49	.62
R^2	.33	.32
Cronbach's α	.87	.85
Eigenvalue	3.95	1.28

eWOM scepticism: All eight items of the eWOM scepticism scale by Zhang et al. (2016), based on a seven-point Likert scale, were entered into factor analysis using Principal Components extraction with Varimax rotation based on Eigenvalues (> 1.00), KMO = .85, χ^2 (N = 170, 28) = 648.73, p < .001. The resultant model explained 68.39% of variance in *eWOM scepticism*. As the scale had to be slightly altered to suit the context of this research, not all original items were used. Thus, only two instead of three factors as in the original scale could be found. All factor loadings are presented in table 3.2. The factors can be described as the following:

Mistrust included four items referring to feelings of suspicion towards eWOM sources and assumptions that these sources have negative intentions. Although one item ('getting people to buy things') loaded slightly higher onto the second factor, textually it appeared to suit better with the first factor. A Cronbach's α of .83 indicates reliability of this factor.

The second factor was labelled truthfulness, following the wording from the original scale. It included four items reflecting the degree to which information from eWOM sources is estimated to be untrue. Again, because it appeared more suitable textually, one item ('intended to mislead') was added to the factor truthfulness, although the loading was slightly higher for the first factor. The factor truthfulness is reliable with a Cronbach's α of .80. For the analysis, it was again necessary to create an overall scale for *eWOM scepticism*, which was found to be reliable with a Cronbach's α of .87.

Table 3.2. Factor and reliability analysis for eWOM scepticism (*N* = 170)

Item	Mistrust	Truthfulness
The source of an online recommendation might be up to something	.87	-
The source of an online recommendations is not who they appear to be	.83	-
Often the same recommendation is made by the same source under different names	.37	-
Online recommendations care more about getting people to buy things	.56	-
Online recommendations are not generally truthful	-	.86
In general, online recommendations don't reflect the true picture of a subject	.37	.78
We can hardly depend on getting the truth from most online recommendations	-	.75
Most online recommendations are intended to mislead	.56	.54
R^2	.36	.33
Cronbach's α	.83	.80
Eigenvalue	4.23	1.24

Note: Rotated with Varimax, factor loadings below 0.30 excluded

3.7 Method of data analysis

The obtained data will be analysed in SPSS by using Hayes' PROCESS macro to conduct an analysis of moderated mediation. Hayes (2018) describes this analysis – also referred to as conditional process analysis – as a combination of both moderation and mediation, with which different effects in a causal research model and the conditions that influence them can be understood. Thus, moderated mediation permits to test not only H6, but all hypotheses as introduced in the previous chapter within one analysis. The PROCESS macro conducts all the necessary regression analyses in one step, providing an overview over all direct and indirect as well as conditional effects (Hayes, 2018).

To be precise, mediation provides insights into both the direct effect that the independent variable has on an outcome, in this case the effect of *recommender type* on *attitude* and *intended behaviour*, as well as the indirect effect through the mediator *source credibility*. Thus, this step provides information on H3 and H4. Moderation in contrast looks into how a moderating variable affects the effect that one variable has on another (Hayes, 2018), which in this case is addressed by H5 and the expected effect of *eWOM scepticism* on the relation between *recommender type* and *source credibility*.

By creating a dummy variable and using indicator coding on the independent variable *recommender type*, moderated mediation analysis can be conducted with a multicategorical independent variable, allowing to draw comparisons between the different recommenders and thus gaining insights on the claims made in H1 and H2. Overall, this makes moderated mediation analysis a sound choice especially for an experimental design like this one with random assignment of participants to one of three conditions (Hayes & Preacher, 2014). As Hayes (2018) explains, mediation analyses were previously conducted with the causal steps approach developed by Baron and Kenny. However, this approach is not recommended to be used anymore and the number of recent scientific articles relying on the method is decreasing due to criticisms of its accuracy (Hayes, 2018).

4 Results

In order to gain insights into the validity of the six hypotheses of this thesis, two separate analyses of moderated mediation using the PROCESS macro in SPSS were conducted, one with *attitude* and one with *intended behaviour* as outcome variable. Each analysis was conducted with 5,000 bootstrap samples. In the following sections, the results of these analyses shall be reported and explained. Thereby, the direct effects between the different variables will be examined first, followed by indirect and conditional effects. In addition, a further exploration of possible mediating factors will be considered, as well as the influence of demographic variables. All reported coefficients in this chapter are unstandardized.

4.1 Direct effects

To begin with, all direct effects of the independent variable *recommender type* on other variables in the model shall be examined. It was hypothesised in the research model that there is a relation between *recommender type* and *source credibility*. The corresponding regression analysis, in which *source credibility* is treated as an outcome variable, showed that the model overall was significant, $F(5, 164) = 4.50$, $p = .001$, $R^2 = .12$. However, when looking at the distinct effects between the different recommenders and *source credibility*, no significant relation can be reported: Neither the comparison of effects between algorithm and music curator on *source credibility* was significant with $a_1 = -.35$, $t(164) = -.41$, $p = .683$, nor the comparison of effects between algorithm and peer with $a_2 = -.30$, $t(164) = -.36$, $p = .717$. This already indicates that *source credibility* cannot function as a mediator in the relation between *recommender type* and *attitude* or *intended behaviour* but a thorough explanation of that shall be given later in this chapter.

Secondly, the relation between *recommender type* and the attitude towards the recommendation shall be examined. Based on research on the concept of algorithm appreciation, it was expected that participants will have a more positive attitude towards recommendations made by an algorithm than towards recommendations made by peers, as stated in H1. However, when replacing peers with experts, participants' favourability was expected to change, preferring recommendations from experts over those from algorithms. This assumption was stated in H2. In the model of moderated mediation that was formulated for this research, this relation is represented as the unconditional direct effect of the predictor on the outcome variable (Hayes, 2018).

The outcome of the corresponding regression analysis indicates that the model is significant, $F(3, 166) = 34.19$, $p < .001$, $R^2 = .38$, thus *recommender type* predicts the attitude towards a recommendation. Comparing the effects that the distinct recommenders had on *attitude*, it is noticeable that music curator and peer recommender had a negative effect.

Both ‘human’ recommenders caused a less positive attitude towards the recommendation in comparison to the algorithm recommender. To be exact, the peer recommender’s effect on attitude towards the recommendation differs significantly from that of the algorithm recommender with c_2 = -.35, $t(166)$ = -2.52, p = .013. This effect is moderately negative in comparison to the effect of the algorithm recommender. These results support the claim of the first hypothesis: participants in this research show a more positive attitude towards algorithmic recommendations than towards recommendations made by peers. Thus, H1 is accepted.

The effect of the recommender type music curator on *attitude* also differs significantly from that of the algorithm recommender with c_1 = -.32, $t(166)$ = -2.15, p = .033. This effect as well is moderately negative compared to the effect of the algorithm recommender. So, the impact of expertise turns out to be opposite of prior expectations and the results refute the claim of the second hypothesis: participants in this research do not show a more positive attitude towards expert recommendations than towards recommendations by algorithms. In fact, algorithms appear to be the preferred source of recommendations. H2 is therefore rejected.

Incidentally, testing for the aforementioned direct effect of *recommender type* on the outcome variable *intended behaviour* did not lead to significant results. Despite the model overall being significant with $F(3, 166)$ = 4.03, p = .009, R^2 = .07, neither the comparison between curator and algorithm with c_1 = -.17, $t(166)$ = -.81, p = .421, nor the comparison between peer and algorithm with c_2 = -.04, $t(166)$ = -.21, p = .830 adverted to a significant effect. Thus, while the type of recommender predicts the attitude towards a recommendation, it does not predict the intended behaviour in reaction to the recommendation.

The remaining direct effect to be analysed is the effect between *source credibility* and both *attitude* and *intended behaviour*. From previous research on the concept, it was inferred that recommenders with higher levels of perceived source credibility will positively affect participants’ reactions to recommendations, which is expressed in H3 and H4. For the analysis of this effect, *source credibility* was treated as a predictor variable.

The results of the respective regression analyses show that both the model for the outcome variable *attitude*, $F(3, 166)$ = 34.19, p < .001, R^2 = .38, and for the outcome variable *intended behaviour*, $F(3, 166)$ = 4.03, p = .009, R^2 = .07 were significant. The detected effect of *source credibility* on *attitude* with b_1 = .63, $t(166)$ = 9.86, p < .001 was significant and positive. In comparison, the effect of *source credibility* on *intended behaviour* with b_2 = .31, $t(166)$ = 3.47, p = .001 was also significant and positive, but weaker. Thus, the perception of *source credibility* predicts *attitude* as well as *intended behaviour*. Both H3 and H4 are accepted because the positive coefficients show that higher degrees of *source credibility* lead to more positive attitudes towards the recommendation and to more positive intentions to interact with it. However, it should be noted that contrary to the assumption stated in both hypotheses – namely that recommenders with higher perceived credibility will positively affect *attitude* and *intended behaviour* – this effect occurred independently of *recommender type*, since it was not found to be a predictor of *source credibility*, as reported in the beginning of this section.

4.2 Indirect and conditional effects

In addition to the examined direct effects, the research model assumed the existence of a number of indirect and conditional effects. These effects are on the one hand the moderation of the relation between *recommender type* and *source credibility* by *eWOM scepticism*, as expressed in H5. On the other hand, H6 made the assumption that perceived *source credibility* mediates the effect of *recommender type* on the acceptance of a recommendation, represented by *attitude* and *intended behaviour*.

Referring back to the first step of the analysis, no significant effect of *recommender type* on *source credibility* was found. Yet, moderation could theoretically still occur, even if the direct effect was insignificant. But as the results show, the interaction terms when predicting *source credibility* were not significant, neither for the comparison between music curator and algorithm (*interaction coefficient*$_1$ = .21, $t(164) = 1.13$, $p = .260$), nor for the comparison between peer recommender and algorithm (*interaction coefficient*$_2$ = .09, $t(164) = .47$, $p = .638$). H5 is accordingly rejected: no moderation of an effect between *recommender type* and *source credibility* could be observed. *eWOM scepticism* however was found to have a significant direct effect on *source credibility* with *eWOM scepticism coefficient* = -.28, $t(164) = -2.27$, $p = .024$. The coefficient in this analysis indicates that the considered effect is negative, meaning higher levels of *eWOM scepticism* among participants made them assign less credibility to any of the recommenders.

The absence of a significant effect of *recommender type* on *source credibility* leads to a rejection of H6: *source credibility* does not mediate the effect of *recommender type* on accepting a recommendation, represented by *attitude* and *intended behaviour*. This means that while the type of recommender was found to have a direct effect on participants' attitude towards a recommendation, the degree of *source credibility* assigned to that recommender does not serve as a valid explanation for why this effect occurred, even though *source credibility* itself influenced *attitude* and *intended behaviour* directly. All the effects in the research model that have been described previously are summarised and visualised in figure 4.1 as follows.

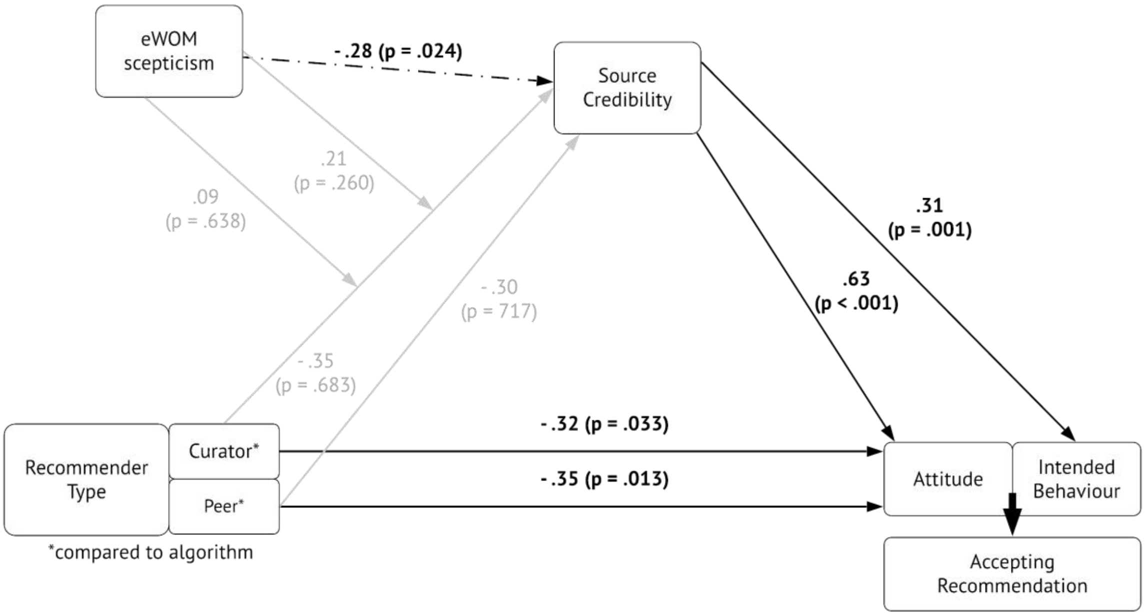

Figure 4.1 Moderated mediation of accepting a recommendation

The discussed procedure represents only one possible approach of determining the occurrence of moderated mediation. Hayes (2015) introduces another way of investigating this matter, which is to look at the index of moderated mediation. This index is described as a “quantification of the association between an indirect effect and a moderator“ (Hayes, 2015, p. 2). Comparing the effects of music curator to that of algorithm on *attitude* ($Index_1$ = .13, *SE* = .12, 95% CI [-.10; .37]) and on *intended behaviour* ($Index_1$ = .07, *SE* = .06, 95% CI [-.05; .21]) shows that both indices of moderated mediation were not significant, since the confidence interval encompassed zero. Similarly, when comparing the effects of peer recommender to that of algorithm on *attitude* ($Index_2$ = .05, *SE* = .12, 95% CI [-.20; .27]) and on *intended behaviour* ($Index_2$ = .03, *SE* = .06, 95% CI [-.10; .15]), the indices were also not significant. Hence, in the case of this analysis, both Hayes’ index of moderated mediation as well as the previously described analysis of the interaction terms lead to the same results, namely that moderated mediation did not occur between the variables under study.

Although the assessment of the index of moderated mediation instead of the interaction terms did not bring forth new or contradicting insights in this case, it can nonetheless be important to consider the method suggested by Hayes (2015) in order to fully comprehend the presence or absence of moderated mediation. Principally, as Edwards and Lambert have put it, moderated mediation can be described as “a mediated effect that varies across levels of a moderator variable” (2007, as cited in Hayes, 2015, p. 2). To account for this definition, PROCESS examined whether there was an indirect effect of *recommender type* on *attitude* and *intended behaviour* through *source credibility* at different levels of *eWOM scepticism*. At first glance, the respective outcome shows that at higher levels of *eWOM scepticism*, the effect of music curator on *attitude* and *intended behaviour* – compared to the effect of an algorithm as recommender – was mediated by *source credibility*. The precise values of the conditional indirect effect can be found in the tables in Appendix D. However, since the index of moderated mediation was not significant, it is inferred that these effects were not significantly different from each other, which in turn means that claims of moderated mediation cannot be supported in this case (Hayes, 2015).

4.3 The impact of expertise and age

In previous literature, source credibility is most often approached as consisting of the two dimensions trustworthiness and expertise (Attaran et al., 2015). These dimensions were also found in this data set, as factor analyses have shown. The aspect of expertise is seen as highly influential, not only in previous research on source credibility but especially in studies that compared different recommenders or advisors (Logg et al., 2019). Hence, an analysis of the same model as above was conducted to cross-check whether expertise by itself could function as a mediator in the relation between *recommender type* and *attitude* or *intended behaviour*.

The results however resemble those of the main analysis: although expertise did predict *attitude* (b_1 = .42, $t(166)$ = 7.26, p < .001) and *intended behaviour* (b_2 = .19, $t(166)$ = 2.44, p = .016), there was no significant effect of the *recommender type* on perceived expertise (a_1 = .09, $t(164)$ = .10, p = .922; a_2 = .34, $t(164)$ = -.38, p = .705). With expertise by itself not being a mediator either, the findings of this analysis emphasise again that source credibility overall is not a suitable concept to provide an explanation for why the acceptance of a recommendation – or at least the attitude towards it – was affected by the recommender type. A conceptually different approach might serve as a more adequate explanation, as shall be reflected on in the discussion that ensues this chapter.

There is a possibility that certain demographic characteristics might have influenced how the users of music streaming platforms come to accepting recommendations by different recommenders. For instance, prior research has found less appreciation of algorithmic advisors among people with lower levels of education, whereas no relation was found between gender and a preference for a specific source of recommendations or advice (Logg et al., 2019; Thurman et al., 2018). An unresolved question however is whether age has an influence on algorithm appreciation: the findings by Logg et al. (2019) indicate that there is no such relation, while Thurman et al. (2018) on the contrary have found people at a higher age to be more favourable for news selection by editors and thus more averse towards algorithmic sources of news selection. In order to provide some clarification to this contradiction, it was examined whether there was an influence between the participants' *age* and their acceptance of a recommendation in this analysis. If an influence occurred generally, it could be further investigated for which recommender types this influence is strongest.

To assess the possibility of an effect of *age* on acceptance of a recommendation, two multiple linear regressions were conducted to predict *attitude* and *intended behaviour* based on the participants' age. The analysis accounted for the *recommender type* as well by including the respective dummy coded variables as independent variables in the regression. The models were not found to be significant, with $F(3, 165) = 1.94$, $p < .126$, $R^2 = .03$ for the outcome variable *attitude*, and $F(3, 165) = .10$, $p < .962$, $R^2 = .002$ for the outcome variable *intended behaviour*. *Age* was neither a significant predictor of participants' attitudes towards a recommendation ($\beta = -.12$, $p = .126$), nor did it predict their intended behaviour in reaction to it ($\beta = -.04$, $p = .597$).

Based on this outcome, it is concluded that *age* did not have an influence on participants' acceptance of a recommendation and that the effects that have been found in the main analysis – namely that *attitude* was predicted by the *recommender type* with more positive results for the algorithm recommender compared to music curator and peer – are consistent across different levels of *age*. In other words, algorithm appreciation was not affected by age in the sample of this survey experiment, which confirms the findings of Logg et al. (2019).

5 Conclusion

In order to provide a cohesive conclusion to this research and to answer the overall research question, the following sections will give an overview over the results and discuss them in the light of the theoretical concepts they relate to. A possible explanation for the observed effect that was not accounted for in the research model will be reviewed. Furthermore, it will be reflected on how this thesis contributes to the current knowledge in different areas of research and what limitations to it are apparent. Finally, indications for future research are given.

5.1 Discussion

5.1.1 Summary of results

First of all, the findings provide an insight into how the different recommenders were perceived by assessing how the participants' attitudes towards a recommendation differed in relation to the source from which they received it. Participants evaluated the same recommendation as useful, accurate and satisfying when it was made by an algorithm, but less so when it was made by another person. In other words, the type of recommender was found to be a cause of attitudes towards recommendations, with a more positive influence of algorithmic than human recommenders. Based on this insight, the presented research confirms the existence of a tendency among people to favour algorithmic sources of recommendations, thus supporting the concept of algorithm appreciation, established by Logg et al. (2019), in the context of music streaming.

Furthermore, some of the conclusions of Yeomans et al. (2019) are refuted, such as the notion that aversion towards algorithmic recommenders or advisors particularly occurs in subjective matters and is caused by a lack of comprehension for the algorithmic recommendation process. Music can certainly be considered a subjective domain that is closely connected to personal tastes and emotions. Still, participants in this survey experiment were more favourable for recommendations from algorithms, thus exhibiting a behaviour opposite of what the notion of algorithm aversion suggests. In addition, participants in each experimental condition were given only limited information about the respective recommender, thus it is unlikely that participants receiving a recommendation from an algorithm had a significantly better insight into the reasoning behind the recommendation than participants receiving a recommendation from a music curator or a peer.

Another crucial insight of this research is that participants showed lower levels of preference for other people as recommenders, regardless of whether the involved recommender was a music curator, i.e. an expert in the field of music, or simply another user of a music streaming platform. Although Logg et al. (2019) have assumed expertise to be a highly influential factor that, if included in the research design, might challenge the applicability of algorithm appreciation, the findings of this survey experiment contradict this thought. While Thurman et al. (2018) have already challenged the idea that expert recommenders are preferred over algorithmic ones, this research confirms their conclusion in a conceptually different area. Thereby, it emphasises the robustness of algorithm appreciation, even when the factor of expertise is involved.

Additionally, the findings of this research provide some clarification on the relation between age and algorithm appreciation. In regard to the context of online news selection, higher age has been found to negatively impact the preference for algorithms as recommenders (Thurman et al., 2018). In this research however, the attitude towards the recommended song was not impacted by how old participants were, thereby confirming the results of Logg et al. (2019) and strengthening the applicability of their concept. This insight furthermore allows for the speculation that being used to consuming music mainly in a digital format, which can be assumed to be higher among younger people, did not affect how participants perceived the algorithm as recommender.

On a side note, the positive impact of algorithmic recommenders in this research pertains only to participants' attitudes, whereas their behavioural intentions in reaction to the recommendation were not affected by the recommender itself. Since accepting a recommendation is understood as a construct consisting of attitude and intended behaviour, the results of this experiment can only partially support claims about the recommender type predicting whether a recommendation is accepted or not. In other words, algorithmic recommenders generally turn out to be the preferred source for music recommendations as participants had more positive attitudes towards the respective recommendations, but recommendations from algorithms are not automatically more likely to be accepted.

Secondly, the impact of source credibility on participants' tendency to accept a music recommendation was scrutinised. As expected, perceived credibility – i.e. higher degrees of expertise and trustworthiness – positively impacted both participants' attitudes and intentions to interact with the recommended song. Hence, prior researchers' conclusions that perceived source credibility elicits positive attitudes towards a communicated message and facilitates the achievement of a communicative aim are confirmed by results of this research (Sternthal et al., 1978). For the considered case of music streaming platforms, it means that the perception of expertise and trustworthiness in a recommender will lead to a more positive attitude towards the recommendation and a higher chance that users will listen to the recommended song or interact with it in other ways. Therefore, it is concluded that source credibility positively affects the tendency of streaming platform users to accept a music recommendation.

Since the environment in which users of music streaming platforms receive recommendations is for the most part characterised by anonymity and unfamiliarity between the users and the sources of recommendations, it was assumed that these contextual factors could evoke a certain degree of scepticism among users towards the intentions of the recommender and the genuineness of the recommendations (Zhang et al., 2016). Thus, in order to avoid potential confounds and to gain a more nuanced insight into how the previously discussed relations come into effect, it was considered that the degree to which participants are sceptical towards online sources of recommendations, i.e. their level of eWOM scepticism, might impact the relation between the recommender type and the credibility that participants assign to it.

Overall, this assumption was not confirmed, but source credibility was found to be directly impacted by eWOM scepticism. More specifically, the detected effect was negative, meaning that higher levels of *eWOM scepticism* led to lower perceived *source credibility*. This effect occurred independently of the recommender type and is therefore considered to rather reflect participants' mistrust of the identity and motivation of sources of online recommendations and reviews in general. Presumably, eWOM scepticism originates from participants' past experiences which they project onto other sources of online reviews and recommendations they encounter (Zhang et al., 2016).

Both the type of recommender and the degree of perceived expertise and trustworthiness have been found to affect attitudes, with the two dimensions of source credibility also impacting intended behaviours in connection with the recommendation. These effects however occurred independently because *source credibility* was not directly influenced by *recommender type*. Although source credibility has been confirmed by previous studies as an influential concept, especially when evaluating eWOM messages such as reviews (Mahapatra & Mishra, 2017), the assumption that it explains the effect of the *recommender type* on *attitude* was not confirmed. While attitudes towards the recommended song differed depending on whether the recommendation was made by an algorithm, a music curator or a peer, the degree to which participants perceived the respective recommender as trustworthy and knowledgeable does not clarify why this difference among participants' attitudes exists.

5.1.2 Perceived personalisation as possible explanation

A critical reflection of the discussed findings leaves a number of questions open for discussion, especially about possible explanations for the effects between the recommender type and attitudes. Besides that, the experiment showed that algorithms are the most preferred recommenders but did not provide a proper reason for this outcome. Considering that age for example was not found to be influential, it is possible that the detected effect is connected to specific characteristics of algorithms and how these characteristics shape the perceptions of algorithms among music streaming users. Although no definitive answers can be given at this point, possible options shall be discussed in the following.

In view of the limited information about the recommenders that was given during the survey experiment, it is inferred that the way how participants evaluated them was largely influenced by preconceived ideas and by previous experiences they have made with this kind of technology (Bucher, 2017). Thus, it is reasonable to first reflect upon algorithms, their purpose and their functionality on streaming platforms and to assess afterwards which experiences could have influenced how users generally perceive algorithmic recommenders.

As outlined, one of the purposes of algorithms as music recommenders is to keep users engaged on music streaming platforms by providing them with guidance to find music that matches their taste (Seaver, 2018). To achieve this, the suggested songs need to be relevant for the users, hence recommendations on music streaming services are often personalised (De Keyzer et al., 2015; Morris & Powers, 2015). Broadly speaking, personalised music recommendations can be described as individualised suggestions that are in accord with a user's specific taste in music (Li, 2016). Generating these personalised recommendations is achieved by using complex algorithmic recommender systems to track and analyse the listening behaviour of users (Möller et al., 2018).

Considering this practice, it appears likely that the concept of personalisation might have had an underlying effect in this research: If participants associated algorithmic recommenders with personalised and thus more relevant music suggestions, it could explain why algorithms were preferred over the other recommenders (De Keyzer et al., 2015). As Senecal and Nantel (2004) have mentioned, any source providing personalised information is generally favoured over sources providing general information.

However, in this survey experiment, every participant – regardless of the assigned experimental condition - was made to believe that the displayed recommendation was individualised based on the musical preferences and listening habits indicated in the first questions. At first, it seems questionable how personalisation could then explain a significant preference for the recommender type algorithm. But, there is a conceptual difference between actual, or in this case alleged personalisation and the degree to which personalisation is perceived (De Keyzer et al., 2015). According to Li (2016), the positive impact of personalisation can only be observed when a message is perceived to be personalised, regardless of the effort that has been made to personalise it. Thus, participants assigned to the recommender type algorithm potentially have perceived a higher degree of personalisation than participants assigned to the music curator or the peer, even though the messages were made to appear personalised to an equal degree. The perceived degree of personalisation in turn is reflected in the attitude towards the recommendation. In practice, personalisation not only impacts how a recommender is perceived, but was also found to positively affect attitudes and purchase intentions, even when the respective messages exhibited rather low degrees of personalisation (Li, 2016).

In conclusion, perceived personalisation might provide an indication of why algorithms had a more positive effect on attitudes towards recommendations than music curators and peers in this survey experiment. This means that the degree to which the recommenders were perceived as making personalised recommendations could be a more useful explanation for the detected effects than how credible participants estimated the recommenders to be. Moreover, if not expertise but perceived personalisation is a driving factor for positive attitudes towards recommendations, it would help to explain why in the case of this research expert recommenders were not perceived significantly better than peers. Overall, this prompts the idea of a conceptual relation not between algorithm appreciation and source credibility, but between algorithm appreciation and perceived personalisation, which is supported by findings of prior research such as by Senecal & Nantel (2004): In their comparative study, recommender systems were found to be the most appreciated source of recommendations, despite being assigned lower levels of source credibility.

5.2 Conclusion and implications

The research question of this thesis asked in what ways different recommender types affect music streaming users' acceptance of music recommendations on streaming platforms. In regard to that question, it can be stated that whether users of music streaming platforms tend to accept a recommendation or not is affected by the source of the recommendation in certain ways. In comparison to music curators and peers, a clear preference for the recommender type algorithm was revealed, since algorithms caused more positive attitudes towards the recommended song among streaming users in this survey experiment. No differences between the attitudes towards recommendations made by music curators and by peers were observed. Moreover, acceptance of the recommendation - i.e. both attitudes and intended behaviour - was found to be affected by the users' perception of source credibility which was in turn negatively influenced by their level of scepticism towards sources of eWOM messages in general.

However, the described effects on acceptance of the recommendation occurred independently of each other in the conducted online survey experiment and therefore, the effects of the recommender type on the acceptance of a recommendation cannot be explained by the level of expertise and trustworthiness that streaming users assigned to each recommender. Perceived personalisation was identified as a possible alternative explanation for the detected effects.

By taking the example of users of music streaming platforms and their reactions to algorithmic recommendations, this research contributes to a broader framework that Logg et al. (2019) have described as 'theory of machine'. This emerging theoretical approach attempts to understand "mechanisms that shape how people expect human and algorithmic judgment, at their finest, to differ" (Logg et al., 2019, p. 100) and builds up on a large body of prior social science research investigating the ways how individuals make sense of the reasoning of others. Researchers aim to develop a similar understanding of these processes when people encounter technologies. The fast progressing development of machine learning and the subsequent prevalence of algorithms in many different areas and frequently used applications has increased the necessity for this kind of research in order to comprehend the social implications of artificial intelligence.

In order to enhance the current level of understanding, it is particularly important to investigate how lay people perceive algorithms, how they experience interactions with algorithms and how they evaluate the respective outcomes (Logg et al., 2019). Thus, the findings of this research not only provide evidence that confirms the notion of algorithm appreciation, but in a broader sense also illustrate that people tend to perceive algorithmic and human reasoning as inherently different, at least in the context of music recommendations. Thereby, the research conducted for this thesis helps to advance the level of knowledge that is sought for under the term 'theory of machine'.

Besides the contribution to this broad field of research, this thesis also sheds a light on the particular context in which it was conducted, namely how the consumption of music is affected by technology. With streaming becoming the most common way to consume music nowadays and recommendations being a highly important feature on the respective platforms, recommendations may play a crucial role in determining what music people listen to and how they discover new music. Considering this issue in view of the insights provided by this research, it means that algorithms as preferred recommendation sources could in the long term to a larger extent replace traditional intermediaries between cultural products and consumers, such as experts like radio DJs and music journalists (Morris, 2015).

Additionally, the findings of this research might also be interesting for professionals in the field and contribute to a further development of recommender systems. Developers and marketers of music streaming platforms could profit from a nuanced knowledge about their customers' opinions and reactions to different recommenders and could, for example, use this in their further work on both functionality and visibility of algorithmic recommenders.

Finally, this thesis complements the understanding about how recommendations are perceived, an issue that is commonly of interest in advertising and marketing research. It provides insights on various related theoretical concepts, such as source credibility and eWOM scepticism. It should be noted that prior studies on consumer reviews, brand attitudes and purchase intentions were applied to the context of this research. Therefore, a number of conclusions on the relations between these concepts can be inferred, for instance, how source credibility affects attitudes, while in turn being impacted by eWOM scepticism. Additionally, the conducted research not only exemplifies how marketing related concepts can be applied in other academic fields, but also how methodological considerations common in advertising and marketing research – such as the usage of vignettes and the establishing of context realism – can be applied in a conceptually different research and contribute to the achievement of findings.

5.3 Limitations

The research for this thesis was conducted after thoroughly assessing theoretical relations between the concepts under study and a multitude of methodological considerations were taken into account to construct a valid and reliable tool of measurement and to avoid confounding factors in the research design. Nevertheless, limitations are inevitable, and findings always have to be interpreted in regard to the specific constraints of a study. Hence, an overview over the most significant limitations shall be given here.

First and foremost, the usage of convenience and self-selection sampling in this research should be seen in a critical light. These sampling methods are generally not recommended to be used in quantitative research, particularly because they do not generate representative samples. Although random assignment ensures the comparability of the experimental conditions and thus compensates for some of the disadvantages of the sampling methods, generalisations of the findings of this research to the population of music streaming users should be made with care. There is a possibility that the detected effects do not apply to all streaming users or that they occur to a lesser extent than they have in this survey experiment.

It is presumed that the participants who took part in this survey experiment were motivated by their interest in music and streaming. This disposition in turn is related to the examined outcome variables, namely *attitude* and *intended behaviour* in reaction to recommendations. It is therefore possible that self-selection bias occurred (Olsen, 2011). For instance, participants in the sample might be more reflective about how they interact with recommendations than people who use music streaming more casually.

The sample furthermore was not balanced in regard to the demographic variable *gender*, as it included a significant higher share of male (60.6 %) than female (37.1 %) participants. No indication in literature or statistics was found showing that music streaming is used more by men than by women, thus this has to be identified as a bias in the sample. To account for a potential confounding influence of *gender* on the examined relations, the variable could have been included in the analysis as a covariate, following the example of De Keyzer et al. (2015).

Besides that, a number of possible limitations stem from the experimental design itself. As discussed previously, it is possible that based on their preconceived ideas of algorithms, participants might have perceived recommendations from the algorithmic recommender as more personalised than they perceived recommendations from music curators and peers. However, this notion is based on speculation, since it is not certain what the aforementioned preconceived ideas exactly are. This limitation could have been avoided by including a question in the survey experiment that provides an insight into this, comparable to what Logg et al. (2019) have done in one of their experiments. A similar approach could have been followed in regard to music curators, because it is possible that participants did not associate this term with experts in the field of music as was intended by the researcher.

Moreover, a number of aspects in the survey experiment had to be simplified due to practicality and the time constraints given for this master thesis. It is open to debate whether results have been impacted by that, yet it is reasonable to mention some of these aspects. For instance, intended behaviour was measured as a proxy of actual behaviours. Despite this being a common practice in research, intended behaviour does not exactly depict actual behaviour (De Keyzer et al., 2015). Therefore, a setting in which participants were given the option to actually perform different interactions with the recommendation might have provided richer and more authentic results.

Furthermore, while the survey experiment was designed with the aim of providing context realism and imitating the process of receiving a music recommendation on a streaming platform, it is not known with certainty to what degree these measures of increasing external validity were effective. Thus, a more elaborate visual design could have for example imitated the interface of a streaming platform. These suggestions however have to be balanced with the necessity to avoid confounding influences in an experiment.

5.4 Future research

Overall, the insights provided by this thesis emphasise two possible directions for future research. They can serve as a starting point to further explore the ways in which people perceive and encounter algorithms in their daily life on the one hand and to understand the behaviour of users in eWOM settings on the other hand.

Regarding the previous discussion on explanations for the observed effects, researchers are encouraged to investigate whether and to what degree perceived personalisation plays a role in the relation between different recommender types and people’s reactions to recommendations. Connected to that, it should also be explored if perceived personalisation correlates with positive perceptions of algorithmic recommenders and what the reasons for this potential correlation could be. Future research may also scrutinise which concepts could serve as potential mediating and moderating factors of the effects between recommender types and accepting recommendations and if established concepts such as personal relevance play a role in this, too (De Keyzer et al., 2015). By doing so, future research could use the discussion given in this work to contribute to the theoretical frameworks of algorithm appreciation and eWOM.

In addition, researchers may further investigate music consumption and the ways how music listeners discover new music on streaming platforms. Particular attention should be paid to how these practices are affected by the prevalence of algorithmic recommendations on the streaming platforms. A relevant issue could be whether there are constraints to the appreciation of algorithms as recommenders, for example because of concerns among the users about the tracking of their listening habits, about exposure to less diverse music or about receiving manipulated or biased recommendations. Studies on these kinds of questions may also draw from findings of research on news consumption, which has been dealing with the phenomenon of filter bubbles in relation to algorithmic news selection for quite some time (Thurman et al., 2018).

While music streaming platforms served as a case for this thesis to examine reactions to algorithmic recommenders, it is suggested for future studies to examine whether effects similar to the ones found in this research also occur in other contexts in which cultural products are consumed via online services and in which recommendations play a vital role, such as streaming services for audio-visual content. When doing so, future researchers could include context-specific factors – i.e. motivations for consumption, the purposes for using the respective platforms and the interest in receiving recommendations – in their analyses to achieve a more elaborate understanding. These factors were partly measured in this survey experiment, but not further considered in the analysis.

Summing up, the insights given in this thesis may provide a starting point for future research to thoroughly investigate both the question of how algorithms are perceived in various contexts, as well as the underlying mechanisms that influence this perception. Such studies would contribute to the understanding of algorithms and their social implications. This suggestion pertains to researchers from different disciplines, including media and communication studies but also sociology and cultural studies. By bringing together results from these fields, it is possible to advance a comprehensive and contemporary theory of how society and technologies like machine learning interact and how both sides mutually impact each other.

References

Arditi, D. (2017). Digital subscriptions: The unending consumption of music in the digital era. *Popular Music and Society*, 1–17. https://doi.org/10.1080/03007766.2016.1264101

Attaran, S., Notarantonio, E. M., & Quigley, C. J. (2015). Consumer perceptions of credibility and selling intent among advertisements, advertorials, and editorials: A persuasion knowledge model approach. *Journal of Promotion Management*, *21*(6), 703–720. https://doi.org/10.1080/10496491.2015.1088919

Bailey, J. E., & Pearson, S. W. (1983). Development of a tool for measuring and analyzing computer user satisfaction. *Management Science*, *29*(5), 530–545. https://doi.org/10.1287/mnsc.29.5.530

Bellman, S. (2017). Experimental design. In J. Matthes, C. S. Davis, & R. F. Potter (Eds.), *The International Encyclopedia of Communication Research Methods* (pp. 1–20). https://doi.org/10.1002/9781118901731.iecrm0098

Bockstedt, J. C., Kauffman, R. J., & Riggins, F. J. (2006). The move to artist-led on-line music distribution: A theory-based assessment and prospects for structural changes in the digital music market. *International Journal of Electronic Commerce*, *10*(3), 7–38. https://doi.org/10.2753/JEC1086-4415100301

Bucher, T. (2017). The algorithmic imaginary: Exploring the ordinary affects of Facebook algorithms. *Information, Communication & Society*, *20*(1), 30–44. https://doi.org/10.1080/1369118X.2016.1154086

Cacioppo, J. T., & Petty, R. E. (1984). The elaboration likelihood model of persuasion. *Advances in Consumer Research*, *11*(1), 673–675. Retrieved from https://search.ebscohost.com/login.aspx?direct=true&db=buh&AN=6434227&site=ehost-live&scope=site

Cheung, C. M. K., Lee, M. K. O., & Rabjohn, N. (2008). The impact of electronic word-of mouth: The adoption of online opinions in online customer communities. *Internet Research*, *18*(3), 229–247. https://doi.org/10.1108/10662240810883290

Ciocca, S. (2017, October). How does Spotify know you so well? *Medium*. Retrieved from https://medium.com/s/story/spotifys-discover-weekly-how-machine-learning-finds-your-new-music-19a41ab76efe

Crawford, K., & Calo, R. (2016). There is a blind spot in AI research. *Nature*, *538*(7625), 311–313. Retrieved from https://www.nature.com/news/there-is-a-blind-spot-in-ai-research-1.20805

Deahl, D. (2018, December). The Verge 2018 tech report card: Streaming music. *The Verge*. Retrieved from https://www.theverge.com/2018/12/31/18156503/2018-tech-recap-streaming-music-spotify-apple-soundcloud-tidal

De Keyzer, F., Dens, N., & De Pelsmacker, P. (2015). Is this for me? How consumers respond to personalized advertising on social network sites. *Journal of Interactive Advertising*, *15*(2), 124–134. https://doi.org/10.1080/15252019.2015.1082450

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, *144*(1), 114–126. https://doi.org/10.1037/xge0000033

Dörr, J., Benlian, A., Vetter, J., & Hess, T. (2010). Pricing of content services – an empirical investigation of music as a service. In M. L. Nelson, M. J. Shaw, & T. J. Strader (Eds.), *Sustainable e-Business Management* (pp. 13–24). Retrieved from http://www.citi.columbia.edu/B8210/read10/pricing%20of%20content%20services.pdf

Dörr, J., Wagner, T., Benlian, A., & Hess, T. (2013). Music as a service as an alternative to music piracy? An empirical investigation of the intention to use music streaming services. *Business & Information Systems Engineering*, *5*(6), 383–396. https://doi.org/10.1007/s12599-013-0294-0

Elson, M. (2017). Question wording and item formulation. In J. Matthes, C. S. Davis, & R. F. Potter (Eds.), *The International Encyclopedia of Communication Research Methods* (pp. 1–8). https://doi.org/10.1002/9781118901731.iecrm0200

Engelmann, I. (2017). Vignette research. In J. Matthes, C. S. Davis, & R. F. Potter (Eds.), *The International Encyclopedia of Communication Research Methods* (pp. 1–5). https://doi.org/10.1002/9781118901731.iecrm0263

Geuens, M., & De Pelsmacker, P. (2017). Planning and conducting experimental advertising research and questionnaire design. *Journal of Advertising*, *46*(1), 83–100. https://doi.org/10.1080/00913367.2016.1225233

Gibbs, A. (2017, September). Curating culture: The brains behind Apple Music's 'New Chicago' playlist. *Forbes Magazine*. Retrieved from https://www.forbes.com/sites/adriennegibbs/2017/09/15/curating-culture-the-brains-behind-apple-musics-new-chicago-playlist/#15112b4838e9

Gursoy, D., Spangenberg, E. R., & Rutherford, D. G. (2006). The hedonic and utilitarian dimensions of attendees' attitudes toward festivals. *Journal of Hospitality & Tourism Research*, *30*(3), 279–294. https://doi.org/10.1177/1096348006287162

Hagen, A. N. (2015). The playlist experience: Personal playlists in music streaming services. *Popular Music and Society*, *38*(5), 625–645. https://doi.org/10.1080/03007766.2015.1021174

Hansen, S. S., Lee, J. K., & Lee, S.-Y. (2014). Consumer-generated ads on Youtube: Impacts of source credibility and need for cognition on attitudes, interactive behaviors, and eWOM. *Journal of Electronic Commerce Research*, *15*(3), 254–266.

Retrieved from https://search-proquest-com.eur.idm.oclc.org/docview/1563701553?accountid=13598

Harvey, R. D., & Harvey, J. P. (2018). Random assignment. In B. B. Frey (Ed.), *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation* (pp. 1–3). https://doi.org/10.4135/9781506326139.n570

Hayes, A. F. (2015). An index and test of linear moderated mediation. *Multivariate Behavioral Research*, *50*(1), 1–22. https://doi.org/10.1080/00273171.2014.962683

Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (Second edition). New York: Guilford Press.

Hayes, A. F., & Preacher, K. J. (2014). Statistical mediation analysis with a multicategorical independent variable. *British Journal of Mathematical and Statistical Psychology*, *67*(3), 451–470. https://doi.org/10.1111/bmsp.12028

Hoewe, J. (2017). Manipulation check. In J. Matthes, C. S. Davis, & R. F. Potter (Eds.), *The International Encyclopedia of Communication Research Methods* (pp. 1–5). https://doi.org/10.1002/9781118901731.iecrm0135

Hovland, C. I., Janis, I. L., & Kelley, H. H. (1953). *Communication and persuasion. Psychological studies of opinion change.* New Haven and London: Yale University Press.

International Federation of the Phonographic Industry (IFPI). (2018). *Music consumer insight report 2018.* Retrieved from https://www.ifpi.org/consumer-research.php

Kjus, Y. (2016). Musical exploration via streaming services: The Norwegian experience. *Popular Communication*, *14*(3), 127–136. https://doi.org/10.1080/15405702.2016.1193183

Lee, M., & Youn, S. (2009). Electronic word of mouth (eWOM): How eWOM platforms influence consumer product judgement. *International Journal of Advertising*, *28*(3), 473–499. https://doi.org/10.2501/S0265048709200709

Li, C. (2016). When does web-based personalization really work? The distinction between actual personalization and perceived personalization. *Computers in Human Behavior*, *54*, 25–33. https://doi.org/10.1016/j.chb.2015.07.049

Logg, J. M. (2017, March 13). *Theory of machine: When do people rely on algorithms?* Retrieved from https://dash.harvard.edu/handle/1/31677474

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, *151*, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

Madden, T. J., Ellen, P. S., & Ajzen, I. (1992). A comparison of the theory of planned behavior and the theory of reasoned action. *Personality and Social Psychology Bulletin*, *18*(1), 3–9. https://doi.org/10.1177/0146167292181001

Mahapatra, S., & Mishra, A. (2017). Acceptance and forwarding of electronic word of mouth. *Marketing Intelligence & Planning*, *35*(5), 594–610. https://doi.org/10.1108/MIP-01-2017-0007

McGinnies, E., & Ward, C. D. (1980). Better liked than right: Trustworthiness and expertise as factors in credibility. *Personality and Social Psychology Bulletin*, *6*(3), 467–472. https://doi.org/10.1177/014616728063023

Möller, J., Trilling, D., Helberger, N., & van Es, B. (2018). Do not blame it on the algorithm: An empirical assessment of multiple recommender systems and their impact on content diversity. *Information, Communication & Society*, *21*(7), 959–977. https://doi.org/10.1080/1369118X.2018.1444076

Morris, J. W. (2015). Curation by code: Infomediaries and the data mining of taste. *European Journal of Cultural Studies*, *18*(4–5), 446–463. https://doi.org/10.1177/1367549415577387

Morris, J. W., & Powers, D. (2015). Control, curation and musical experience in streaming music services. *Creative Industries Journal*, *8*(2), 106–122. https://doi.org/10.1080/17510694.2015.1090222

Music Industry Blog. (2018, September 13). Mid-year 2018 streaming market shares [Blog post]. Retrieved from http://www.midiaresearch.com/blog/mid-year-2018-streaming-market-shares/

Netemeyer, R. G., Haws, K. L., & Bearden, W. O. (Eds.). (2011). *Handbook of marketing scales: Multi-item measures for marketing and consumer behavior research* (3rd ed). Los Angeles: SAGE.

Neuman, W. L. (2014). *Social research methods: Qualitative and quantitative approaches* (7. ed., Pearson new internat. ed). Harlow: Pearson.

Ohanian, R. (1990). Construction and validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attraction. *Journal of Advertising*, *19*(3), 39–52. Retrieved from https://search-proquest-com.eur.idm.oclc.org/docview/236621970?accountid=13598

Olsen, R. (2011). Self-selection bias. In P. J. Lavrakas (Ed.), *Encyclopedia of Suvey Research Methods* (p. 809). Retrieved from http://methods.sagepub.com/base/download/ReferenceEntry/encyclopedia-of-survey-research-methods/n526.xml

O'Reilly, D., Larsen, G., & Kubacki, K. (2013). *Music, markets and consumption*. Oxford: Goodfellow Publishers Ltd.

Pallant, J. (2005). *SPSS survival manual: A step by step guide to data analysis using SPSS for Windows (Version 12)*. Crows Nest, NSW, Australia: Allen & Unwin.

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Cambridge: Harvard University Press.

Privitera, G. J. (2015). *Statistics for the behavioral sciences* (Second edition). Los Angeles: SAGE.

Purnawirawan, N., De Pelsmacker, P., & Dens, N. (2012). Balance and sequence in online reviews: How perceived usefulness affects attitudes and intentions. *Journal of Interactive Marketing*, *26*(4), 244–255. https://doi.org/10.1016/j.intmar.2012.04.002

Recording Industry Association of America (RIAA). (2018). *RIAA 2018 year-end music industry revenue report.* Retrieved from https://www.riaa.com/reports/

Ruel, E., Wagner, W. E., III, & Gillespie, B. J. (2019). *The practice of survey research: Theory and applications* (pp. 101-119). https://doi.org/10.4135/9781483391700

Saldaña, M. (2017). Sampling, nonrandom. In J. Matthes, C. S. Davis, & R. F. Potter (Eds.), *The International Encyclopedia of Communication Research Methods* (pp. 1–5). https://doi.org/10.1002/9781118901731.iecrm0224

Schivinski, B., & Dabrowski, D. (2016). The effect of social media communication on consumer perceptions of brands. *Journal of Marketing Communications*, *22*(2), 189–214. https://doi.org/10.1080/13527266.2013.871323

Seaver, N. (2018). Captivating algorithms: Recommender systems as traps. *Journal of Material Culture*, 135918351882036. https://doi.org/10.1177/1359183518820366

Senecal, S., & Nantel, J. (2004). The influence of online product recommendations on consumers' online choices. *Journal of Retailing*, *80*(2), 159–169. https://doi.org/10.1016/j.jretai.2004.04.001

Singleton, M. (2018, June). Three years later, Tidal is still waiting for its big wave. *The Verge*. Retrieved from https://www.theverge.com/2018/6/28/17510390/tidal-streaming-service-subscribers-spotify-apple-music

Soll, J. B., & Larrick, R. P. (2009). Strategies for revising judgment: How (and how well) people use others' opinions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(3), 780–805. https://doi.org/10.1037/a0015145

Spotify. (2019). *Shareholder letter Q1 2019*. Retrieved from https://investors.spotify.com/financials/default.aspx

Sterba, S. K., & Foster, E. M. (2011). Self-selected sample. In P. J. Lavrakas (Ed.), *Encyclopedia of Survey Research Methods* (pp. 807–808). Retrieved from http://methods.sagepub.com/base/download/ReferenceEntry/encyclopedia-of-survey-research-methods/n526.xml

Sternthal, B., Dholakia, R., & Leavitt, C. (1978). The persuasive effect of source credibility: Tests of cognitive response. *Journal of Consumer Research*, *4*, 252–260. https://doi.org/10.1086/208704

Thurman, N., Moeller, J., Helberger, N., & Trilling, D. (2018). My Friends, Editors, Algorithms, and I: Examining audience attitudes to news selection. *Digital Journalism*, 1–23. https://doi.org/10.1080/21670811.2018.1493936

Waterfield, J. (2018). Convenience sampling. In B. B. Frey (Ed.), *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation* (pp. 1–3). https://doi.org/10.4135/9781506326139.n155

Webb, N. G. (2018). Variables, control. In M. Allen (Ed.), *The SAGE Encyclopedia of Communication Research Methods* (pp. 1–4). https://doi.org/10.4135/9781483381411

White, H. A., & Andsager, J. L. (2017). Third-person effect: Basic concept. In P. Rössler, C. A. Hoffner, & L. van Zoonen (Eds.), *The International Encyclopedia of Media Effects* (pp. 1–10). https://doi.org/10.1002/9781118783764.wbieme0051

White, T. B. (2005). Consumer trust and advice acceptance: The moderating roles of benevolence, expertise, and negative emotions. *Journal of Consumer Psychology*, *15*(2), 141–148. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1207/s15327663jcp1502_6?casa_token=M7KT-mwvLRcAAAAA:vfZkpq03du-VnmJwkmdgKaxmHgeID-exIf-GFTfxQ5JY8O91amaAK0KSwpYsMVkM8aHYLgO5sye39Fo

Wikström, P. (2013). *The music industry: Music in the cloud* (2nd edition). Cambridge, UK: Polity Press.

Wojcicki, S. (2017, June 22). Updates from VidCon: more users, more products, more shows and much more [Blog post]. Retrieved from https://youtube.googleblog.com/2017/06/updates-from-vidcon-more-users-more.html#gpluscomments

Wrench, J. S. (2017). Quantitative methodology. In J. Matthes, C. S. Davis, & R. F. Potter (Eds.), *The International Encyclopedia of Communication Research Methods* (pp. 1–10). https://doi.org/10.1002/9781118901731.iecrm0197

Yan, T. (2008). Nondifferentiation. In P. Lavrakas (Ed.), *Encyclopedia of Survey Research Methods* (pp. 1–3). https://doi.org/10.4135/9781412963947

Yeomans, M., Shah, A., Mullainathan, S., & Kleinberg, J. (2019). Making sense of recommendations. *Journal of Behavioral Decision Making*. https://doi.org/10.1002/bdm.2118

Zhang, X. (Jerry), Ko, M., & Carpenter, D. (2016). Development of a scale to measure skepticism toward electronic word-of-mouth. *Computers in Human Behavior*, *56*, 198–208. https://doi.org/10.1016/j.chb.2015.11.042

APPENDIX A

Fabricated album covers used in the pre-test of the survey experiment

Figure A1. Album Cover *Between Planets* by fictional artist *Blair*.

Figure A2. Album Cover *Dogma Beware* by fictional artist *Boots & Laces*.

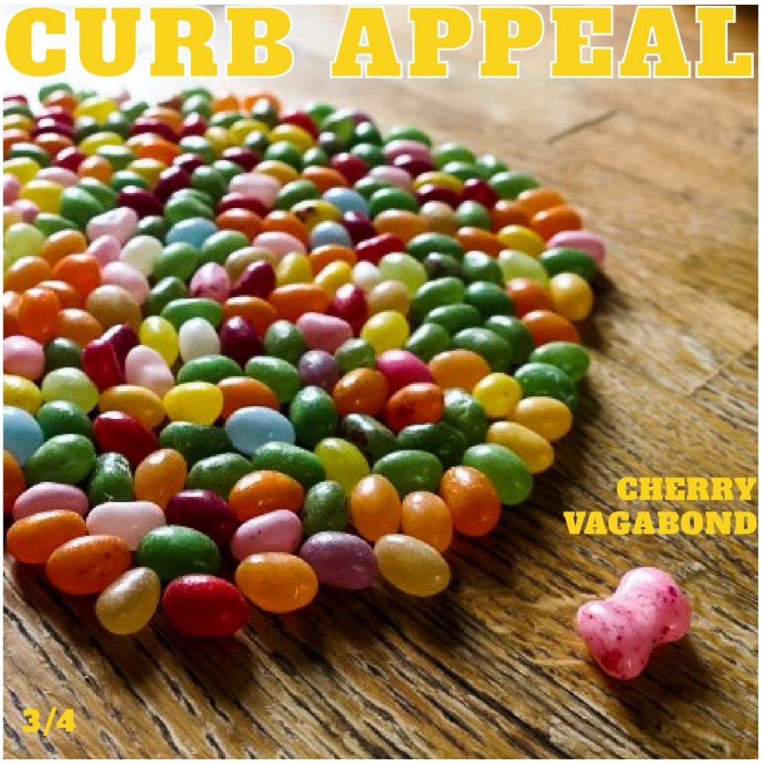

Figure A3. Album Cover *Cherry Vagabond* by fictional artist *Curb Appeal*.

Figure A4. Album Cover *Absent Shores* by fictional artist *Dream of Abyss*.

Figure A5. Album Cover *Decisions, Decisions* by fictional artist *Inescapable Immortality.*

Figure A6. Album Cover *Hell & Hope* by fictional artist *Pawn Panic.*

Figure A7. Album Cover *Night in the Ship* by fictional artist *Place Stamp Here*.

Figure A8. Album Cover *Space* by fictional artist *The Sockets*.

Figure A9. Album Cover *Academy & the Child* by fictional artist *Tipsy Jones*.

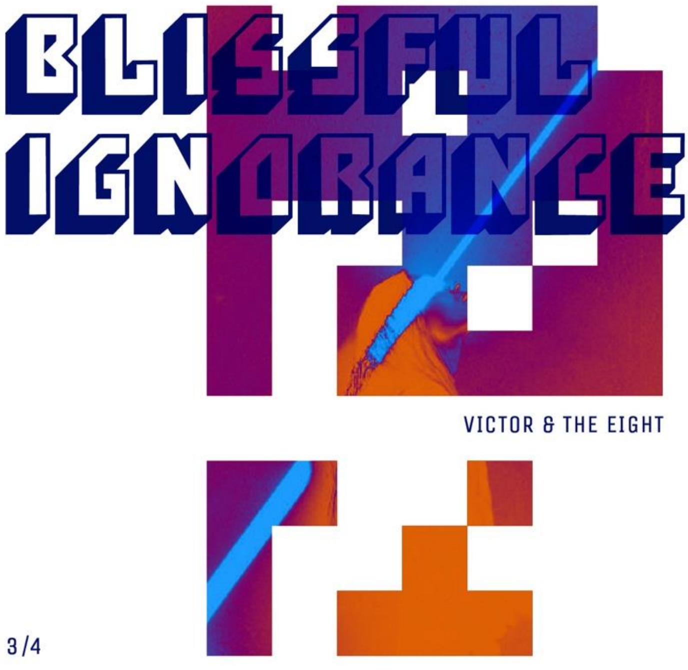

Figure A10. Album Cover *Blissful Ignorance* by fictional artist *Victor & the Eight*.

APPENDIX B

Outcome of the pre-test: familiarity and appeal of the three fabricated album covers that were considered to be used in the survey experiment as well as likeliness to interact with them

Appendix B1. Outcome for *Dogma Beware* by fictional artist *Boots & Laces*

(*N* = 26)

	s^2	*M*	*SD*
Familiarity with the artist	.49	4.58	.70
Likeliness to listen to a song from this album based on the cover	.87	3.08	.94
Likeliness to look up the artist based on the cover	.78	3.12	.88
Likeliness to add a song from this album to a playlist based on the cover	.81	3.32	.90
The album cover matches with musical taste	.83	3.12	.91
The album cover looks appealing	1.15	3.23	1.07

Appendix B2. Outcome for *Space* by fictional artist *The Sockets*

(*N* = 26)

	s^2	*M*	*SD*
Familiarity with the artist	.57	4.62	.75
Likeliness to listen to a song from this album based on the cover	1.04	2.81	1.02
Likeliness to look up the artist based on the cover	.99	3.12	.99
Likeliness to add a song from this album to a playlist based on the cover	.99	3.42	.99
The album cover matches with musical taste	.93	3.15	.93
The album cover looks appealing	1.20	2.92	1.20

Appendix B3. Outcome for *Blissful Ignorance* by fictional artist *Victor & the Eight* (*N* = 26)

	s^2	*M*	*SD*
Familiarity with the artist	.40	4.65	.63
Likeliness to listen to a song from this album based on the cover	1.18	3.31	1.09
Likeliness to look up the artist based on the cover	1.22	3.50	1.11
Likeliness to add a song from this album to a playlist based on the cover	1.13	3.62	1.06
The album cover matches with musical taste	1.13	3.58	1.07
The album cover looks appealing	1.22	3.50	1.11

APPENDIX C

Questionnaire used to conduct the online survey experiment

Title: Music Consumption on Streaming Platforms

Introduction:

Thank you for participating in this survey, your responses are very important to me. I am studying music listening habits of users of music streaming platforms for my master's thesis in Media, Culture and Society at Erasmus University Rotterdam. Specifically, I am interested in ways of discovering new music on streaming platforms.

Please answer the questions intuitively and in regard to your own experiences and opinion, there are no correct or incorrect answers. Completing the survey will take no longer than 10 minutes. I assure that all of your answers are recorded and processed anonymously, kept under confidentiality and will only be used for the purpose of my research.

Your participation is very valuable for my thesis, so thank you very much again for taking the time to fill in this survey. With best regards, Ina Weber

Q1 By checking this box you confirm that you have read the information above and want to take part in this survey. Note that you are free to abandon the survey at any time.

- I confirm!

Q2 As a start, please select the music streaming platform(s) you currently use from the list below (multiple answers possible).

- Amazon Music
- Apple Music
- Deezer
- Google Play Music
- Pandora
- Qobuz
- Soundcloud
- Spotify
- Tidal
- YouTube
- Other, please specify: ________________
- I do not use any music streaming platforms.

Q3 Please state the approximate number of hours you spend per week listening to music on streaming platform(s). ________________

Q4 Do you have a paid subscription for the music streaming platform that you use the most?

- Yes
- No

<u>Instruction</u>:

Please imagine the following situation:

You want to try out another music streaming platform for which you just have created a new account. In order to receive music recommendations on this streaming platform that match your taste, you are asked a few questions about your taste in music and your listening habits. Based on your answers, a song will be picked for you.

Q5 Please select up to 3 of your favourite music genres from the list below.

- Country
- Dance
- Electronic
- Folk & Blues
- HipHop / Rap
- Indie / Alternative
- Jazz
- Latin
- Metal
- Pop
- Reggae
- R'n'B / Soul
- Rock
- Singer – Songwriter
- World Music
- Other, please specify: ________________

Q6 In which situations do you most often listen to music? Please state how much you agree with the following statements.

I listen to music ...

	Strongly disagree (1)	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree (7)
... with my full attention and I don't do anything else on the side.	○	○	○	○	○	○	○
... while doing sport.	○	○	○	○	○	○	○
... on my way to work / university / school.	○	○	○	○	○	○	○
... to relax.	○	○	○	○	○	○	○
... when I'm with friends.	○	○	○	○	○	○	○
... to concentrate on the work I am doing.	○	○	○	○	○	○	○
... just to have something playing in the background.	○	○	○	○	○	○	○
... while doing housework like cooking or cleaning.	○	○	○	○	○	○	○
... to fall asleep.	○	○	○	○	○	○	○

Q7 For which purposes do you use music streaming platforms? Please state how much you agree with the following statements.

I use music streaming platforms to ...

	Strongly disagree (1)	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree (7)
... create my own playlists.	○	○	○	○	○	○	○
... have access to music from a big variety of artists.	○	○	○	○	○	○	○
... receive music recommendations.	○	○	○	○	○	○	○
... to see what music others are listening to.	○	○	○	○	○	○	○
... listen to preselected playlists tailored for different contexts, moods or genres.	○	○	○	○	○	○	○
... find information about different artists.	○	○	○	○	○	○	○
... listen to the playlists of my friends or other users.	○	○	○	○	○	○	○
... discover new music.	○	○	○	○	○	○	○

<u>Instruction:</u>

The data about your musical preferences and listening habits has been recorded. Based on your answers, a song on the new music streaming platform will be picked for you. Please wait a moment until the process continues.

Presentation of the stimuli:

An **algorithm** has selected this song for you
The new music streaming platform uses **algorithms** to make music recommendations for its users. Based on your answers from earlier, the algorithm has selected the song *Dogma Beware* by an upcoming artist called *Boots and Laces* as likely to match your taste.
The algorithm is programmed to process user data such as preferences and listening habits. In combination with the analysis it makes of different songs, the algorithm is able to make predictions about the users' musical preferences. Based on that, it can recommend songs to the users which are likely to appeal to them.

A **music curator** has selected this song for you
The new music streaming platform works together with **music curators** to make music recommendations for its users. Based on your answers from earlier, a music curator has selected the song *Dogma Beware* by an upcoming artist called *Boots and Laces* as likely to match your taste.
As part of their job, music curators select music and compile playlists for specific occasions. They are familiar with different genres and artists which enables them to make predictions about the users' musical preferences. Based on that, they can recommend songs to the users which are likely to appeal to them.

Another streaming **user** has selected this song for you
The new music streaming platform makes it possible to receive recommendations from **other users** who have a similar taste in music. Based on your answers from earlier, another user has selected the song *Dogma Beware* by an upcoming artist called *Boots and Laces* as likely to match your taste.
By creating public playlists or by sharing songs, users of streaming platforms can engage in different practices of recommending music to others. As frequent listeners themselves, they are able to make predictions about others' musical preferences. Based on that they can recommend songs to others which are likely to appeal to them.

Instruction: In the following, please keep the recommendation you just saw in mind and tell me your opinion about it based on your first impression.

Q8 (condition peer) Please indicate how likely you are to engage with the recommended song by stating how much you agree with the following statements. I would ...

	Extremely unlikely (1)	Unlikely	Somewhat unlikely	Neither likely nor unlikely	Somewhat likely	Likely	Extremely likely (7)
... listen to this song.	○	○	○	○	○	○	○
... add this song to one of my playlists.	○	○	○	○	○	○	○
... look up the page of the artist who performs this song.	○	○	○	○	○	○	○
... skip / ignore this song.	○	○	○	○	○	○	○
... listen to the album of this song.	○	○	○	○	○	○	○
... search for similar songs.	○	○	○	○	○	○	○

Q9 (condition peer) Please give me your general opinion about the recommendation by stating how much you agree with the following statements.

	Strongly disagree (1)	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree (7)
Recommendations from other users of streaming platforms are valuable.	○	○	○	○	○	○	○
Recommendations from other users of streaming platforms provide me with information about	○	○	○	○	○	○	○

what music to listen to.							
Recommendations from other users of streaming platforms are helpful for me when I'm looking for new artists/songs.	○	○	○	○	○	○	○

Q10 (condition peer) Please indicate how you think other users generally react to this kind of recommendation by stating how much you agree with the following statements.
I have the impression that other music streaming users ...

	Strongly disagree (1)	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree (7)
... are satisfied with recommendations from other users.	○	○	○	○	○	○	○
... find recommendations from other users to be matching their taste.	○	○	○	○	○	○	○
... would appreciate receiving recommendations from other users.	○	○	○	○	○	○	○

Instruction: Now think again about the recommendation you saw and about the process that was carried out to select this recommendation for you.

Q11 (condition peer) Please indicate how much you agree with the following statements.
I think other users of streaming platforms ...

	Strongly disagree (1)	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree (7)
... are being honest when making music recommendations.	○	○	○	○	○	○	○
... are reliable when it comes to music.	○	○	○	○	○	○	○
... are sincere about music.	○	○	○	○	○	○	○
... are trustworthy in questions about music.	○	○	○	○	○	○	○
... are experienced with music.	○	○	○	○	○	○	○
... are knowledgeable about different songs and artists.	○	○	○	○	○	○	○
... are qualified to make music recommendations.	○	○	○	○	○	○	○
... are biased towards certain songs and artists.	○	○	○	○	○	○	○

Q8 (condition music curator) Please indicate how likely you are to engage with the recommended song by stating how much you agree with the following statements. I would ...

	Extremely unlikely (1)	Unlikely	Somewhat unlikely	Neither likely nor unlikely	Somewhat likely	Likely	Extremely likely (7)
... listen to this song.	○	○	○	○	○	○	○
... add this song to one of my playlists.	○	○	○	○	○	○	○
... look up the page of the artist who performs this song.	○	○	○	○	○	○	○
... skip / ignore this song.	○	○	○	○	○	○	○
... listen to the album of this song.	○	○	○	○	○	○	○
... search for similar songs.	○	○	○	○	○	○	○

Q9 (condition music curator) Please give me your general opinion about the recommendation by stating how much you agree with the following statements.

	Strongly disagree (1)	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree (7)
Recommendations from music curators are valuable.	○	○	○	○	○	○	○
Recommendations from music curators provide me with information about	○	○	○	○	○	○	○

what music to listen to.							
Recommendations from music curators are helpful for me when I'm looking for new artists/songs.	○	○	○	○	○	○	○

Q10 (condition music curator) Please indicate how you think other users generally react to this kind of recommendation by stating how much you agree with the following statements.
I have the impression that other music streaming users ...

	Strongly disagree (1)	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree (7)
... are satisfied with recommendations from music curators.	○	○	○	○	○	○	○
... find recommendations from music curators to be matching their taste.	○	○	○	○	○	○	○
... would appreciate receiving recommendations from music curators.	○	○	○	○	○	○	○

<u>Instruction:</u> Now think again about the recommendation you saw and about the process that was carried out to select this recommendation for you.

Q11 (condition music curator) Please indicate how much you agree with the following statements.
I think music curators ...

	Strongly disagree (1)	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree (7)
... are being honest when making music recommendations.	○	○	○	○	○	○	○
... are reliable when it comes to music.	○	○	○	○	○	○	○
... are sincere about music.	○	○	○	○	○	○	○
... are trustworthy in questions about music.	○	○	○	○	○	○	○
... are experienced with music.	○	○	○	○	○	○	○
... are knowledgeable about different songs and artists.	○	○	○	○	○	○	○
... are qualified to make music recommendations.	○	○	○	○	○	○	○
... are biased towards certain songs and artists.	○	○	○	○	○	○	○

Q8 (condition algorithm) Please indicate how likely you are to engage with the recommended song by stating how much you agree with the following statements. I would ...

	Extremely unlikely (1)	Unlikely	Somewhat unlikely	Neither likely nor unlikely	Somewhat likely	Likely	Extremely likely (7)
... listen to this song.	○	○	○	○	○	○	○

... add this song to one of my playlists.	○	○	○	○	○	○	○
... look up the page of the artist who performs this song.	○	○	○	○	○	○	○
... skip / ignore this song.	○	○	○	○	○	○	○
... listen to the album of this song.	○	○	○	○	○	○	○
... search for similar songs.	○	○	○	○	○	○	○

Q9 (condition algorithm) Please give me your general opinion about the recommendation by stating how much you agree with the following statements.

	Strongly disagree (1)	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree (7)
Recommendations from algorithms are valuable.	○	○	○	○	○	○	○
Recommendations from algorithms provide me with information about what music to listen to.	○	○	○	○	○	○	○
Recommendations from algorithms are helpful for me when I'm looking for new artists/songs.	○	○	○	○	○	○	○

Q10 (condition algorithm) Please indicate how you think other users generally react to this kind of recommendation by stating how much you agree with the following statements.
I have the impression that other music streaming users ...

	Strongly disagree (1)	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree (7)
... are satisfied with recommendations from algorithms.	○	○	○	○	○	○	○
... find recommendations from algorithms to be matching their taste.	○	○	○	○	○	○	○
... would appreciate receiving recommendations from algorithms.	○	○	○	○	○	○	○

<u>Instruction:</u> Now think again about the recommendation you saw and about the process that was carried out to select this recommendation for you.

Q11 (condition algorithm) Please indicate how much you agree with the following statements.
I think algorithms ...

	Strongly disagree (1)	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree (7)
... are being honest when making music recommendations.	○	○	○	○	○	○	○
... are reliable when it comes to music.	○	○	○	○	○	○	○

... are sincere about music.	○	○	○	○	○	○	○
... are trustworthy in questions about music.	○	○	○	○	○	○	○
... are experienced with music.	○	○	○	○	○	○	○
... are knowledgeable about different songs and artists.	○	○	○	○	○	○	○
... are qualified to make music recommendations.	○	○	○	○	○	○	○
... are biased towards certain songs and artists.	○	○	○	○	○	○	○

<u>Instruction:</u> Now in the last part of the survey, we are leaving the new streaming platform behind and I would like to know more about your general opinion on online recommendations.

Q12 Please indicate how much you agree with the following statements.

	Strongly disagree (1)	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree (7)
We can hardly depend on getting the truth from most online recommendations.	○	○	○	○	○	○	○
Online recommendations are not generally truthful.	○	○	○	○	○	○	○
In general, online recommendations don't reflect the true picture of a subject.	○	○	○	○	○	○	○

Online recommendations care more about getting people to buy things.	○	○	○	○	○	○	○
Most online recommendations are intended to mislead.	○	○	○	○	○	○	○
The source of an online recommendation might be up to something.	○	○	○	○	○	○	○
The source of an online recommendations is not necessarily who they appear to be.	○	○	○	○	○	○	○
Often the same recommendation is being made by the same source but under different names.	○	○	○	○	○	○	○

Q13 When you think back to the beginning of the survey, do you still remember what the recommended song was?

- Yes
- No

Q14 If you do, please state the title of the song and/or the name of the artist below.

____

Q15 When you think back to the beginning of the survey, do you still remember who recommended the song to you?

- Yes
- No

Q16 If you do, please state below who recommended the song to you.

____

Q17 Please state the year you were born:

____

Q18 What is your gender?

- Male
- Female
- Other
- Prefer not to say

Q19 What is the highest level of education that you have obtained?

- Less than high school
- High school graduate
- Professional degree
- University / University of Applied Sciences Bachelor's degree
- University / University of Applied Sciences Master's degree
- Doctorate

Q20 In which country do you currently reside?
▼ Afghanistan ... Zimbabwe

Q21 Last question: while answering the survey, did you search for additional information on the artist and/or song that was recommended to you?

- Yes
- No

Debriefing:

Your answers have been recorded, thank you very much for filling out this survey!
Before you exit the survey, I would like to give you some background information on the study and clarify some things.
For my master thesis, I am researching how people react to different music recommendations. What you saw earlier was not an actual recommendation for you made by an algorithm, a curator or a peer. Both the artist and the song as well as the album cover have been completely made up for this study. Any resemblances to existing artists and/or songs are accidental and were not intended. In order to not influence you and the answers you give, I had to make up this story and could not inform you about the true purpose of the survey earlier. Therefore, if you pass this survey on to other people, please do not share this information with them.
I hope this message clarified any potential confusion that might have come up on your side during the survey. If you have any further questions, comments or complaints, please do not

hesitate to get in touch with me (481707iw@students.eur.nl) or my supervisor at Erasmus University, Dr. Joao Ferreira Goncalves (ferreiragoncalves@eshcc.eur.nl).

You can close the browser window now.

Best Regards, Ina Weber

APPENDIX D

Effect of *recommender type* on *attitude* and *intended behaviour* mediated by *source credibility* at different levels of *eWOM scepticism*

Appendix D1. Relative conditional indirect effect *recommender type* on *attitude*

Recommender Type	Low eWOM scepticism (3.50)	Moderate eWOM scepticism (4.44)	High eWOM scepticism (5.33)
X1	.24	.37*	.49*
X2	.002	.05	.10

Notes:
X1 refers to the effect of music curator in comparison to the reference group (algorithm)
X2 refers to the effect of peer in comparison to the reference group (algorithm)
Levels of eWOM scepticism refer to 16^{th}, 50^{th} and 84^{th} percentiles
Significance: *$p < 0.05$ (95 % level of confidence for all confidence intervals)

Appendix D2. Relative conditional indirect effect *recommender type* on *intended behaviour*

Recommender Type	Low eWOM scepticism (3.50)	Moderate eWOM scepticism (4.44)	High eWOM scepticism (5.33)
X1	.12	.19*	.24*
X2	.001	.03	.05

Notes:
X1 refers to the effect of music curator in comparison to the reference group (algorithm)
X2 refers to the effect of peer in comparison to the reference group (algorithm)
Levels of eWOM scepticism refer to 16^{th}, 50^{th} and 84^{th} percentiles
Significance: *$p < 0.05$ (95 % level of confidence for all confidence intervals)