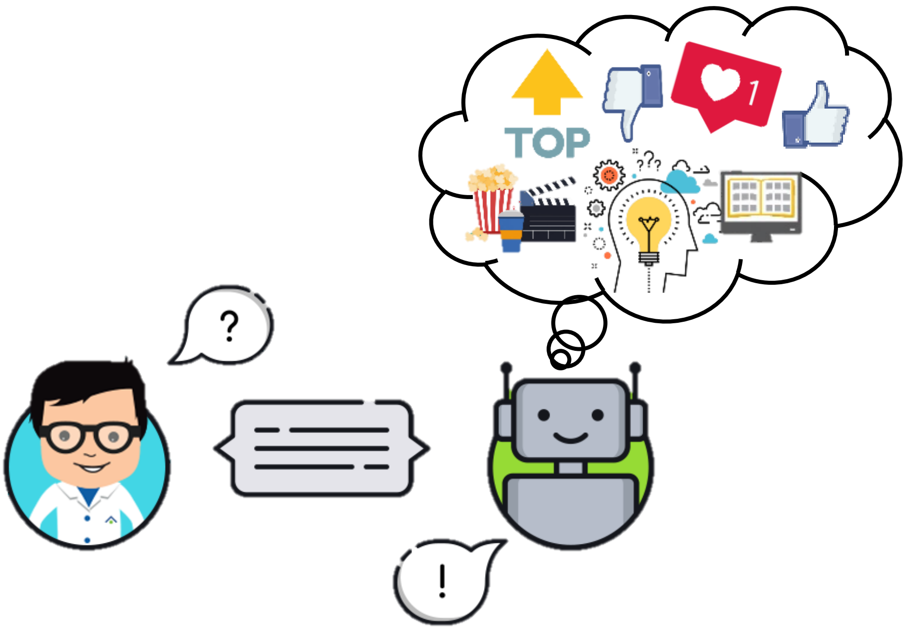
Francesca Feliciotti

PERSONALIZED CONVERSATIONAL RECOMMENDER SYSTEM IN A MOVIE PLATFORM: THE IMPACT ON USER SATISFACTION



TU Delft

Personalized conversational recommender system in a movie platform: the impact on user satisfaction

By

Francesca Feliciotti

Student Number: 4737695

in partial fulfilment of the requirements for the degree of

Master of Science
in Management of Technology

at the Delft University of Technology,
to be defended publicly on September 26th, 2019

Chairman: Dr. Stephan Lukosch, Section Systems Engineering and Services
First Supervisor: Dr. Laurens Rook, Section Economics of Technology and Innovation
Company Supervisor: Dennis de Vries, Cognitive Engagement, Deloitte Amsterdam

Acknowledgments

This work represents the conclusion of my graduation project of the MSc Management of Technology at the Delft University of Technology. I am grateful for all the choices, experiences and support that led me to achieve this milestone.

I would like to thank my supervisor Dr Laurens Rook for his thoughtful insights and his professional guidance. Although this thesis has been an individual work, Lauren’s continuous and valuable support and enthusiastic encouragement never made me feel I was alone during the project.

I would also like to thank my Chairman Dr Stephan Lukosch for his critical feedback, that proved to be of great help.

I would like to thank my colleagues of the Cognitive Engagement team for giving me the chance to conduct my thesis at Deloitte. In particular, I would like to single out my supervisor Dennis de Vries for his great support and for providing me with an insight into the world of technology consulting.

Furthermore, I would like to thank my family, who always encouraged me and was there for me despite the miles away. To my parents, Paolo and Nadia, for their advice that I always pretend not to listen to but which guide me when making my choices. To my brother Stefano because we're each other's biggest fans. With the hope that you can begin to believe in yourself, that you dare to keep all your dreams alive and that you never stop fighting to make them come true.

Abstract

During the last years, as virtual assistants such as Siri (Apple), Google Assistant, Amazon Alexa, spread into everyday situations, conversational recommender systems were proposed as an interactive recommendation process to connect with the user. However, there is little knowledge about the personalization of conversational recommender systems as a way to increase the satisfaction of the users. The current research focuses on the users' experience with a movie platform. It argues that users satisfaction can only be improved if the conversational recommender system knows the preferences and the Openness to Experience of the users, and therefore gives custom-made recommendations. Specifically, the relationship between the degree of Openness to Experience and serendipitous or accurate recommendations is investigated. The present study demonstrates that overall people low on Openness to Experience significantly prefer accurate suggestions rather than serendipitous ones. Instead, people high on Openness to Experience do not have a significant preference regarding accurate or serendipitous recommendations. Furthermore, to explore whether a conversational recommender system increases user satisfaction, this study explores to what extent the satisfaction of users in recommender system is dependent on (traditional vs. conversational) recommender system mode of interaction. Results indicate that conversational recommender system exerts a positive impact on user satisfaction.

Contents

1. Introduction

1.1 Background

Since Alan Turing's (1950) renowned question "Can machines think?", conversational understanding is the way through which it is possible to evaluate the intelligence of a machine. In the 1960s, Weizenbaum invented ELIZA, a natural language processing computer program developed to study natural language conversation with a computer (Brandtzaeg & Følstad 2017). In the early 1980s, inspired by ELIZA Richard Wallace, created ALICE, that led to the advance of Artificial Intelligence Markup Language (AIML). AIML is an Extensible Markup Language dialect used to link words and expressions sent by the user with topic categories (Radziwill & Benton, 2017). Progressing from the deterministic responses of primitive systems such as ELIZA and ALICE, present day systems use advanced rules-based pattern matching algorithms to interact with users (Radziwill & Benton, 2017). In particular, today's algorithms perform advanced activities that require the ability to 'learn by doing' through the use of machine learning algorithms. Recent developments in machine learning and artificial intelligence are stimulating the interest for human-like autonomous agents such as conversational agents (or: chatbots). Currently, such artificial intelligence applications are entering the daily life of users (Lortie & Guitton, 2011).

Specifically, the most famous Internet companies such as Microsoft, Amazon, Facebook, and Google have created conversational software agents, known as chatbots, to assist their customers. Chatbots are virtual agents that can provide services and data through natural language user interfaces (Brandtzaeg & Følstad, 2017). These conversational software agents are able to execute tasks and provide conversational responses based on natural language inputs (Radziwill & Benton, 2017). Chatbots provide a whole series of services that not only simplify the lives of users but also improve the productivity of companies, through the use of increasingly advanced technologies.

As virtual assistants such as Siri (Apple), Google Assistant, Amazon Alexa, spread into everyday situations (Sun & Zhang, 2018), the interest in creating a more human-aware chatbot is becoming crucial. According to Brandtzaeg et al. (2017), developers and designers indispensably need to understand how people react to the chatbot experience and what stimulates their future usage. However, despite the initial optimism concerning the launch of chatbots by Facebook and Microsoft, their use is less widespread than expected. This might be, because that most available chatbots come up short to fill users' needs due to nonsensical replies, unclear purposes, or fail to relate personally to the people they are interacting with (Brandtzaeg & Følstad, 2017). Hence, it is necessary to investigate how chatbots tailored to users' personality can enhance their experience and influence the development of their desires.

The current research focuses on the users' experience, arguing that it can only be improved if the chatbot is adapted to the preferences and interests of the users and, therefore, give personalized recommendations. Thus, integrating machine-learning recommendation techniques into conversational systems and adapting the recommendations to the personality of the users could guarantee the success of the chatbots. This technology, known as conversational recommender system, is still emergent. The ways to personalize conversational recommender systems based on individual differences, and how this may influence user satisfaction have not been studied yet, and are the focus of this study. Furthermore, it has not yet been explored whether conversational recommender systems increase user satisfaction or whether the user prefers to adopt a traditional recommender system.

1.2 Problem definition

Today there is a growing attention for online personalization of recommender system. In fact, with all the hype in personalization marketing, existing studies have mostly focused on using real-time data and explicit customer inputs to provide personalized messages (Looney, Jacobson & Redding, 2011). In fact, the explosive growth of e-commerce and multimedia platforms was followed by the need of furnishing recommendations coming from the accurate selection of an entire range of possible alternatives (Ricci, Rokach, & Shapira, 2015). Recommendations have to be personalized or adjusted according to the customer's tastes and interests, with different people usually receiving different item suggestions (Jannach, Resnick, Tuzhilin & Zanker, 2016). That implies knowing customers' personality and collecting user history and interests. In this respect, virtual assistants for customer service have turned out to be more and more popular business tools to engage with a target audience and to grow profitably and efficiently (Kurilchik, 2017).

The chatbot is one of the most direct and important user-oriented methods to interact with customers, collect their data and achieve personalization. The main function of this technology is to communicate with human beings through text messages and to be able to deeply comprehend the conversation in order to respond adequately (Peters, 2018). Chatbots' main task these days is to work in customer service, client support, information and entertainment (Brandtzaeg & Følstad, 2017). Since chatbots are not limited by time, constraints, and fatigue, they can offer 24/7 support to the consumer. As a consequence, they can potentially ameliorate the customer service provided by companies (de Haan, 2018). However, the main challenge is to implement a chatbot able to interact with different users according to their different personalities and interests. A conversational recommender system would be able to accomplish that: *it would use different approaches to deal with the different conversational styles of the clients.*

Recently, there has been a growing search for attention in more user-oriented methods, as a way to collect consumer's data and achieve personalization (Tkalcic & Chen, 2015). However, despite the attention that literature has given to personalization, only a few works have perceived that using consumers' preferences and personalities is a way to accomplish that. Up to now, the study on chatbots has been basically centered on algorithms for moving forward the accuracy of the answers, while the personalization of the chatbot based on user's personality and interests has not been widely explored. Some "offline experiments" have been conducted about the latter (Chen, Wu & He, 2013), but the relationship between personality factors and users' preferences in recommendations is still at its infancy. Filling this research gap helps to comprehend if the impact of a personality-based conversational recommender system can improve the users satisfaction.

1.3 Knowledge gap

In recent years, the emergence of chatbots has given the incentive to research on conversational recommender agents. These interactive assistants are meant to assist users in finding the candidate items that match their tastes (Sun & Zhang, 2018).

Christakopoulou et al. (2016) proposed a preference elicitation framework to understand what to ask to quickly learn new users' preferences. However, the afore-mentioned model has a very limited action space. It focuses on the cold start[1] without aiming to capture the longer-term dependencies and re-enforce user satisfaction. Afterwards, the work of Sun and Zhang developed a unified framework to combine recommender and dialogue system technologies together for building a chat agent (Sun & Zhang, 2018). However, it does not take into consideration the users' personality and how it influences the dialogue.

Regarding the personalization of the conversational recommender system, the current study builds on the work of Tintarev et al. (2013): supporting the idea that given that personality influences behavior, tailoring to personality may benefit the performance of the system. The innovation consists in applying this customization to a conversational recommender system to elicit a positive reaction from users. In fact, it has been studied that when the computer interacts matching the personality of the user, people are more positively inclined to interact with the computer (Nass & Moon, 2000). In the present research, this finding is put to the test if people prefer to have a conversations with the recommender system rather than receiving recommendations in the traditional way. Moreover, the conversations for a conversational recommender system adapted to the user's personality.

1.4 Movie Platform

Motivated by prior work, this research evaluates the user's experience when using a personality-based conversational RS, more specifically in the context of a movie platform. It has been decided to conduct the study on a movie platform because people often ask others for movies recommendations. This makes movie recommendations a stimulating scenario for recommender systems and has led to significant study in this area. In fact, some of the most famous movie platforms, such as MovieLens and Netflix, are focusing on personalized recommendation from user data, providing many valuable lessons learned (Amatriain & Basilico, 2015).

In recent years, recommender systems have become popular tools to assist customers in their online choices (Ricci, Rokach, & Shapira, 2015). In fact, these systems assist customers in finding the most interesting items from a huge amount of data. For media products, online movie recommendations support users to find their favorite movies. One of the foremost and broadly utilized applications to assist users in finding their favorite movies from an enormous movie catalog is Movielens (Wang, Yu, Feng, & Wang, 2014). This web-based recommender system and virtual community recommend promising movies to target customer by finding other users with similar movies rating and history (Wang, Yu, Feng, & Wang, 2014). In 2006, Netflix saw the strategic value of improving recommendations and made considerable progress with respect to the applications of Movielens for the movie prediction (Jannach, Resnick, Tuzhilin, and Zanker, 2016). Netflix opened a competition and established a prize for the research team that could program for the best collaborative filtering algorithm to predict user ratings. Netflix, putting attention on accuracy, has certainly driven recommender systems research forward. Yet, academia and industry are both focusing on creating new movie recommendation algorithms and extensions to better predict users tastes (Wang, Yu, Feng, & Wang, 2014). In the present research, the MovieExplorer platform that was developed by the Free University of Bozen-Bolzano (Unibz) will be used. This movie recommendation platform uses four different algorithms to provide a user with new movie suggestions after a search or browse task.

In the movie domain, users are accustomed to receiving suggestions based on their interests and movies previously watched (content filtering), or other profiles of similar consumers (collaborative

[1] The cold-start problem occurs when the RS cannot make reliable predictions because the system has not gathered sufficient information yet (Quijano-Sánchez, Bridge, Díaz-Agudo, & Recio-García, 2012).

filtering). Users' opinion is often a crucial component in determining if a technology is successful. So, the fact that the user is already familiar with receiving movies suggestions makes it easier to obtain feedback about the conversational RS service.

The conversational RS is built in such a way that the recommender system can provide the chatbot with information about previously watched movies and user preferences. On the other hand, the chatbot conversing interactively with the user can update the status of the recommender system. The knowledge of chatbot integrated within recommender system is still emergent. There are still several issues that need to be investigated in further research. Within this context, it is still unknown how to adapt the human-chatbot interaction to the personality of the users. The present research aims at addressing the abovementioned knowledge gap, tailoring to the language and the recommendations offered to users personality. In the scripted conversations, the movies recommendations are provided using the recommender algorithms of MovieExplorer. However, the number of serendipitous and accurate recommendations are filtered and adapted to the degree of Openness to Experience.

1.5 Research Objective

The goal of this research is to delve into the experience of the clients when they can have personalized answers to their queries. The idea is to study the effect on users satisfaction when interacting with a conversational recommender system, that is able to adapt to the different personalities of the user while providing suggestions that are in line with their preferences. An effective implementation of a conversational recommender system can examine user preferences, through the data gathered by the recommender system, and have specific conversations according to the different tastes of the customers (Akhil & Joseph, 2017). Thus, not only will it know the preferences and the movies previously watched but it will also know the personality of the users. Personality affects the way people make decisions. In fact, it has been shown that people with similar personalities may have similar preferences (Cantador, 2013). Therefore, it is necessary to treat each personality differently. From the various personality dimensions that have been acknowledged, the famous "Big Five" Five trait domains—Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness to Experience offers a comprehensive model of personality traits (DeYoung, 2006).

In particular, the research focuses on Openness to Experience. This trait of personality is seen as the need for originality and variety (McCrae, 1996). Before starting to interact with the chatbot, users will do a personality test and, then, will be separated into two different groups according to their level of Openness to Experience: high and low. The chatbot tailors the recommendations based on the fact that users score high or low in the Openness test personality. The research focuses only on this trait because it reflects the propensity to creativity, intellectual curiosity, originality and variety of experiences (Cantador, 2013). A high score in Openness to Experience entails a strong degree of novelty. A low score in Openness to Experience depicts a propensity for familiarity.

Recommendations in the present research will be personalized on Openness to Experience as it follows:

- For the person high on Openness to Experience the conversational recommender system will give more diversified and serendipitous suggestions. Serendipitous recommendations are surprising and relevant suggestions (Kaminskas & Bridge, 2014). People that are Open to Experience may consider the accuracy of the recommender system as a potential side effect, as the conversational recommender system may suggest movies that are too similar to what the user watched previously (Rook, Sabic, & Zanker, 2018).

- For the person low on Openness the conversational recommender system will suggest accurate recommendations, that are suggestions that do not deviate too much from the user's tastes and that do match with the user's profile.

There are several significant aspects of this research. Firstly, it is important to explore if conversational RSs exert a positive impact on user satisfaction by also making a comparison between automated RS and conversational RS. This has hardly been compared before, so it is the first study carrying out the aforementioned analysis. Secondly, there are no examples of conversational recommender systems that adapt to the Openness to Experience of the users based on varying accuracy and serendipitous recommendations. Thirdly, it is crucial to understand how the conversational recommender system should make personalized recommendations according to the users Openness to Experience. Different personalities should be approached differently in order to provide relevant recommendations. Thus, the current study looked at different wording, based on LIWC and Yarkoni (2010;2012), to talk differently based on the degree of Openness to Experience. Since the research on conversational recommender system is still embryonic, the current work aims at providing recommendations to help the development of personalized recommended conversations.

1.6 Research Question

On the basis of the research objectives already presented, this thesis work aims to answer the following main research question:

Can a personalized conversational movie recommender system enhance user satisfaction?

In order to formulate an answer to this main research question, the satisfaction of a user with movie recommendations will be investigated, derived from two methods: automated vs. conversational. Plus, to deeply understand if a personalized conversational recommender system increases customer satisfaction, three sub-questions have been formulated. An analysis of each sub-question will follow, with the scope of giving further support in answering the main research question.

1. *How does openness to experience influence reactions to diverse and accurate recommendations?*

Before building the conversation with the conversational recommender system, it is relevant to study how people with a different degree of Openness to Experience react to accurate or serendipitous recommendations. The idea behind is that the personality of the participants has an impact on the way in which the recommendations are perceived and, depending on this, it is possible to appreciate the originality or accuracy of the movies suggested differently.

2. *How can the conversational RS be tailored to Openness to Experience?*

Once the reactions that people high and low on Openness to Experience to the different kinds of recommendations have been analyzed, the conversational recommender system must be designed to match users' Openness to Experience. This sub-research question has a theoretical relevance and it is fundamental to build the second part of the experiment. Thus, it is not statistically tested but it gives insights on how to create scripted conversations. The personalization of the conversation according to the Openness to Experience is done in two ways:

- Tailoring the tone of the conversation according to the degree of Openness to Experience of the user. The conversations are build using the Linguistic Inquiry and Word Count (LIWC) dictionary and literature on linguistic personalities, such as the study of Yarkoni (2010).
- Adjusting the recommendations according to Openness to Experience, the idea is that matching the degree of accuracy of the suggestions to the personality can make the user more willing to use the system and to accept the recommendations. The afore-mentioned gives an insight on how openness can influence user satisfaction in conversational recommendations.

3. *What are the reactions of the participants to automated or conversational RS in terms of user satisfaction?*

This sub-research question explores if a personalized recommender system mode (traditional vs automated) has a different impact on user satisfaction. It studies whether there is a difference in user satisfaction with the two RS tools and analyzes how users perceive the recommendations provided by the different RSs. This analysis helps to understand the impact that a conversational RS, which is able to recognize customer interests and personality, can have on users. Participants' reactions would contribute to the analysis of their satisfaction and can be further used as feedback to improve conversational systems in future research. To answer this question, surveys will be conducted after performing the experiments, where the participants interact with both automated and conversational recommender systems.

1.7 Research approach

The current research develops a user study (a so called "offline experiment") to estimate the impact of a personalized conversational recommender system on user satisfaction. This user study will be divided into two phases. The first one will be conducted on the movie platform, MovieExplorer, that was introduced before. On MovieExplorer, the movie recommendations will be provided using a traditional recommender system. The second phase will be about recorded conversation, personalized for each participant. In this phase, movies will be recommended to users through a conversational recommender system. The figure below represents how the experiments will be conducted, conceptually.

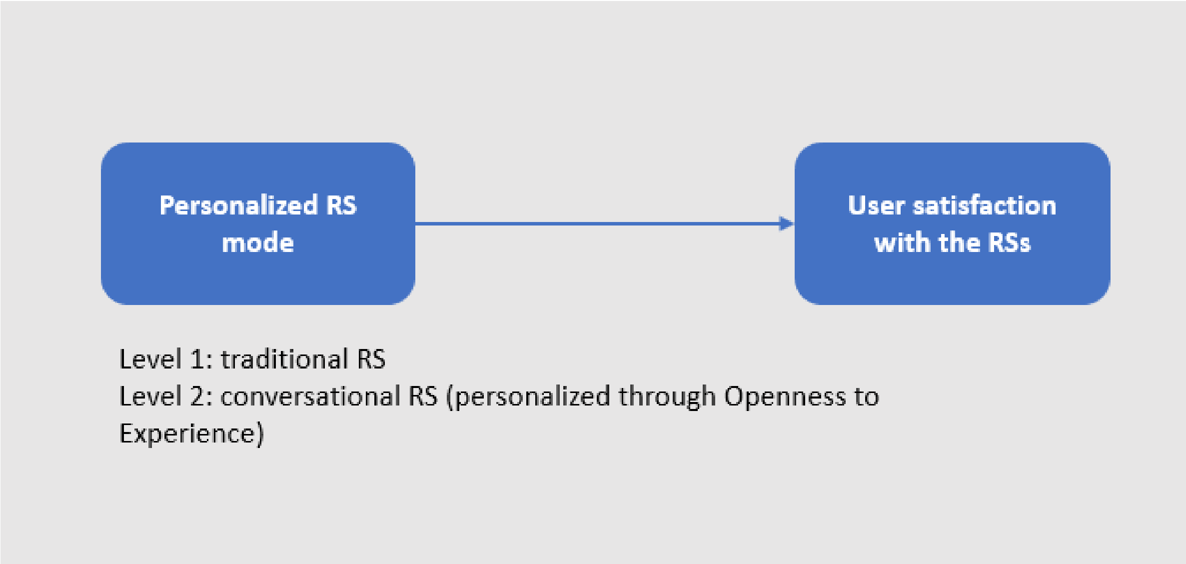


Figure 1: conceptual model

The most appropriate manner to reach the set goals already presented -- to successfully test user satisfaction and to effectively answer the above-mentioned questions -- is to use the following approach.

The first sub research question will highlight how to use the data of the recommender system to build the scripted conversations and provide suggestions to the users. It is fundamental to understand how the conversational recommender system provides the preferences that have been previously collected by the recommender system. This is also highly related to the result of the personality test. Understanding the participants' personality could be a step forward for effective communication and could increase the chances that users accept the advice provided by the conversational recommender system.

Second, after building the recorded conversational recommender system, the experiment, which includes surveys of the participants, will be conducted. The final objective will then be to analyze if the conversational recommender system that gives suggestions knowing the personality of the user and its preferences can improve the users satisfaction. An overview of the research approach is visualized in Table 1 below:

Research sub-question	Method
How does openness to experience influence reactions to diverse and accurate recommendations?	Pretested first experiment (Movie Explorer), results used to build the scripted conversations
How can the conversational RS be tailored to Openness to Experience?	Literature review + user study (building the conversations)
What are the reactions of the participants to automated or conversational RS in terms of user satisfaction?	Second experiment (recorded conversations)+ survey

Table 1: overview research approach

1.8 Data gathering

In order to conduct the experiments, participants will be approached and gathered from the researcher's personal network (colleagues and friends). The drawback of this process is that surveys could be limited in terms of volume and time. This could raise issues of representativeness (Opdenakker, R., 2006). As a consequence, it is necessary to contact a diverse group, in order to be able to draw the most truthful picture possible. Through the first experiment, it will be possible to find out if there is a relationship between accuracy and serendipity. Also, the first experiment will be set up to assess the Openness to Experience score for each participant. This is needed to answer sub-research question one later on. To answer the second sub-question, participants will be preselected on their high/low scores on Openness to Experience. In this second phase, scripted conversations, personalized both in terms of text and in term of type on movies, will be offered. Finally, to answer the third question and understand if participants appreciated more automated or conversational recommender system, reactions and user satisfaction will be captured through a survey.

This research compares two different systems of recommendations (automated vs conversational) and analyses the impact of a personalized conversational recommender system. It also gives an insight into how conversations can be personalized.

1.9 Flow Diagram

The following Flow Diagram visually illustrates the structure of the thesis project and helps to find a balance in the research load of the sub-questions. All research methods to answer those two sub-questions are mentioned in the previous paragraphs. The figure below shows the entire outline of the thesis visually. To each sub-question corresponds a research method, explicitly mentioned in the paper.

Research approach	Objective	Outcome
Literature review	Introduction	Problem statement, main and sub research questions and knowledge gap
Literature review + users' study	How Openness to Experience influences the reactions to diverse and accurate recommendations SQ1	Analyze the relationship between accuracy/serendipity and Openness to Experience
Experiment + users' surveys	How to tailor the conversation to Openness to Experience SQ2	Design the scripted conversation based on personality
Experiment + users' surveys	Explore users' reactions SQ3	Assess the users' satisfaction with the two RSs
	Conclusion, reflections and future research	Answer to the main research question

Table 2: Flow diagram

1.10 Report structure

The report is structured as follows. As already seen, **Chapter 1** provided the reader with an introduction related to this research, reporting the aims and the objectives of the study. Next, the literature review is reported in **Chapter 2**. This chapter documents the state of art about conversational RS. Plus, it explains the main factors playing a role in the personalization of conversational RSs. **Chapter 3** discusses the research methodology, including the experiment design, relevant information on materials used for the two experiments, the explanation of the procedure followed and the measures taken into account. Subsequently, **Chapter 4** describes the results of the experiment. The discussion of the scientific and practical relevance and analysis of the empirical findings follows in **Chapter 5**. This chapter reports also the limitations and suggestions for future work. **Chapter 6** offers a conclusion that covers the overall significance of the study.

2. Literature review

This section examines the literature analyzed and reviewed to support the thesis and to understand the relative concepts and theories at stake. This chapter is also necessary to provide enough evidence to back the present research. It can offer some insights on the most significant elements that have to be considered to provide an answer to the following main research question: "*Can a personalized conversational movie recommender system enhance user satisfaction?*".

2.1 Recommender Systems

Recommender Systems (RSs) are software tools and techniques that can suggest the most relevant items available to users. The recommendations offered by a recommender system are intended to assist the users in several decision-making tasks, such as what products to buy, what movies to watch, or what news to read (Ricci, Rokach & Shapira, 2015). RSs take into account users' preferences and constraints to predict the most suitable items for them. In order to perform such a computational task, RSs gather data and information from users about their interests, preferences and online behavior (Ricci, Rokach & Shapira, 2015).

In the 1960s, there was already the idea of utilizing a computer in order to filter information, taking into account the tastes of the user. This idea was enclosed in the term "selective dissemination of information" (Hensley, 1963). At the beginning, the systems utilized keywords explicitly selected by the users to categorize and filter the documents. Thereafter, to represent documents and their corresponding representation of users interests, more sophisticated techniques like weighted term vectors[2] or more elaborated document analysis methods like latent semantic indexing[3] were applied (Jannach, Resnick, Tuzhilin & Zanker, 2016). RSs based on these techniques are usually named "content-based filtering" approaches. In 1994, Resnick et al. introduced a system called GroupLens, which made automated forecasts of the items that users would like according to the nearest-neighbor technique[4]. In 1999, Shafer et al. reported on how the main commerce websites were already adopting RS to assist their customers find items to purchase. Over the last decade, the explosive growth of the "Internet of Things", the large amount of web-base data, and the development of new e-business facilities (rating products, item comparison, etc.), have often overwhelmed users, leading them to make poor choices (Ricci, Rokach & Shapira, 2015). RSs have proven to be a powerful tool for users to filter through large amount of information and product spaces (Ekstrand, Riedl & Konstan, 2011) and for reducing information overload (Rook, Sabic & Zanker, 2018).

Ricci et al. (2015) analyzed several reasons to adopt this technology:

---

[2] Weighted term vector is a technique used in searches of information retrieval, user modeling and text mining.

[3] Latent semantic analysis is a method that investigates the interactions between a collection of documents and the terms they contain in natural language processing.

[4] The nearest-neighbor technique consists in representing "objects as points in a high-dimensional metric space, and then uses nearness with respect to the underlying distance as a means of indexing and similarity-based comparison" (Kleinberg, 1997, p. 601)

• *Increase conversion rate*. Visitors that browse through websites are more willing to accept the recommendations and consume items. This object is reached because the recommended items are more likely to match the desires of the user.

• *Promote more diverse items.* An RS can suggest items that differ from the most popular ones in the respective category, which might be difficult to discover without a precise recommendation. This is made possible as the RS already knows the interests of its users. However, without an RS the service provider cannot take the risk of advertising items which may not fit the taste of a specific user.

• *Increase user satisfaction.* A well-designed RS plays a role in improving the experience of the user. Indeed, the combination of interesting suggestions and a usable interface will improve the user's subjective evaluation of the system. If users find the recommendations relevant and the human computer interaction is properly designed, they will also enjoy using the system and will be more likely to accept the recommendations.

• *Increase user fidelity.* The RSs acquire user information and capture their earlier interactions with the site, as for example the user's ratings of items. Therefore, the more time the user spends interacting with the website, the more precise the user's model becomes. As the collection of user's preferences progresses and improves, so does the performance of the recommender output in matching the needs and tastes of the users.

2.1.1 Accuracy in recommendations

The accuracy of predictions about the desires and interests of users is an essential principle in RS research, and it is typically used to assess its performance. The accuracy is based on the identification of all the features and signals available in the data. This involves all types of temporal effects regarding the time-drifting nature and the dynamics of user-item interactions (Ricci, Rokach & Shapira, 2015). Moreover, to improve accuracy it is necessary to analyze hidden feedback, like which are the items that users choose to rate (regardless of rating values). In fact, users do not randomly select the items to rate, and the selection they make can reveal important aspects of the preferences of the users, going beyond the numerical values of the ratings. In the most popular RSs, accuracy is seen as the way to give appropriate suggestions and help users in finding what they want among the large available number of items. For instance, the RS used by Netflix with its focus on accuracy makes optimum predictions for as-yet unrated items (Jannach, Resnick, Tuzhilin & Zanker, 2016).

Usually, the accuracy of recommendations in an RS based on collaborative filtering is obtained by analyzing the behavioral patterns previously disclosed by all users (Rook, Sabic, Zanker, 2018). It is believed that two users who agree on the same item will likely be in agreement on another item too. Collaborative filtering starts with certain item/user affinity scores (Christakopoulou, Radlinski & Hofmann, 2016), while the content-based filtering categorizes the costumers according to the features of the items they dislike or like. Hence, higher accuracy is reached by suggesting to the users, items similar to those that they have already rated (Christakopoulou, Radlinski & Hofmann, 2016).

In the movie domain, recommender systems are broadly adopted as personalization tools (Adomavicius, & Tuzhilin, 2005). The movies are usually suggested based on the following metrics: most viewed on the site, similar users' behavior, demography of the customers, investigation of the past behavior of the customers as a way to predict their future behavior. Generally, these techniques enable the personalization of the site, giving individual suggestions to each customer (Schafer, Konstan & Riedl, 1999). However, a potential downside of accuracy is that the RS could suggest items too similar to what the user previously liked. Such recommendations could not satisfy the user anymore

(Rook, Sabic, Zanker, 2018). For example, once a recommendation list contains one The Lord of Rings movie, there is a reduced value from suggesting its sequels, as the user is likely already aware of them (Jannach, Resnick, Tuzhilin & Zanker, 2016). Thus, suggesting items whose scores are high when matched against the user profile can sometimes be unfavorable to user engagement (Jannach, Resnick, Tuzhilin & Zanker, 2016). Nevertheless, finding the right mix of familiar and original items can be challenging, and further research is needed. This issue is also known as *lack of serendipity* to highlight the tendency of the RSs to create recommendation lists with a limited degree of originality (Ricci, Rokach & Shapira, 2015). To address this issue, academics are exploring other quality metrics that take advantage of the originality and unexpectedness of recommended items and that are willing to find the right mix of similarity and diversity (Rook, Sabic, Zanker, 2018). The next section will provide a better understanding of diverse recommendations and serendipity.

2.1.2 Alternative recommendation measures: diversity and serendipity

RS research has traditionally focused on accurately predicting users' ratings for unseen items. However, accuracy is not the only important objective of recommendations and has its drawbacks, as already discussed in Sec. 2.2.1. Ensuring that the suggested items are new and/or diverse has become crucial for giving recommendations that are both surprising but relevant (Kaminskas & Bridge, 2014). The experience of receiving unexpected and surprising recommendations is known as serendipity. Serendipitous recommendations assist users in finding a surprisingly interesting item that they may not have otherwise discovered (Lops, De Gemmis & Semeraro, 2011). For instance, suggesting a movie played by the user's favorite actor is a new recommendation if the user is unaware of it. On the other hand, it may not be fully serendipitous, since the users would have probably discovered the movie on their own (Ricci, Rokach & Shapira, 2015). Non-trivial and surprising suggestions are vital to make the system look more lively. The introduction of surprising recommendations could help reveal the unexpressed wishes of the users that they did not know existed but those that perfectly matched their lifestyle. For instance, suggesting the purchase of milk and bread in a grocery shop is obvious and will not produce added value for the users or additional sales for the company (Jannach, Resnick, Tuzhilin & Zanker, 2016). The RS should consider factors such as novelty and unpredictability of an item in order to include it in its recommendations.

Several researchers have attempted to build algorithms to achieve the ideal balance between similarity and diversity. Hurley and Zhang (2011) formulated the tradeoff between diversity and similarity as a problem of binary optimization and defined the parameter of the input control to explicitly adjust the two metrics. To introduce diversification in the recommendation lists, Ziegler et al. (2005) implemented the intra-list similarity metric. Hence, researches are increasingly focusing on serendipity in recommendations, exploring individual differences in user personality together with suggested items. However, adjusting the degree of diversity within the set of multiple recommendations based on users' personality has been rarely explored. The present work will give an insight into how the conversational RS could use users' personality as a way to differentiate movie recommendations.

2.2 Chatbot

Chatbots are "machines conversation system[s] which interact with human users via natural conversational language" (Shawar & Atwell, 2005, p. 489). This technology combines language models and computational algorithms to mimic human conversations (Hill, Ford, & Farreras, 2015). Chatbots can recognize the inputs given by users and provide a reply that is considered to be the most intelligent response to the input. Functionally, the approach to natural language processing used by chatbots nowadays is an extension of the same technique adopted by ELIZA (Hill, Ford, & Farreras, 2015). Several chatbot architectures and technologies have recently emerged, each attempting to accurately represent the language spoken by people in real conversational settings (Hill, Ford, & Farreras, 2015). In 2016, Microsoft, for instance, introduced its idea of conversations as a platform and launched the first chatbots on Skype (Følstad, & Brandtzæg, 2017). Meanwhile, Facebook was developing chatbots for Messenger. Later on, Google Assistant was introduced with the aim of helping users with everyday questions, regarding the weather forecast, finding the nearest restaurant, etc.

Yet conversations with virtual assistants are not often smooth and break down fast. However, the main technology companies believe that chatbots are the future way of human-computer interaction. In fact, currently, mobile messaging applications have around 1.5 billion users worldwide. The typical interaction remains between two humans, with a machine at each end. With advances in AI, it may be possible to utilize the natural-language interaction to connect users with machine agents better known as chatbots (Følstad, & Brandtzæg, 2017).

The Web is changing, and Chen et al. (2012) have identified the following three phases, placing the chatbots as the next big thing:

- In Web 1.0, search engines such as Google and Yahoo, and e-commerce businesses such as Amazon and eBay play a characteristic role. These online platforms allow business organizations to interact directly with their customers. Furthermore, these platforms record the users' searches and interactions through cookies and server logs, in order to understand customers' desires and to identify new business opportunities.
- Web 2.0 is based on social and crowd-sourcing systems. Here, the user's online activities are captured in its entirety, which is then analyzed by Web tools to reveal their browsing and purchasing patterns. This recorded data spans from various social media interactions like online forums, blogs, social networking, etc. to virtual worlds and social games (O'Reilly, 2005).
- Web 3.0 is the coming era of mobile interfaces and human–computer interactions. Thus, in the near future, a webpage or a dedicated application may be replaced by chatbots, that can become the favored user interface for different activities (Følstad, & Brandtzæg, 2017).

2.2.1 Chatbot Features

Currently, the main purpose of chatbots is to cope with redundant requests and waiting times, reducing the costs of traditional customer service and providing immediate 24/7 customer service. This service is provided to improve the experience of the users and to build a connection between the customer and the company. Chatbots are an excellent tool for quick response to users. They assist users by replying to questions that are hard to find, thus saving time and providing entertainment (Dahiya, 2017). The chatbot must be built in such a way to be simple, easy to use and easily understandable.

Chatbots, in general, exploit *natural language technologies* mainly *to engage users in information-seeking and task-oriented dialogs* (Kerly, Hall & Bull, 2007, p.181). Specifically:

- *Information-seeking* systems offer relevant answers to the users' queries. An example of this is asking for the status of an order. When the customer asks the system to trace its order, the system retrieves the required information and presents it to the user (de Haan, 2018).

- *Task-oriented* systems are designed to help users to execute tasks. For instance, users can directly ask the chatbot to place an order. The users explain what they are looking for, along with their preferences to the chatbot and, if needed, the chatbot asks further information. Once all specifics have been processed, the users give their approval and the order is placed (de Haan, 2018).

In order to improve task-oriented systems and have deeper knowledge about users preferences, the current research will focus on conversational recommender systems, which are a sub-category of the aforementioned. Dynamically adapting the conversations and users' preferences while the dialogue evolves is the challenge to be faced.

2.2.2 Conversational Recommender System

Recommender systems are currently part of our daily life, however, in tight interaction with the users, they begin to show their limits. In the recent years, due to the new wave of deep learning approaches, conversational recommender systems have been proposed as a way to build a connection between users and the company (Anelli, Basile, Bridge, Di Noia, Lops, Musto & Zanker, 2018). In line with the observations made in 2.2.1, these systems are task-oriented conversational chatbots that assist users to find information, products or services (Sun & Zhang, 2018). Moreover, they engage with their users to help them in articulating their preferences and expressing their opinion about the past recommendations in order to improve the future recommendations (Anelli, Basile, Bridge, Di Noia, Lops, Musto & Zanker, 2018). Good recommendations, gained by interactively soliciting and recognizing users' intentions, can create business opportunities. Thus, combining recommendation techniques into conversational systems could benefit both users and companies. In fact, this could improve the conversion rate of a shopping/sales chatbot and better fulfil users' needs (Sun & Zhang, 2018). However, despite the commercial potential, the research on this topic is still very limited and the existing solutions are rudimentary.

One of the first steps towards conversational recommendation systems was made by Christakopoulou et al. (2016), who developed a dialog system that is able to infer user preferences and understand, which questions are better for the system to give faster recommendations. This conversational recommender system uses offline data, that are *a priori* available sets of user-item ratings to learn the interdependencies among users and items (Christakopoulou, Radlinski & Hofmann, 2016). These data are used to initialize the model. Afterwards, the mechanism selects a question to ask the user. Based on the answer of the user all the model parameters are updated. Therefore, the understanding of the system about the users' and the items in the question changes. Then, the mechanism selects the next question, the user replies and so on, until the system comes up with the list of recommended items. Fig. 2 below shows the described technique. The above-mentioned dialog system only asks if a user likes an element or if the user prefers element A over element B, while a typical task-oriented dialogue system usually requires facets from users. On the basis of this study, Sun & Zhang (2018) tried to maximize the long-term utility by reinforcing the learning framework and asking questions that request facets from users. However, the work has still room for further improvement.

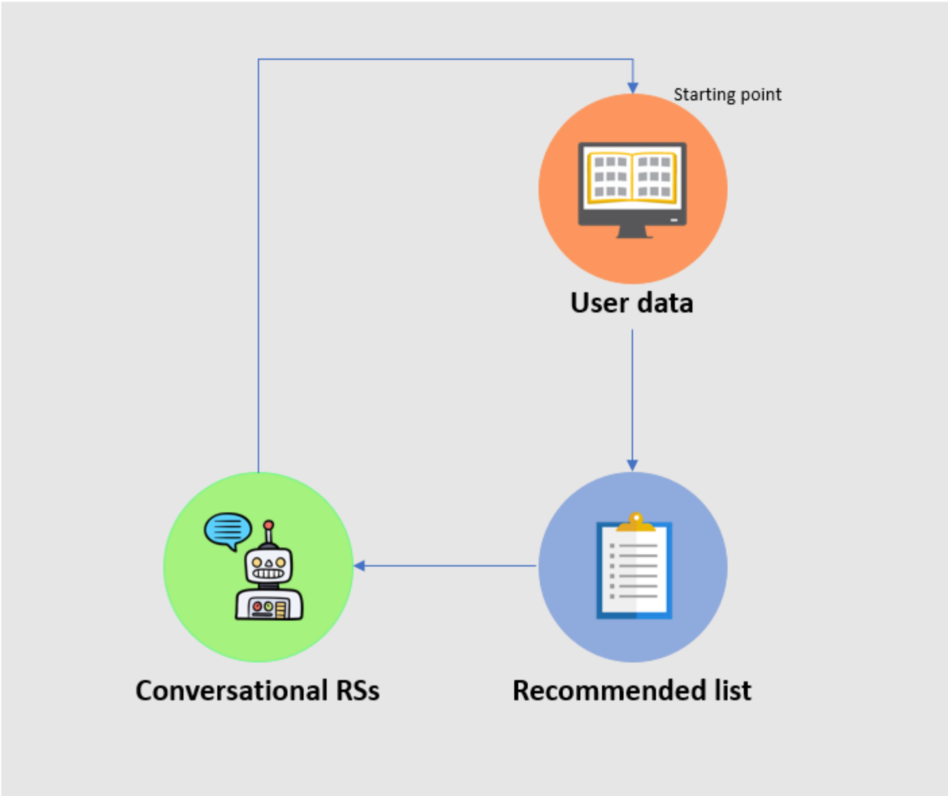


Figure 2: Conversational RS mechanism

In the upcoming era of chatbots and natural-language user interfaces, conversational touch-points will become the most pivotal factor for differentiating the user' service (Følstad, & Brandtzæg, 2017). As a result, interactions, services and content, that were earlier differentiated based on the user interfaces, will now be under the same conversational threads. This means that interaction with the digital system will not occur through swiping, scrolling or button clicking, but rather through text strings in natural language (Følstad, & Brandtzæg, 2017). This research reflects on the need for analyzing the personalization of human-computer interaction to meet the challenges and the opportunities of the future era of chatbots.

2.3 Conversational RSs and Personalization

Conversational RSs could take human-computer interaction to the next level, creating a close tie between the user and the computer based on the exchange of information. It is therefore important to first discuss what online personalization is, and how it has been studied so far.

Online personalization is based on the concept of storing as many historical customer session data as possible, and then querying those data stored as customers navigate through a website (VanderMeer, Dutta, Datta, Ramamritham, & Navanthe, 2000). It is possible to query online navigation patterns to provide customized services, recommendations and real-time responses to the users. Online personalization plays a key role in assisting clients to go through the information on the web and facing the information infrastructure, that is constantly changing. Moreover, it has also demonstrated to be useful also for businesses and organizations, since it is crucial for intelligent information search and data management (Abu-Dalbouh, 2016). The whole process of online personalization is centered on end-user and on the support it needs to access, search, and find information (Abu-Dalbouh, 2016).

Previous researches on human-computer interaction revealed that people are more willing to interact with a human-like computer that “knows” them and offer personalized service (Lee & Nass, 2004; Harianto, 2018). So, nowadays, almost every human-computer interaction is personalized.

Personalization consists in using technology and customer data to tailor online business interactions with the individual customer. With the increased availability of information, it has become imperative and more feasible to determine what information is interesting for the customers while minimizing the search through irrelevant information (Foltz, & Dumais, 1992). The use of customer data either previously obtained or provided in real time about the customer can help businesses to meet the needs of the customer. Successful personalization applications rest on getting acquainted with consumers’ personal interests and behaviors that are typically gathered from their online data and, stored in the form of consumer profiles (Adomavicius, & Tuzhilin, 2005). Personalization consists of creating customer loyalty by establishing significant one-to-one relationships. In this context, it is crucial to understand the desires of each individual and contribute to meeting a goal that proficiently responds to the needs of each individual in a given situation. Cultivating one-to-one relationships makes future communications smoother and more efficient, benefiting both business and clients in the long run (Fan & Poole, 2006). It is widely expected that personalization can create benefits for the customer (Vesanen, 2007).

Personalization requires the ability of the service provider to adjust its offerings to the individual customer. In return, the customer may let the service provider know to what extent, the offerings satisfied their needs. In fact, “the customer is always a co-creator of value” (Vargo & Lusch, 2008, p. 3). Communication between the client and the service provider is based on understanding what the client uniquely desires, and what the service provider can do, uniquely, for that specific client (Ball, Coelho & Vilares, 2006). Interactivity is essential to meeting customer’ high-level needs and establishing connection between a market beneficiary and a market provider (Vargo & Lusch, 2008). When a company interacts effectively with its customer, customer’s feeling of being personally cared and addressed for should increase (Ball, Coelho & Vilares, 2006). Thus, personalization strategies have always been enticing for e-commerce websites that recognized personalized services as a way to improve customer experience and to support recommendations of off-the-shelf goods (Goy, Ardissono, & Petrone, 2007). Today’s multimedia platforms are all tailored to their different customers; Spotify has personalized playlists and Netflix suggests different series to different people. It has been widely accepted that services that fit the customer’s requests better are more satisfying than one-size-fits-all (Ball, Coelho & Vilares, 2006). There are several advantages related to personalization such as improved preference match, recommendations, services and communication. In addition, personalization may encourage the customers to believe that the firm is more caring towards them, increasing their trust and satisfaction with the service offered.

In a fast-changing environment, where the fleeting attention of the user develops into a main factor, it is indispensable for the technology providers to personalize the system and design engaging experiences (Attfield, Kazai, Lalmas, & Piwowarski, 2011). In this regard, human computer interaction is seen as the key to improve user satisfaction and converse rate. During an interactive mediated activity, users satisfaction is considered as one of the most important dimensions (O’ Brien & Toms, 2008) of their experience (Hassenzahl and Tractinsky, 2006; Law, Roto, Hassenzahl, Vermeeren, & Kort, 2009). Interactive mediated activity refers to the human activity backed by digital interactive technologies, for instance, the Internet, virtual reality systems or computer apps (Bouvier, Sehaba & Lavoué, 2014). In the latest years, there has been a proliferation of user satisfaction definitions suggested in the literature. Many decades ago, Howard and Sheth (1969) defined user satisfaction as “the buyer’s cognitive state of being adequately or inadequately rewarded” for the items purchased

(Howard & Sheth, 1969, p.145). According to Giese and Cote (2000) users satisfaction at a specific moment is an emotional or cognitive reaction to a specific focus (products, consumer experience, etc.) at a specific time. Hunt (1977) referred to user satisfaction as an evaluation that resulted in the consumer experience being at least as good as it was supposed to be. At the moment of designing user-centered web applications, satisfaction is a decisive component. It is strictly linked to the quality of the user's experience, highlighting the positive features of their interaction and, in specific, the phenomena connected with technology captivation (Attfield, Kazai, Lalmas & Piwowarski, 2011).

2.4 Personalization through Individual Differences in Personality

Personality influences all areas of people lives. It governs who they are and how they make decisions (Smith, Dennis, Masthoff, & Tintarev, 2018). According to the psychological literature, personality is determined by a set of features that define how people are different from each other. Human behavior and experience are influenced by personality traits, that show the "consistent patterns of thoughts, feelings, and actions" in individuals (McCrae & Costa, 1995, p.235). It is believed that personalities should be used also in the businesses, in order to create customized services. In fact, consumers will be more influenced by messages that match their personality. This will be reflected in a greater change of attitude towards the direction that the message is seeking (Nass & Moon, 2000). As will be analyzed in the next section, personalized systems that adjust to end users should consider the personality of the user to achieve successful results (Smith, Dennis, Masthoff, & Tintarev, 2018).

Personalized offerings can impact the satisfaction of the consumer. Recent studies and developments in personalization have sought to develop new user-oriented approaches to achieve better accuracy and satisfaction of recommendations (Liang, Lai, & Ku, 2006; Kang, Shin, & Gong, 2016). However, despite the attention given to personalization (Tkalcic & Kunaver, 2009), only a few works in the previous research have perceived that using consumers' personalities is a way to accomplish that. In this context, personalization by adjusting the recommendations to the personality of the consumers should enhance communication and increase user satisfaction.

As personality-based recommenders have a deeper understanding of the users from a psychological point of view, they can provide more personalized and tailored services or information. However, adoption of RS to a user's personality is still in the developing stage. The existing researches and studies have demonstrated that the above-mentioned RSs can be applicable to a wider range of people and that they can also be more successful (Hu & Pu, 2009). Actually, the same can be affirmed about chatbots: conversations should be adapted to personality. In fact, if a computer changes personality traits and becomes more familiar to the user it is interacting with, the individual is more positively disposed to the computer than when its personality traits and its answers are the same for all the users (Nass& Moon, 2000). In other words, if computers' "personality" matches that of the user, people are more likely to interact with the system. Several tests have been run on how this could be achieved in so-called "offline experiments". Tkalcic and Kunaver (2009), for instance, use personality to measure the user similarity for collaborative filtering recommender systems. Hu and Pu (2009) also discovered that recommender systems that take user personality into consideration are more effective in enhancing user allegiance to the system and reducing their cognitive effort compared to non- personality based systems (Chen, Wu & He, 2013). Still, most of these offline experiments on personalizing RS have used the Five Factor Model or the Big Five to achieve personalization. The next section introduces the Big Five in more detail. In particular, it focuses on how recommendations can be personalized based on Openness to Experience.

2.5 Openness to Experience and Conversational RSs

The Big Five model of personality, is one of the most widely adopted personality theories. It introduces a five-factor model of personality. The five factors are usually indicated as follows: (a) Agreeableness, (b) Extroversion, (c) Neuroticism, (d) Conscientiousness, and (e) Openness to Experience (McCrae and Costa, 1999). Each of these Big Five personality traits is composed of more specific features and traits. The personality traits Agreeableness and Extraversion represent the social dimensions of the Big Five model (Rook, Sabic, & Zanker, 2018). Agreeableness is characterized by the propensity to get along with other people, demonstrate helping behaviors and emotionally support others (Graziano, Habashi, Sheese, & Tobin, 2007). Extraversion taps into the extent to which someone is cheerful, warm and outgoing (Costa and MacCrae, 1992). Trait Neuroticism describes the propensity of a person to experience distress and negative emotional states (McCrae and John 1992). Trait Conscientiousness denotes the individual differences in impulse, reliability, determination, self-control and level of thoughtfulness (McCrae and Costa, 1995). Likewise, Openness to Experience refers to the creativity, curiosity and desire to novel experiences and ideas.
Of the five factors, Openness to Experience is particularly related to the need for novelty, diversity and intrinsic appreciation for new experiences over familiarity and routine (McCrae, 1996). The desire for new experiences provides an incentive to accept dissonance and uncertainty (McCrae & Costa, 1997). People that score high in openness are more receptive to emotions and have a greater intellectual curiosity (Cantador, 2013). Meanwhile, people that score low on Openness to Experience are more conservative and have more common (i.e., average, mundane) interests in life (Blumer & Doering, 2012).

It is stimulating to study the impact of Openness on user satisfaction with proactive recommendations because Openness reflects the interest in art (such as movies) and engagement with intellectual ideas. However, existing studies reveal mixed, and often contradictory, information about the influence of Openness to Experience on user satisfaction (Rook, Sabic & Zanker, 2018). For a movie platform, users high on Openness, as a result of their willingness to have new stimuli, probably prefer a chatbot that suggests new and original movies to them. Conversely, users low on Openness probably prefer one that suggests more accurate movie recommendations (more similar to their tastes). The research believes that a personalized conversational RSs could play a central role in satisfying customer wants and needs (Kang, Shin, & Gong, 2016), especially, in the movie domain, where interactivity and engagement are fundamental aspects. Plus, in the present research, the conversational RS will use different words based on the degree of Openness to Experience. Previous studies presented the relation between language and personality (Fast & Funder, 2008), and explored a relationship on personality domains such as the Big Five with words usage (Yarkoni, 2010). Specifically, words that are top on high Openness to Experience are: *novel* and *bizarre,* whereas those suitable to low Openness to Experience are: *afterwards* and *little* (Yarkoni, 2010).

Recommender systems based on people's personality are an increasing area of interest, and some links between client's tastes and personality have already been identified (Tintarev, Dennis, & Masthoff, 2013). The novelty lies in applying this personalization to conversational RSs and examining how it affects user satisfaction. User satisfaction is analyzed to explore the importance of personalization through Openness to Experience for conversational RSs technology. Overall, adapting recommendations to the degree of Openness to Experience of the users serves to moderate the impact of conversational recommender system on their satisfaction.

Tkalcic and Chen (2015) stated that the research activities are approximately split into: (1) the implicit drawing out of Big Five personality characteristics from user data, and (2) the explicit calculation of Big Five personality characteristics through tests (Rook, Sabic & Zanker, 2018). Overall, however, the effects of Big Five personality traits on satisfaction with diverse recommendations is not mature enough and inconclusive. Regarding trait-Openness to Experience, for instance, some studies reported no statistically significant evidence of the correlation between Openness to experience and preference for a more diverse set of recommended items (Tintarev, Dennis & Masthoff, 2013). In some other cases, the fact that there is no significant correlation is only partially confirmed (Chen, Wu & He, 2013). Thus, it is interesting to explore how matching the suggestions to the degree of Openness to Experience of users can impact their satisfaction in conversational and interactive recommendations. With this regard, it has been assumed that people high (vs. low) in Openness to Experience would prefer to receive diversified (vs. homogeneous) recommendations because they are more willing to have new experiences.

Building on these loose threads of evidence, the mockup of the conversational RS will be built in such a way to suggest more highly accurate recommendations to people low on Openness to Experience and more diverse movies suggestions to people high on Openness to Experience. This will be explained in more detailed in Section 3.4.

2.6 Summary Literature Review

Recommender System	(RSs) are software tequinques that ca n provide users with the most appropr iate products. RSs have proven to be a strong instrument for consumers to fi lter through big quantities of data and product spaces and reduce overload of data.	Rook, Sabic & Zanker, 2018 Ricci, Rokach & Shapira, 2015
Accurate recommendations	Accurate recommendations are dependable and useful recommendations, provided based on users profiles.	Rook, Sabic & Zanker, 2018
Serendipitous recommendations	Serendipitous recommendations are unexpected and surprising recommendations, that can help users in finding surprisingly interesting items that they may not have otherwise discovered.	Lops, De Gemmis & Semeraro, 2011 Ge, Delgado-Battenfeld & Jannach, 2010
Chatbots	Chatbots are conversational machine systems that interact through natural conversational language with users.	Brandtzaeg & Følstad, 2017

Conversational RSs	Conversational RSs are task-oriented conversational chatbots that assist users to find information, products or service through a dialog system	Sun & Zhang, 2018 Christakopoulou, Radlinski & Hofmann, 2016 Anelli, Basile, Bridge, Di Noia, Lops, Musto & Zanker, 2018
Online personalization	Online personalization is the ability of the service provider to adjust its offerings to the individual customer' needs and interests.	Adomavicius & Tuzhilin, 2005 Tintarev, Dennis, & Masthoff, 2013 Tintarev & Masthoff, 2012; 2015
Openness to Experience	Openness to Experience is one of the Big Five Personality. It refers to the interest in art, engagement with intellectual ideas, curiosity and desire to novel experiences.	McCrae and Costa, 1996; 1999 Cantador, 2013

Table 3: Summary Literature review

3. Method

3.1 Participants

Participation was on a voluntary basis, and participants had to give informed consent to having their data collected for research purposes right after reading the purpose of the study (see below). Participants were informed that the generated data would have been used for didactic purposes to generate statistical results in anonymized and aggregated form. The initial sample consisted of 113 participants (58 men, 45 women, 10 other; *Mage* = 24.76 years, *SD* = 3.32) that had been collected over a period of several weeks. Tables 4 and 5 provide the descriptives of the initial sample.

Nationality	Frequency
Italian	63
Spanish	10
Greek	8
Indian	7
Others	25

Table 4: Nationality full sample

Education	Frequency
Master	38
Bachelor	45
High school	14
College	3
Doctorate	2
Others	10

Table 5: Education full sample

The second part of the experiment was conducted with some of the participants who had already completed the first part. For the second part of the experiment only a small part of the participants, that joined the first part was asked to participate. Participants were randomly selected from participants of the full sample, that were willing to continue with the experiment. So, 22 (12 men, 10 women; *Mage* = 24.68 years, *SD* = 2.82) participants followed the second experiment, 11 of them low on Openness to Experience and the other 11 high on Openness to Experience. Tables 6 and 7 provide the descriptives of the second sample.

Nationality	Frequency
Italian	10
Spanish	1
Greek	3
Indian	1
Others	7

Table 6: Nationality second sample

Education	Frequency
Master	7
Bachelor	12
High school	3

Table 7: Education second sample

3.2 Experimental design

The experiment consists of two phases (two different modes of providing recommendations: fully automated vs. conversational) at the end of which a list of recommended movies was presented to the participants. In the first phase, an interactive platform for movie search and selection that was developed by researchers of the university of Bolzano offered fully automated movie recommendations based on a list of favorite movies provided by the participants. In the second phase, the movie recommendations additionally also had a dependence on the participants' Openness to Experience and offered a similar set of movie recommendations as in phase one, but in a personalized conversational manner. In each phase, once the recommendations are presented, the participants were asked to rate them on a scale of 1 to 10 stars. Finally, user satisfaction with regard to the two methods of recommendation was measured by means of a detailed survey.

3.3 Materials

3.3.1 First part

The first part of the experiment was conducted on MovieExplorer (a movie platform and database, which will be introduced in detail below). After filling out a pre-survey on Openness to Experience and the demographic questions, the participants were sent to the landing page where the participants were asked to complete the following task:

"Your task is to select a top 10 of movies taken from the provided database. For each selected movie, please first indicate if you have seen the movie before, and how you would rate it. Next, add the movie to your Top-10 list on the right side of the screen. Once you have completed your top 10, hit the button to get recommendations. In return, you will receive some movie recommendations from us."

The MovieExplorer is a repository of movies built and hosted by the recommender systems group of prof. Zanker at the faculty of Computer Science in Bolzano, Italy. The movie repository is spread over multiple pages, with each page showing a list of 10 movies. The movies are by default sorted in reverse chronologically order, i.e. newest first. The participants also have the option to change the sorting based on popularity and rating. They can navigate to the previous or the next page by using the corresponding buttons displayed on the top. The participants can select the movies either from the list described or also search for movies by name and genre by using the control widget on the top right. The selected movies are automatically filled in the selected list on the right. The participants have full freedom to change opinion and remove movies from the selected list or revisit movies already rated. A screenshot of the interface is shown in Fig 3.

Figure 3: landing page

Clicking on a movie will open a page that contains detailed information regarding the movie such as title, poster, genre, release year, the average rating on the MovieExplorer database, a short synopsis and the clickable movie trailer. The platform also provides the participants with a list of movies similar to the one selected. The participants have the possibility to close the items panes and head back to the landing page. Fig. 4 gives an overview of what is described above.

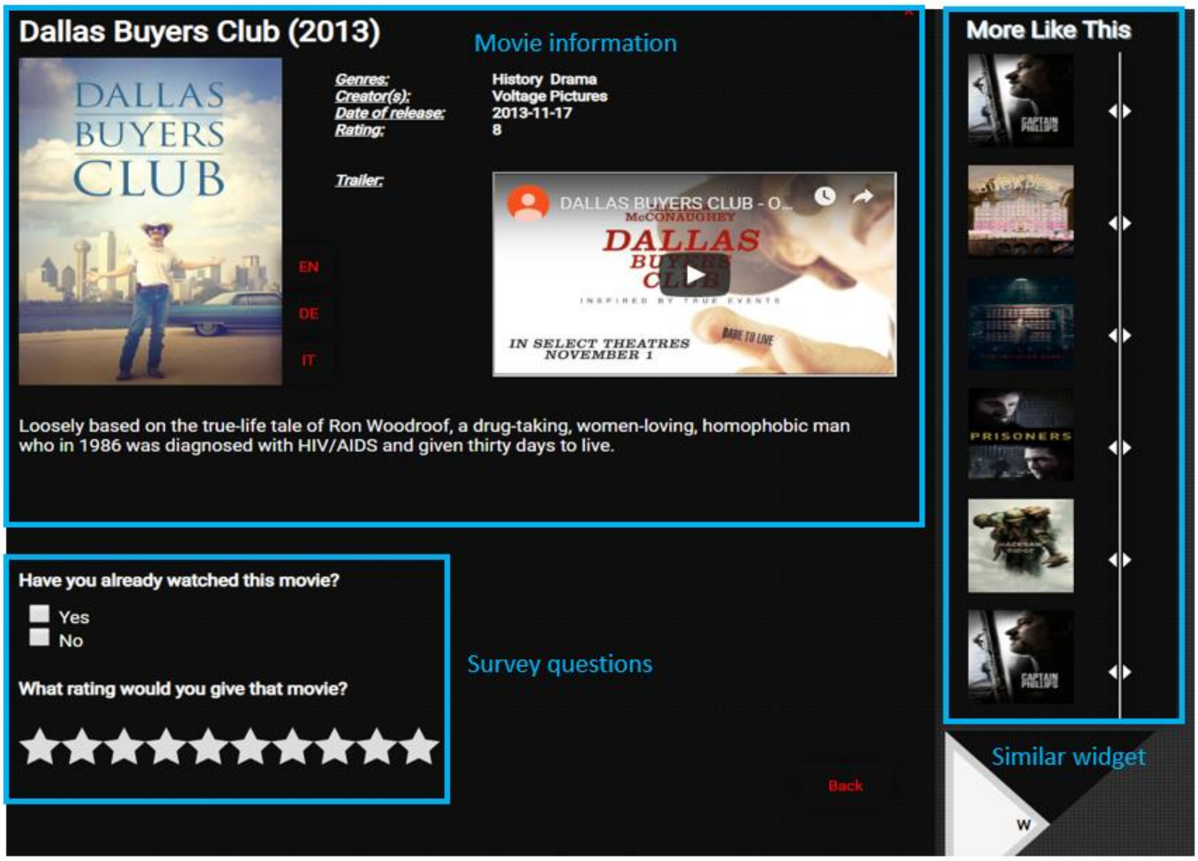


Figure 4: movie information

Once the top-ten movies list is completed, the participants hit a button to request movie recommendations. Next, they receive a list of twenty recommended movies. Participants have to rate the recommendations (ranging from 1 to 10 stars), indicate whether they have already watched the suggested movies and whether they would watch it.

3.3.2. Second part

The second part of the experiment was presented in the form of a scripted (i.e., storyboarded) conversational movie recommender system. The recorded conversations between the participants and the conversational RS were built with BotMock. The conversations were personalized according to the high or low degree of Openness to Experience of the participants -- as was previously collected with the help of a validated personality test (Donnellan, Oswald, Baird, & Lucas, 2006). Openness to Experience was taken into consideration in playing a more personalized and engaging service, that provides highly diversified movies suggestions and used different language based on the personality. Regarding the words used in the scripted conversations, depending on a person's personality, some words were used (or, if applicable, not used) in preference to others. These words were taken from the well-established personality categories for Openness to Experience of the LIWC dictionary (Pennebaker et al., 2001), which would then be utilized to build the conversations. The study of Yarkoni (2010) also provided important insights into the most suitable words to adopt in conversations personalized for high and low Openness to Experience. A similar approach was also taken in a recent study in the research domain of recommender systems, in which conversations for conversational agents were scripted to approach personalization (Smith, Dennis, Masthoff, & Tintarev, 2018). Below is an example of the two types of conversations depending on whether the participant was high or low on Openness to Experience. Each participant received a personalized scripted conversation with different movies recommendations based on their Openness to Experience and to the top-10 movie list they provided during the first part of the experiment.

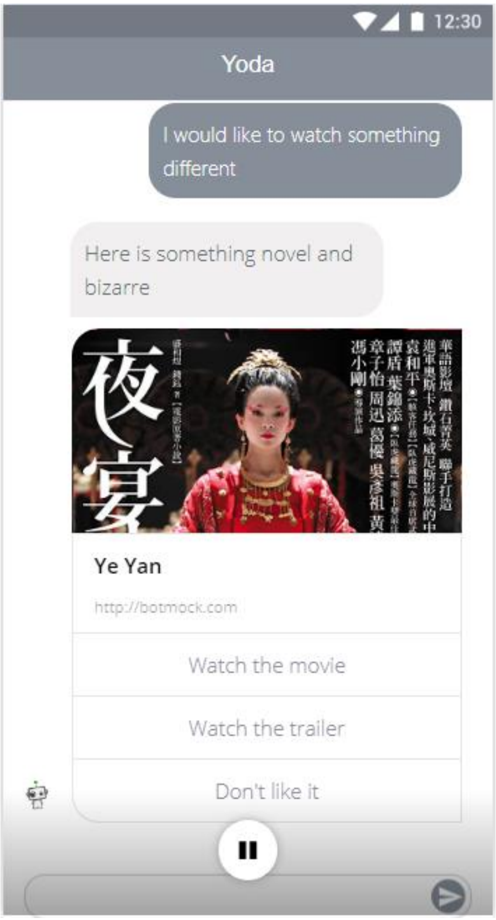


Figure 5: Conversational RS people High on Openness to Experience

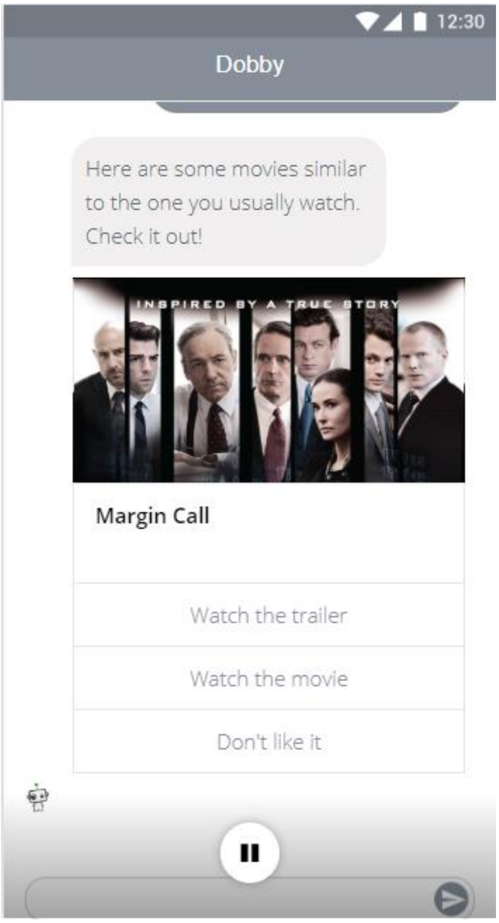


Figure 6: Conversational RS people Low on Openness to Experience

Also, the movie recommendations themselves were tailored towards to the Openness of Experience of the participants. MovieExplorer is the first functioning tool in the recommender systems research community to offer users both accurate and serendipitous (surprising) recommendations. This unique feature of MovieExplorer offered the possibility to customize the movie recommendations according to the personality of the participants. Specifically, the conversational RS gave a larger number of accurate recommendations for people low in Openness to Experience (thus different movies that were similar to these people's taste in movies as captured in their previously registered movie top 10) and more serendipitous movies (thus movies that were surprisingly different to these people's taste in movies as enclosed in their movie top 10) to people high in Openness to Experience.

3.4 Procedures

All participants received a URL link to join the first phase of the experiment hosted on the MovieExplorer website. On the landing page, participants were welcomed and provided with an introduction of the first part of the experiment, containing the instructions as follows:

"*Welcome to MovieExplorer*
Hello!
How do you choose what movie to watch? The purpose of this study is to improve the ways in which algorithms provide movie recommendations. In the study, you will be asked to provide a top-10 of

movies. In return, you will receive some movie recommendations that we hope you'll like. You can help the research on this topic by taking about 20 minutes to complete the following survey."

First, a pre-survey on Openness to Experience and demographics were presented to gain information about the participants' personality. Note that this information, though presented at the beginning of the first phase, would only be utilized during the second phase. Next, the participants were asked to provide their own top-10 movies from the MovieExplorer platform database. Also, they had to specify whether they had watched the movie before and indicate their appreciation for the movie with a value ranging from 1 to 10 stars. Once the participant had completed the movie top 10 and hit the "submit" button, the system gave a sequence of 20 automated movie recommendations based on the participants' own movie preferences. For each of these recommendations, the participants had to again indicate whether they had already watched the movie, they would watch it and rate the recommended movie from 1 to 10 stars.
The second phase of data collection and measurement was guided by personalized scripted conversations. As a first step towards personalization, individual scores on Openness to Experience were converted into low and high values with the help of a median split. Second, each participant selected for the second part received an invitation to join the second part of the experiment. In this phase, participants were exposed to a 59 second recorded (storyboarded) conversation. The participants were asked to watch and read the video conversation, and to imagine like they were the ones interacting with the chatbot. This approach and these instructions were similar to the user-as-a-wizard method that is often used for the development of personality-based conversational recommendations (Smith, Dennis, Masthoff, & Tintarev, 2018; Tintarev, Dennis, & Masthoff, 2013). Participants received conversations based on their own (high or low) answering pattern on Openness to Experience (i.e., their low or high median split value), retrieved from the pre-survey conducted in the first phase. Based on this, the participants high in Openness to Experience received movie recommendations that were more diverse and serendipitous than indicated in their own top-10 movie list from the first phase. Specifically, 70%[5] of the suggested movies would be original (surprising), while the remaining 30% would be accurate (similar to one's preferences). Thus, the larger part of movie recommendations would be characterized by serendipity. In contrast, participants scoring low in Openness to Experience would receive movie recommendations more accurate and in line with their tastes and frame of reference. Again, this was deduced from their own top-10 movie list provided during the first phase. Mirroring the high Openness to Experience condition, people low on Openness would receive 70% of precise and accurate movie recommendations, while only 30% of the movie recommendations would be serendipitous. As in the MovieExplorer part at the first phase, participants had to express their appreciation for each newly recommended movie from 1 to 10 stars. Also, as in phase one, participants were asked to indicate if they had already watched the recommended movie.

Finally, after the first two phases were completed, the participants were directed to a seven-item post-survey containing more specific questions regarding the way in which the two types of recommendations had been experienced and evaluated. Since each phase had a unique way of providing recommendations, the objective of the survey was to gauge the user satisfaction at each phase. In such a way as to determine whether the conversational RS increases user satisfaction.

[5] The present study is the first one trying to adapt the numbers of accurate and serendipitous recommendations to the user's Openness to Experience. The author considered it appropriate to give more serendipitous recommendations to high on Openness to Experience people, believing 70% to be a good percentage to differentiate.

3.5 Measures

3.5.1 Openness to Experience

Openness to Experience is one of the dimensions in the Big Five personality traits. Many assessment tools have been created to measure the Big Five personality traits. In this research, the Mini-IPIP scale (Donnellan, Oswald, Baird, & Lucas, 2006) was used to evaluate the trait Openness to Experience. In general, the Mini-IPIP consists of 20 questions, in groups of four corresponding to each of the personalities. Given that the current research takes into consideration only the Openness to Experience dimension, only the questions corresponding to Openness to Experience (shown in Tab. 8) were used in the pre-survey. Each question was rated on a 5-point Likert scale anchored at 1 (strongly disagree) and 5 (strongly agree). Finally, the responses were collated to assess the overall (high or low) degree of Openness to Experience of the participants.

Item
Openness
Have a vivid imagination
Am not interested in abstract ideas
Have a difficult understanding abstract ideas
Do not have a good imagination

Table 8: Mini IPIP (Donnellan, Oswald, Baird, & Lucas, 2006).

3.5.2 Openness to Experience-based recommendations

MovieExplorer offers recommendations based on four different algorithms (MatrixFAC400, BPRMF, ItemBasedKNN and MostPopular). The MatrixFAC400 and BPRMF give serendipitous recommendations, the ItemBasedKNN provides accurate suggestions and the MostPopular advises the well-liked movies. After measuring the participants' Openness to Experience, the correlation between the serendipity/accuracy in recommendations and the degree of Openness to Experience of the participants was measured. After seeing that there is a correlation, this result was used in the conversational recommender system, which was designed to give different recommendations depending on the personality of the participants.

The conversational RS incorporated Openness to Experience to adjust the degree of diversity within the set of recommendations. In this way, the recommendations were personalized to the individual users' needs. The diversity-adjusting strategy based on personality features was inspired by studies conducted by Smit et al. (2018) and Chen et al. (2013). One of the findings of Smit et al. (2018) is that people with a high level of Openness to Experience are more willing to vary their movie genres. As for the variation of actors, Chen et al. (2013) studied that it also is positively correlated to Openness to Experience. Therefore, in the current study more genre and actor variations were offered to people high on Openness to Experience. The insight of previous literature was used together with the algorithms from MovieExplorer to offer people in the conversational RS stage customized movie recommendations. It has been possible to customize the number of accurate and serendipitous movies according to the degree of Openness to Experience of the participants. In such a way that the difference was not only in the way the recommendations were provided but also in the recommendations themselves.

3.5.3 Build the conversations based on Openness to Experience

The scripted conversations were personalized to match the high or low degree of Openness to Experience of the participants. The difference in personalization, therefore, was not only in the choice of the algorithmic set of movie recommendations, which were more or less accurate and serendipitous depending on the personality profile (further details in Sec. 3.5.3) but also in the choice of the specific words (vocabulary) used in the scripted conversations.

There is a lot of evidence in the literature for a systematic association between Openness to Experience and individual differences in word use. These associations were first identified in the work of Pennebaker and colleagues (2001). The scripted conversations were based on the study of Yarkoni (2010) and of Masthoff and colleagues (Smith, Dennis, Masthoff, & Tintarev, 2018; Tintarev, Dennis, & Masthoff, 2013), which also link Openness to Experience to the words' usage. Analyzing the previous literature gave an insight into the second sub research question, providing an answer on how to customize the conversations according to the different degree of Openness to Experience. Specifically, Table 9 shows several words, that appear in the scripted conversations, and their correlation to Openness to Experience according to Yarkoni (2010).

Openness to Experience and word-level correlations
feeling (0.28), novel (0.29), bizarre (0.32), afterwards (-0.24), let (-0.24), am (0.22), time (-0.24)

Table 9: Openness to Experience and word correlation (Yarkoni, 2010)

3.5.4 User satisfaction

User satisfaction is well-documented in the literature, and it is often described as an emotional reaction to experiences associated with a particular item or service (Westbrook & Reilly, 1983). Accordingly, this study focused on analyzing the participants' responses to the movie recommendations received and the (RS) service offered. As mentioned earlier, user satisfaction with the recommended movies was measured twice. After the first phase, user satisfaction was measured each time the participants had received a new movie suggestion in MovieExplorer on a 10-point rating scale (1 being "not at all satisfactory", 10 being "very satisfactory"). At the second phase, satisfaction was measured in the same way, but then after each time, a participant had received movie recommendations through conversation. – and, again, satisfaction was measured on a 10-point rating scale. In other words, satisfaction with the recommendations provided by the two different RS tools was measured. Finally, a seven-question self-report survey was provided to assess if the participants had preferred a conversational RS over a traditional (automated) RS. More specifically, the satisfaction of the user with a different way of providing recommendations was also assessed after usage.

4. Results

This section presents the results obtained from the two experiments introduced in the previous chapter. First, the description of the sample and its satisfaction with the recommendations provided by MovieExplorer is carried out. Next, the relationship between Openness to Experience and the accuracy vs serendipity in recommendations is examined. Then, user satisfaction with both conversational RS and traditional RS is investigated. Finally, the lessons learned during the research will be raised.

4.1 Data cleaning

The data were collected through the interactive experiment on MovieExplorer, and were coded, and keyed in prior to data analysis. The data from MovieExplorer were received in a .csv format and they were read and analyzed statistically with the help of Python. Pandas, the open-source Python library, was used to open, analyze and read the .csv files. Next, Numpy Python library provided the basic tools to compute multidimensional arrays. SciPy Python library was used for statistical computing. Matplotlib is a Python 2D plotting library, which was used to produce the plots. Outliers (N = 5; 4.43%) were removed before conducting any statistical analysis.

4.2 Sample characteristics

Recall that the overall sample characteristics of the participants in study were (58 men, 45 women, 10 others; *Mage* = 24.76 years, *SD* = 3.32). During the first experiment (with Movie Explorer) participants were asked to fill in a questionnaire in order to assess their degree of Openness to Experience. The results of this validated (Mini-IPIP) scale were used to build the high and low Openness to Experience scores with the help of a median split (Md = 4.00). Every overall participant's score lower or equal to the median was assigned the value 0 (and labelled "low OE"); instead, every overall score higher than the median was assigned the value 1 (and labelled "high OE"). Figure 7 shows how the full (Movie Explorer) sample was divided based on high vs. low Openness to Experience. This factor was used to test the extent to which high and low scores for Openness to Experience lead to different satisfaction levels with movie recommendations received. Please note that later, the high or low scores on Openness to Experience were also used to invite participants to partake in the second (conversational recommendation) stage of the study- in which based on the degree of Openness to Experience, participants receive different movie suggestions.

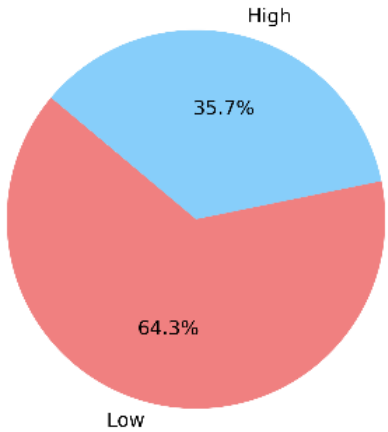


Figure 7: Openness to Experience full sample

4.3 Descriptive Analysis of the MovieExplorer platform

This section explores how the participants appreciated the automated movie recommendations given by MovieExplorer. Figure 8 shows the overall satisfaction with automated movie recommendations based on participant ratings (Mn = 6.94; SD = 0.78). In general, participants are satisfied with recommendations provided by MovieExplorer.

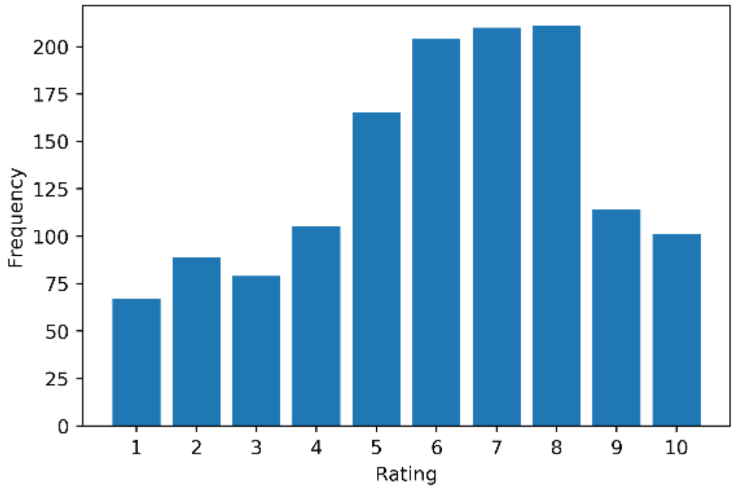


Figure 8: Overall satisfaction ratings

After each movie recommendation received, participants had been asked whether they would like to watch that movie or not. Figure 9 shows that the participants, in general, were favorable towards the automated movie recommendations received. Overall, most participants indicated that they would like to watch the movies suggested by MovieExplorer (compared to ''no'' or ''unknown''). The chi-square test on the aggregated set of answers showed that the difference between people that wished to watch the provided recommendations and those who did not, or were not sure about it, was significant ($\chi 2 = 165.42$, $p < .0001$).

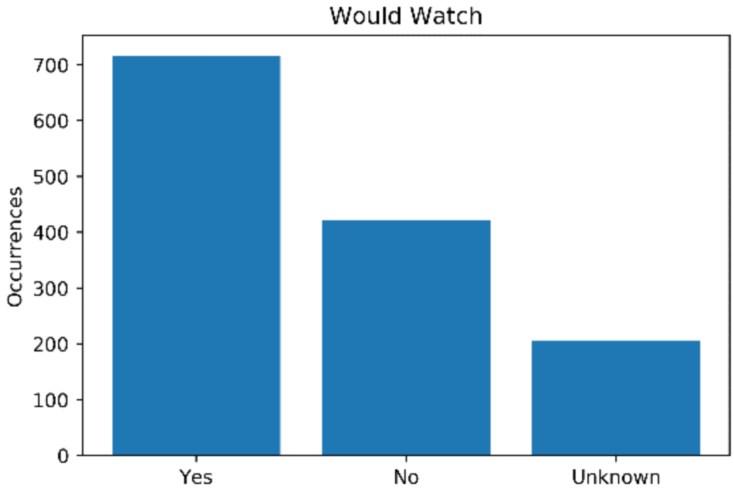


Figure 9: Overall satisfaction

Also explored was the possibility that overall satisfaction with movie recommendations in the sample was different for people high on Openness to Experience and people low on Openness to Experience.

A chi-square test of independence was calculated to measure to what extent people low on Openness to Experience who voted "yes" differed from those "no" and "unknown" ($\chi 2 = 57.27$, $p < .0001$). Figure 10 offers the descriptive plot for low Openness to Experience, showing that unlike from what it may appear in the figure there is a significant difference from people who wanted to watch the movies and those who do not want to or are not sure about it.

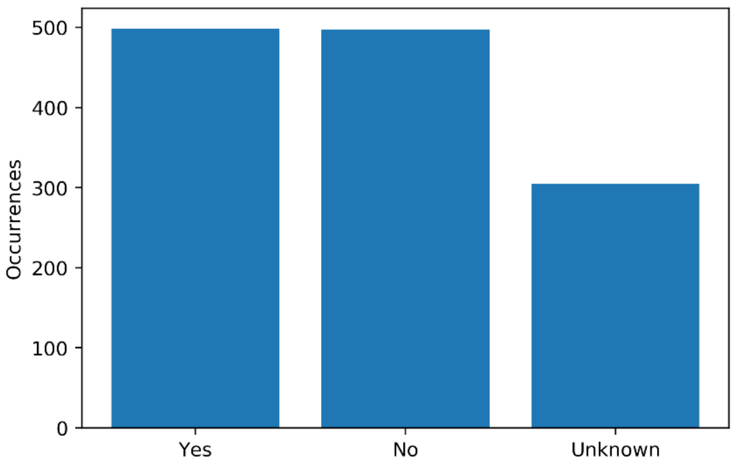


Figure 10: Satisfaction Low Openness to Experience

The same analysis was conducted for people high on Openness to Experience. The chi-square on the difference between the ratings "yes" and "no" and "unknown" also was significant for people high on Openness ($\chi 2 = 173.94$, $p < .0001$). Figure 11 offers a descriptive plot, which shows that people high on Openness to Experience are willing to watch the recommended movies. This is in line with the fact that people high on Openness to Experience are more curious and eager for new experiences.

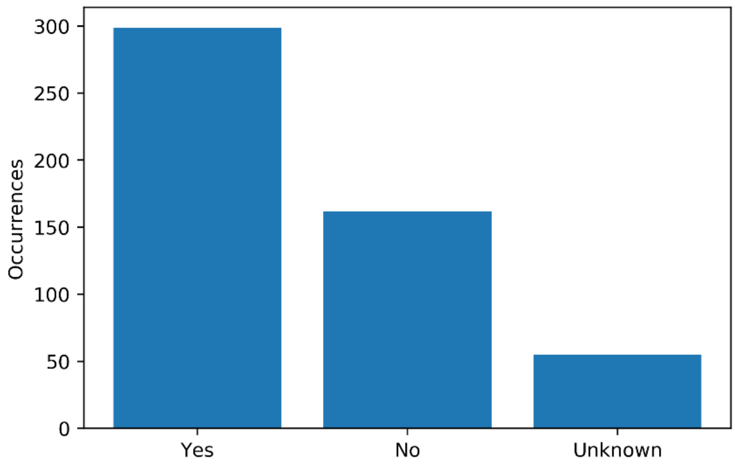


Figure 11: Satisfaction High Openness to Experience

These results describe how the participants in the sample related to the MovieExplorer platform and the movie recommendations they received in the first stage of the experiment. Furthermore, they are essential for providing an answer to our main research question and hypotheses stating that there is a difference between user satisfaction with a traditional recommender system and a conversational recommender system.

4.4 Openness to Experience and diversity in recommendations

In this paragraph, the first sub-question of this study is tested and analyzed – i.e., the relationship between the degree of Openness to Experience and the preference for accurate or serendipitous movie recommendations. The aforementioned median split on high vs. low Openness to Experience was used to explore the appreciation for accurate and serendipitous movie recommendations based on personality.

Before conducting any statistical testing on accurate and serendipitous recommendations, the distributions of ratings for accurate recommendations (Fig. 12) and serendipitous (Fig. 13) were analyzed. The normal test was conducted to test whether the sample of the ratings for accurate recommendations ($s^2 + k^2 = 10.82$, $p = 0.004$) and the sample of the ratings for serendipitous recommendations ($s^2 + k^2 = 6.43$, $p = 0.04$) differ from a normal distribution.

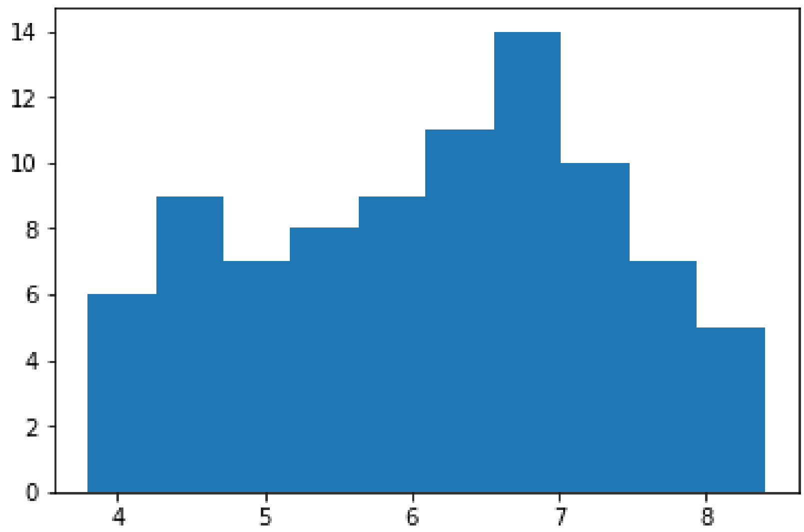


Figure 12: Continuous distribution ratings accurate recommendations

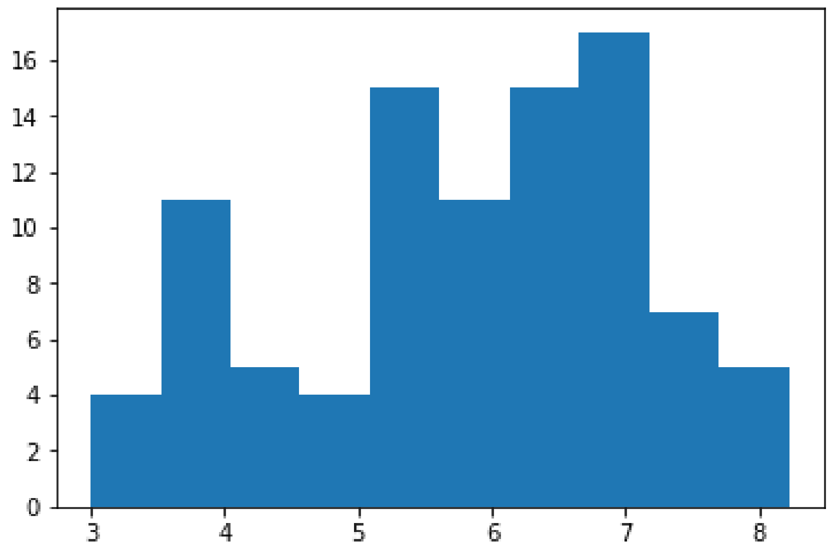


Figure 13: Continuous distribution ratings serendipitous recommendations

Analyzing the distribution of the samples was fundamental to understand how to proceed with the statistical analysis. The ratings that people gave to the recommended movies are rank order data, meaning that it was necessary to opt for a non-parametric procedure. The Mann-Whitney test is a common test to use in this case. The downside of using a Mann-Whitney test, in this case, is that it has less statistical power than in a parametric test (known as 'underpowered test'). But, since the data is ranked data, it was necessary to stick with Whitney test. Therefore, Type-I errors ("the number of times a test will find a significant effect when there is no effect to find" (Field, 2013, p.283)) may have occurred.

In Figure 14, the blue dots represent the mean score of the ratings that the participants gave to the accurate movies in Movie Explorer. The blue dots are on two different columns. The column at zero represents the low on Openness to Experience participants, while the column valued at one represents the participants high on Openness to Experience. Figure 14 shows that, overall, participants low on

Openness to Experience better appreciated accurate movie recommendations than participants high on Openness to Experience. The trend was in the theoretically assumed direction, but not statistically significant (Mann-Whitney U= 981.00, p= 0.37).

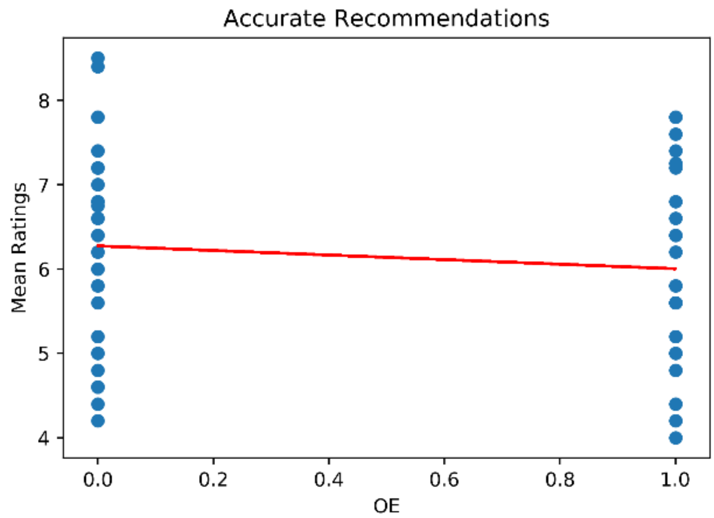


Figure 14: Accurate recommendations Low vs high Openness to Experience

The same analysis was carried out for serendipitous movie recommendations. In Figure 15, the blue dots represent the mean ratings the participants gave to the serendipitous movies in Movie Explorer. As before: the blue dots correspond to the appreciation of serendipitous movies based on low and high on Openness to Experience. As theoretically assumed, participants high on Openness to Experience better appreciated serendipitous movie recommendations than participants low on Openness to Experience. Yet, there was not sufficient statistical relevance (Mann Whitney U= 948.00, p= 0.27).

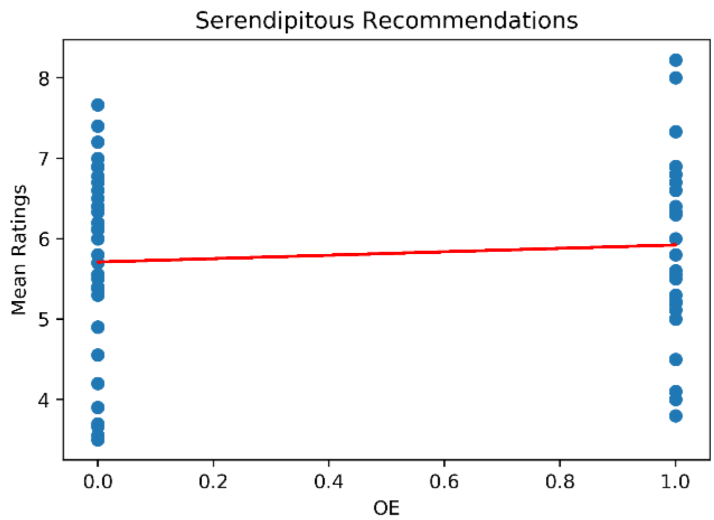


Figure 15: Serendipitous recommendations Low vs high Openness to Experience

Because the trends reported above were in the theoretically assumed direction, an additional test was devised, based on another difference measure. This time, the difference between the ratings of the movies suggested with an accurate algorithm and a serendipitous algorithm (i.e., overall accuracy ratings *MINUS* overall serendipitous ratings) was analyzed. Thus, a Mann–Whitney test was conducted twice to examine if this proxy measure was the same for (1) people low on Openness to Experience and for (2) people high on Openness to Experience.

For low Openness to Experience, the Mann–Whitney test was significant (Mann–Whitney U= 742.00, p-value< 0.01), which indicated that the distribution of the overall ratings of accurate movies and serendipitous movies are not equal for people low on Openness to Experience. For people low on Openness to Experience, the ratings of the accurate recommendations were significantly higher than the ratings of the serendipitous ones.

For high Openness to Experience, the Mann–Whitney test was marginally significant (Mann–Whitney U= 297.5, p < 0.10). This indicated that- formally the distribution of the overall ratings of accurate movies and serendipitous movies are approximately equal for people high on Openness to Experience. People high on Openness to Experience could under some conditions rate accurate movies recommendations and serendipitous ones differently. Still, for the Mann–Whitney test there was not enough evidence against the null hypothesis.

To conclude, the analysis carried out so far answer to the first sub-question. The results highlight that overall people low on Openness to Experience significantly favor accurate suggestions over serendipitous ones. While people high on Openness to experience do not show a significant preference, meaning that they are willing to watch in the same way accurate and serendipitous movies suggestions. However, when analyzing the ratings that people high on Openness to Experience and low on Openness to Experience have given to accurate and serendipitous movies, this difference is not significant. This shows that there is not enough statistical evidence to accept the hypothesis that high on Openness to Experience and low on Openness to Experience people voted in different ways for accurate movies and serendipitous movies.

4.5 User satisfaction

The main research question inquires whether conversational RS increases user satisfaction. In order to answer to the above mentioned, it has been analyzed if people are more satisfied with recommendations received from a personalized (conversational) movie recommender system than with those provided by an automated movie recommender system. Table 10 summarizes the extent to which participants were satisfied with the proactively delivered movie recommendations of the two types of RS. The table shows the mean of the answers given by the participants for each question, where 1 corresponded to "*not at all true for me*" and 10 to "*very true for me*".

Questions	Mean score	Standard deviation
I preferred the movies I received from the automated RS over those of the conversational RS	4.47	2.09
I liked the movie recommendations of the automated RS more than those of the conversational RS	4.65	1.71
The automated RS knew better than the conversational RS how to provide movie recommendations	4.41	1.94
The conversational RS fulfilled my expectations better than the automated RS	6.94	1.66
I felt that the automated RS knew me better than the conversational RS	4.53	2.00
Movie suggestions made by the automated RS were more accurate than those made by the conversational RS	4.59	2.20
I felt more connected with the conversational RS than with the automated RS	6.76	2.21

Table 10: Users satisfaction questionnaire

The ratings the participants gave to the movies were used to analyze the users' satisfaction. The majority of the participants preferred the conversational RS as a tool to receive suggestions more than the traditional RS. The users' satisfaction with the two RS and with the movies suggested by the two latter is now analyzed more accurately.
Table 11 shows in more detail how the participants voted for the automated RS and conversational RS. Further analysis of these data and will follow in the next section.

	Full sample	Low Openness to Experience	High Openness to Experience
% of participants that have preferred the conversational RS both as a tool and as a recommendations provided	0.65	0.50	0.78
% of participants that have preferred the conversational RS as a tool but rated higher the recommendations provided by the automated RS	0.18	0.25	0.11
% of participants that have preferred the automated RS both as a tool and as a recommendations provided	0.12	0.13	0.11
% of participants that have preferred the automated RS as a tool but rated higher the recommendations provided by the conversational RS	0.06	0.13	0

Table 11: analysis ratings automated RS and conversational RS

Figure 16 shows that the majority of participants in the present study preferred the conversational RS (82.4%) over the traditional RS (17.6%) as a tool to receive movie suggestions. A chi-square test of independence was calculated to measure to what extent automated RS differs from conversational RS as a tool ($\chi 2 = 98.32$, $p < .0001$). This confirmed that conversational RS enhance user satisfaction.

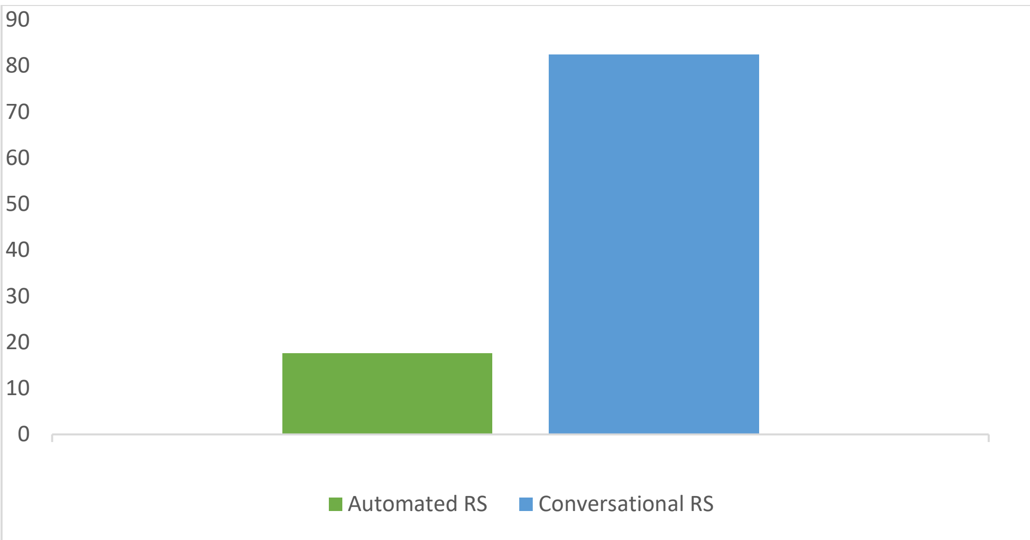


Figure 16: Automated RS vs Conversational Rs as tool full sample (N=22)

Regarding the items recommended, Figure 15 illustrates that the majority of participants in the present study preferred the movie suggestions provided by the conversational RS (70.6%) over the ones suggested by the automated RS (29.4%).

An independent T-test ($t = 1.31$, $p = 0.20$) was conducted to check if the recommendations provided by the automated recommender system ($M = 6.27$, $SD = 1.15$) differed significantly from the ones provided by the conversational RS (M = 6.67, SD = 0.56). Even though the pattern was in the predicted direction, this result was not significant. This is due to the fact that overall people preferred the recommendations provided by the conversational RS, but the ratings given to the movies suggested by the conversational RS were not significantly higher compared to the ones given to the automated RS.

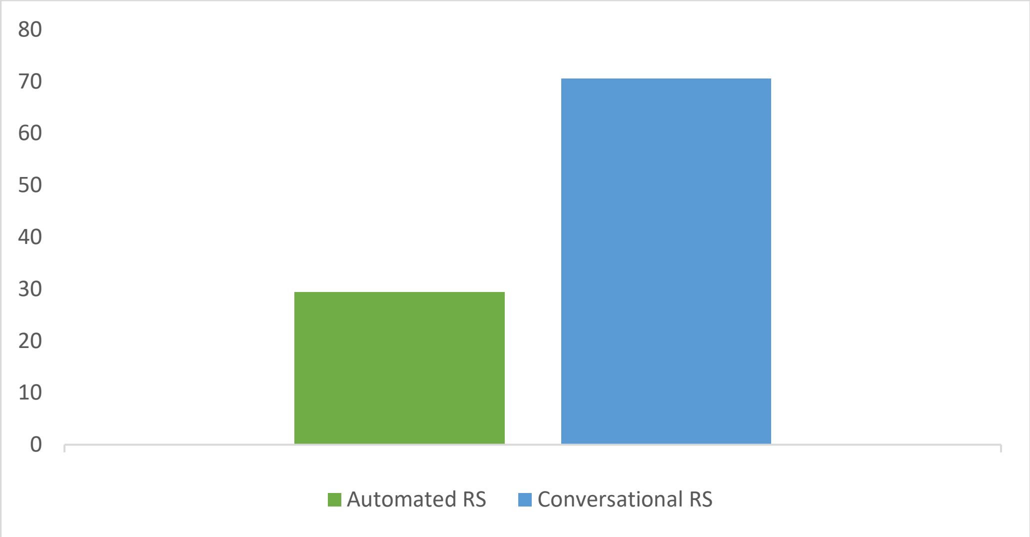


Figure 17: Automated RS vs Conversational Rs recommendations full sample (N=22)

To explore the impact of personalization of RS based on personality, the same analyses were repeated by differentiating the between people High on Openness to Experience and Low on Openness to Experience.

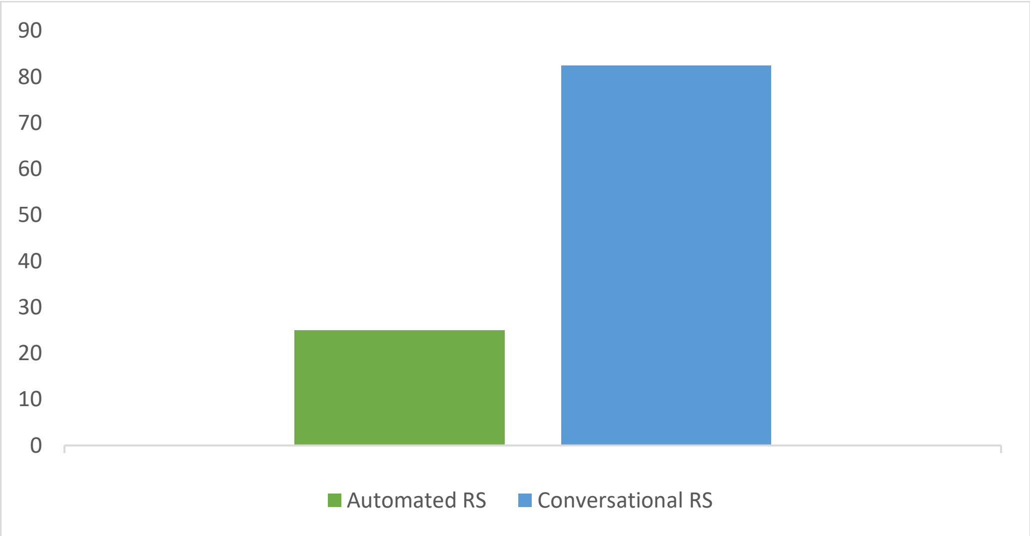


Figure 18: Automated RS vs Conversational Rs as tool Low Openness to Experience

A chi-square test of independence showed that the difference between people low on Openness to experience who voted the automated recommender system and the conversational RS was significant ($\chi 2 = 53.67$, $p < .0001$). This finding indicates that people prefer to interact with the computer on human terms rather than receiving the recommendations in a traditional way.

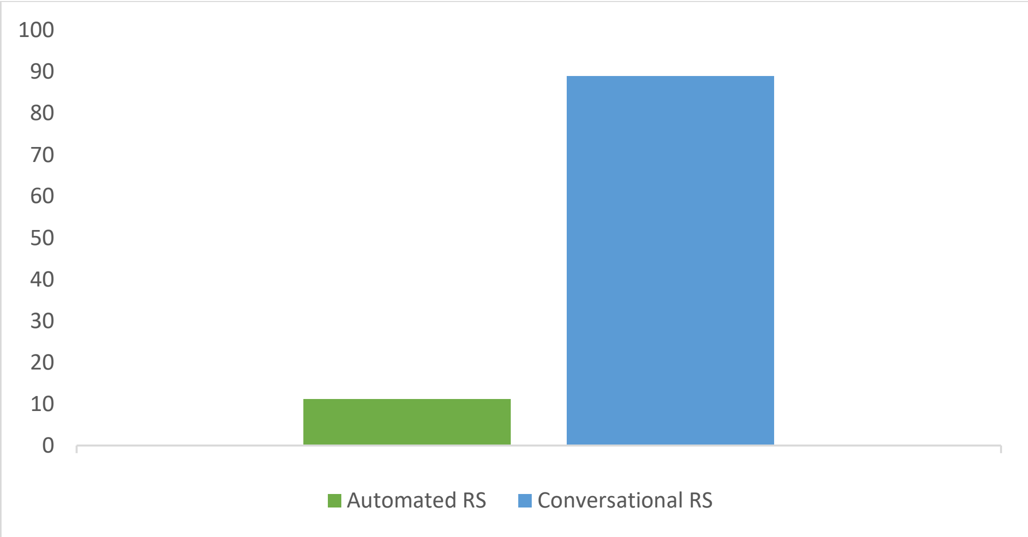


Figure 19: Automated RS vs Conversational Rs as tool High Openness to Experience

A similar pattern was observed for people high on Openness to Experience. Ratings for the automated RS differed significantly significantly from the ones of conversational RS also for people High on Openness to Experience. ($\chi 2 = 44.64$, $p < .0001$). This means that people high on Openness to Experience prefer to receive suggestions while conversing with the RS and to have the chance to express opinions about tentative recommendations.

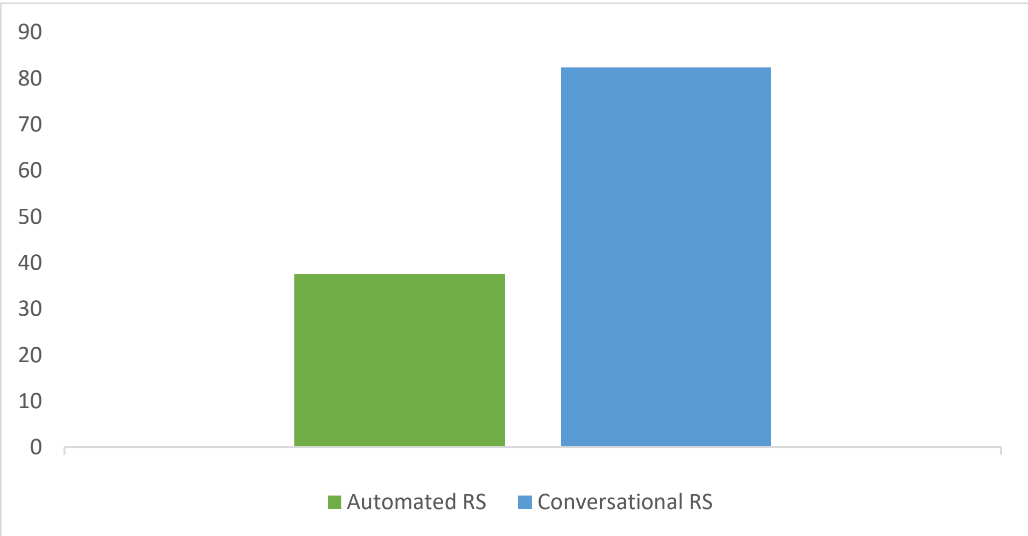


Figure 20: Automated RS vs Conversational Rs recommendations Low Openness to Experience

As reported above, in general, no significant difference was found for the way in which participants in the study appreciated the recommended items from the two RS systems. An independent T-test ($t = 1.15$, $p = 0.27$) was conducted to check if the recommendations provided by the automated recommender ($M = 5.60$, $SD = 0.83$) system differed significantly from the ones provided by the conversational RS ($M = 6.93$, $SD= 0.28$) for people Low on Openness to Experience.

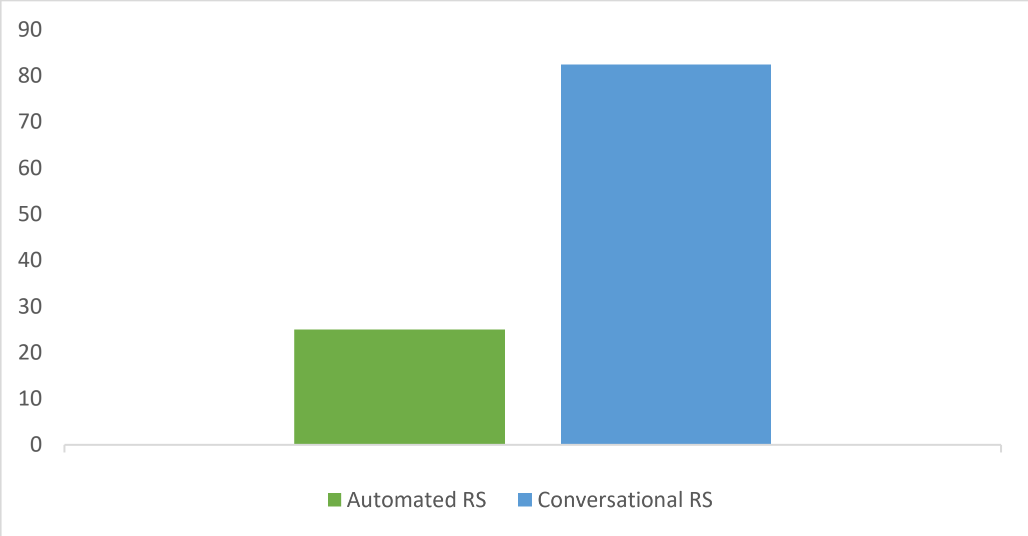


Figure 21: Automated RS vs Conversational Rs recommendations High Openness to Experience

Again, an independent T-test (t= 0.77, p = 0.45) was conducted to assess if the recommendations provided by the automated recommender (M = 6.15, SD = 1.34) system differed significantly from the ones provided by the conversational RS (M = 6.54, SD = 0.70) for people Low on Openness to Experience.

The results of Section 4.4 provide an answer to the third sub-question. It has been demonstrated that overall participants prefer the conversational RS than the automated RS. Although the participants did not show a significant difference in the votes given to the movies suggested by the automated RS and the conversational RS, they preferred the conversational RS as a tool. It is possible to generalize these findings and answer the main sub research question by stating that personalized conversational RS can contribute to improving user satisfaction.

5. Discussion

The present study explored to what extent the satisfaction of users in RS was dependent on (automated vs. conversational) RS mode of interaction. It also investigated if conversational RS was influenced by the Openness to Experience of people, the accuracy and serendipity of recommendations. It was hypothesized that a conversational RS that gives suggestions based on people's high or low Openness to Experience increase users' satisfaction.

5.1 Scientific relevance

In this study, several theoretical contributions were made. First, an overview of the state-of-the-art of chatbots, and of conversational RSs, in particular, was provided (Christakopoulou, Radlinski & Hofmann, 2016; Sun & Zhang, 2018). The key outcome of this study is that people were more satisfied with conversational (vs. automated) RS. The results therefore especially contribute to the awareness that conversational RSs exert a positive impact on user satisfaction. That such a positive effect emerged clearly since overall participants preferred the conversational RS compared to the automated one. This empirically gives credits to the idea raised by Anelli et al. (2018) that researchers should go beyond the traditional accuracy goal of the automated recommender system and that RS should invite the users to have an interactive interaction with the RS to guide the recommender in making further recommendations. This research confirms this idea and hints towards the opportunity for conversational RS to become the next level of human-computer interaction.

Furthermore, the study gives insight into the effect of online personalization. To the author's knowledge, the current research is the first-ever known academic attempt to adapt a conversational RS and the nature of the recommended items to users' Openness to Experience. Thus far, only the conversations themselves had been modified to a person's high or low Openness to Experience (Tintarev, Dennis, & Masthoff, 2013). In the present study also the recommended items themselves were adapted to a person's personality. Specifically, the relationship between the degree of Openness to Experience and serendipitous/accurate suggestions was novel, and never explored before. As a promising first finding, the present study demonstrated that overall people low on Openness to Experience significantly prefer accurate suggestions rather than serendipitous ones. Interestingly, while people high on Openness to Experience do not have a significant preference regarding accurate or serendipitous recommendations, the present study reported a trend towards a preference for serendipity. In other words, people low on Openness to Experience have a strong preference to watch movies that are similar to the once they have previously watched; people high on Openness to experience are willing to watch in the same way both movies more similar to their tastes as well as original movies. This suggests that serendipitous recommendations also are appreciated by high Openness to Experience people. However, when zooming in the personality difference, this effect disappears. People high and low on Openness displayed a similar response pattern on recommendations received. Yet, to personalize recommendations Openness to Experience may not be the most appropriate measure. It would be stimulating to investigate in future research another personality measure (e.g. Curiosity and Exploratory Inventory) as a way to go for personalized recommendations. This construct was defined as a major personality trait to explore individual differences in curiosity (Litman & Silvia, 2006) and as preferences for novel, complex, or ambiguous stimuli (Pearson, 1970). Since Curiosity and Exploratory Inventory is widely recognized as a trait related to the desire of new knowledge and experience (Litman & Silvia, 2006), people that have this kind of personality might appreciate serendipitous recommendations more than people that do not show this personality trait.

5.2 Practical relevance

The findings of this research have significant implications for future organizational practice. It was demonstrated in this research that users of a movie platform would prefer to receive suggestions from a conversational RS instead of an automated one (as is nowadays also used in the major movie platforms such as Netflix; see Amatriain & Basilico, 2015). People, however, are nowadays demanding more natural conversation, they want context and humanlike understanding. Thus, conversational RS has the potential to be the new mode of interaction with users on an ecommerce or multimedia platform. This highlights that organizations should invest more in natural language development technology with the aim of building systems that will enable users to access services in a conversational way (McTear, 2016). In such a way that the users can continually interact with the conversational RS to specify their requirements and refine their preferences (Tintarev & Masthoff, 2015). Conversational RS can help users to articulate short and long term preferences and to express an opinion about the given recommendations in order to guide the future ones (Anelli, Basile, Bridge, Di Noia, Lops, Musto, & Zanker, 2018).

Second, developers have to focus on building a conversational RS that is not only able to capture user preferences as input but is also able to communicate with them smoothly and without misunderstandings. Rather than continuing to develop increasingly complex algorithms to explore users preferences, designers should focus on the interaction between user and computer (Nass & Moon, 2000). In order to ensure clear and effective communication, it is also recommended to use different words depending on the personality. Adapting the language to the personality can help the user to feel more understood and stimulate the trust towards the system. However, the customization of the language has been a minority component of the experiment conducted. Therefore, further experiments and research are necessary to understand the relationship between language, users' personalities and reactions to the recommendations received. Individual differences in personality can be linked to individual differences in linguistic style (Yarkoni, 2010), and this relationship could be used more extensively by conversational RSs developers to build conversational touch-points. With this regard, several linguistic inventory and word count (LIWC) software packages have been created to enhance the human computer interaction and to enable a behavioral analysis of the users. The application prospect of developing such conversational RS with personalized recommendations as well as linguistic customization is believed to provide an improved experience and satisfaction with the users (Zanker, Rook, & Jannach, 2019).

5.3 Limitations

The results reported hitherto should be considered in the light of some limitations.
Firstly, the results are derived from a single study. One serious drawback of single studies is the lack of generalization (Driskell & Salas, 1992). Secondly, the small size of the second part of the experiment does not allow for a clear generalized statement for our findings. Still, the small sample did -in the present study- not negate the recognition that conversational RS can play an important role in enhancing users' satisfaction.

The study was further limited as the experiments were run on a voluntary basis, with no incentives or prize. Thus, the experiment was based on the good faith of the people to respond seriously and loyally. Participants had the free choice of deciding the experiment anywhere and at any time according to their convenience. Uncontrolled experiments are more keen to be less attentive and more prone to inaccuracy than subjects in a lab with an experimenter (Oppenheimer, Meyvis, & Davidenko, 2009). The results might have been different if performed in a full-fledged laboratory (Sekaran & Bougie, 2016).

Moreover, the second part of the experiment was conducted with scripted conversations on movies. Instead of having real interaction with the conversational RS, participants were asked to only observe the conversation. So, users did not freely interact with the system as they would do in real life, that might have yielded unexpected effects to real experiments. User Interface mockups are adopted to depict scenarios where the real-life data is employed for illustrating the use case instead of abstract descriptions (Urbieta, Torres, Rivero, Rossi, & Dominguez-Mayo, 2018). Yet, conversational RS have several features (e.g. active learning in conversational recommenders) that cannot be valued by only looking at the mockup.

In this research, the user satisfaction with a traditional RS and a conversational RS was compared. However, the visual and information design contribute to the user’s experience (Cyr, Kindra & Dash, 2008) of the RSs and consequently impact their satisfaction. So, if the design of the two RSs had been different, probably even the users' preferences for the two systems would have been different.

The duration of the research further challenged the current study, as it was relatively short; participants were observed over a moderately short period of time. They could not really acquire knowledge from experience and come to trust the RS on its merits. On the other hand, also the RSs work that the more they are adopted by the users, the more they know the preference of the people and refine the suggestions. So, this set up did not make possible to analyze how the conversational RS might have come to be used over time. This is something to consider for future experiments and iterations with RSs.

5.6 Future work

Research on conversational RS is still in its infancy. There are numerous challenges to be tackled by future research; some of these are discussed in the following.

First, other methods of experimentation are welcomed. In the present study, the online experiment was conducted with a mockup - therefore, in a protected environment, free of misunderstandings between user and bots. Usually, the number of domains about which a chatbot can converse is limited and this can create misunderstandings with the user (Bakarov, Yadrintsev, & Sochenkov, 2018). In this context, the conversations were scripted, and the user was not free to switch to chit-chatting mode and talk about an unknown domain. Future research could pursuit doing a field experiment, where participants have the chance to interact freely with the conversational RS. It would be interesting to detect the difference between these two kinds of experiments. A field experiment would lead to a more accurate analysis because users can truly understand if they feel more connected with the conversational RS and if the conversation leads to misunderstandings. The “real-world” setting can provide an extent of generalizability (Sekaran & Bougie, 2016) to the results obtained in this study.

Furthermore, participants should be observed for a longer period of time. In fact, the more RSs are adopted, the more they retain information about previous interactions that are highly relevant for predicting future actions (Ricci, Rokach, & Shapira, 2015). In the long run, RSs can adapt real-time to the participants' responses and give more appropriate recommendations, while participants interacting for longer with the conversational RS could increase their trust and accept more willingly the recommendations provided. Repeated measurements allow the estimation of the variability within a case over time to provide the sources of information about user satisfaction (Hayes, & Blackledge, 1998). Besides, it gives both the RS the opportunity to learn more about the user's needs and wishes and the user the opportunity to trust the recommendations provided. A longitudinal study would also give the opportunity to the conversational RS to customize its language according to the user. As stated before, in the present study, the role of the words adopted, and the language used by the conversational RS was a minority component. Depending on whether the participants were low or high in Openness to Experience the conversational RS adopted specific words, based on the study of Yarkoni (2010) and of Masthoff and colleagues (Smith, Dennis, Masthoff, & Tintarev, 2018; Tintarev, Dennis, & Masthoff, 2013).

However, future research should need to focus more on the language adopted by the conversational RS, because that could be the key for the success of this technology. In order to capture people's conversations and behaviors and tailor the conversations to them, an Electronically Activated Recorder (EAR) could be adopted. The EAR is a technology for obtaining behavioral data in field research (Mehl, Pennebaker, Crow, Dabbs, & Price, 2001). Studying people under real circumstances, unobtrusively record snippets of natural language is important to assess users behaviors and personality. From recordings, real-word conversations can be transcribed, and words usage and social information related to the participants can be reliably coded and, further, used from the conversational RS to talk in a more personal way. In fact, these psychoinformatics can help to understand which language to use depending on the person (i.e. colloquial, formal) and which words arouse certain emotions to a particular individual (Yarkoni, 2010; 2012).

It is crucial to make recommendations according to the personality of the user (Nunes, Cerri & Blanc, 2008). The present work argued that the role conversational RSs could play on users' satisfaction could be tailored to recommendations based on Openness to Experience. In doing so, the present research focused only on the Openness to Experience to achieve personalization in conversations. This has not previously been done in the literature. However, it would be fascinating to explore also the impact Conscientiousness, Neuroticism, Agreeableness and Extraversion – the other dimensions of the Big Five or Five Factor Model of Personality -- could have in proactive recommendations. In particular, a collection of much larger samples of people would offer the chance to model personality trait clusters together with conversational RS and proactive recommendations (Rook, Sabic, & Zanker, 2018).

Recommendation systems are adopted by numerous Internet-focused companies in different application domains. This work has been conducted in the movie domain and demonstrated that users of a movie platform would prefer to interact with the RS. It would be interesting to investigate the effects of conversational RS on user satisfaction also in other domains. Today, almost every e-commerce site and application has incorporated recommendations to enhance user experience. Music platforms are also another active domain area for recommendations. Users of Apple's iTunes or Spotify are provided with personalized mixes and playlist (Amatriain & Basilico, 2015). Even social network companies have proposed several recommendation paths. Twitter, for instance, introduced the "Who to Follow" recommendation algorithm to suggest new social connections (Amatriain & Basilico, 2015). So, it would be stimulating to investigate if the preference for conversational RS applies to all domains or is confined to just a few of them.

A final opportunity for the prospective research for researchers and developers within the new world of conversational RS is to provide needed guidance on ethics and privacy (Følstad, & Brandtzæg, 2017). Any new technology entails new ethical implications. As conversational RSs are still at its infancy it is important to draw a clear line between intentions in design and the eventual use of technology (Albrechtslund, 2007). Designers must pay particular attention to the potential uses not intended in the design process, as these may be ethically undesired. For instance, managers could misuse these conversational RSs and align users' interests to their business needs, instead of helping them to overcome the problem of overloading information. Interactive persuasive technologies, such as the conversational RS, adapt to individuals (Kaptein & Eckles, 2010). Practitioners should be concerned with adaptive persuasive systems that adjust to individual differences in the effectiveness of particular means, as influencing strategies. In this context, it is crucial for designers of conversational RSs to anticipate as many ethical scenarios as possible in order to build a service that is trusted. Important insights of dialogue systems for the tasks of a movie conversational RS were provided. Now, it is up to the human computer interface research community to contribute to the further development and personalization of the conversational RSs.

6. Conclusion

The present study reported the first evidence that conversational RSs boost users satisfaction. Such a positive impact emerged clearly when comparing the reactions users had reported while interacting with a traditional RS and a conversational RS. Users appeared willing to have more natural conversations services and to interact more with the RS in order to specify and refine their preferences. In order to investigate how to personalize the conversational RS, also an analysis was conducted to examine the impact that accurate and serendipitous recommendations have on users. In general, people low on Openness to Experience prefer accurate recommendations, while people High on Openness to Experience appreciate with equal pleasure both accurate and serendipitous recommendations. Using these insights, scripted conversations with the conversational RS were built in such a way as to use different words and provide recommendations with a different degree of accuracy depending on whether people were high or low on Openness to Experience. It would be interesting to explore in future research also other personality measures and to personalize conversations accordingly. Moreover, additional in-depth studies are needed regarding the customization of the conversational recommender system, both in terms of the choice of recommendations and the words used.

7. Appendices

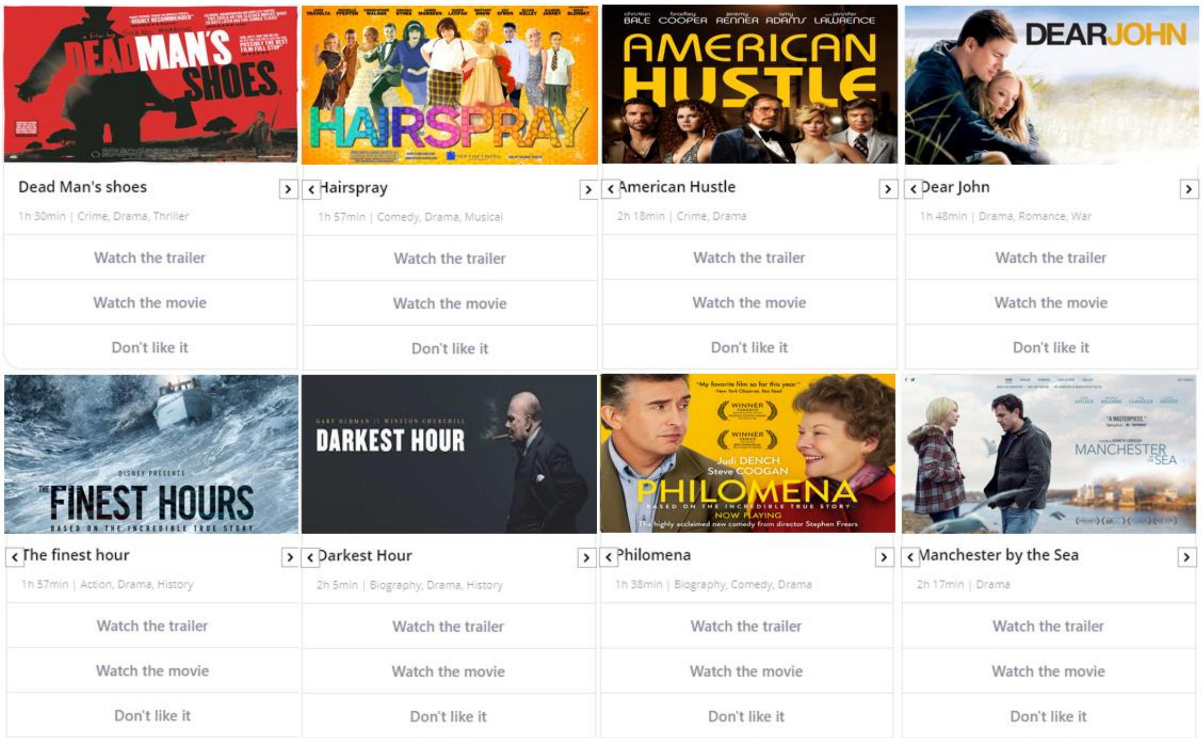
7.1 Sample conversation script low Openness to Experience

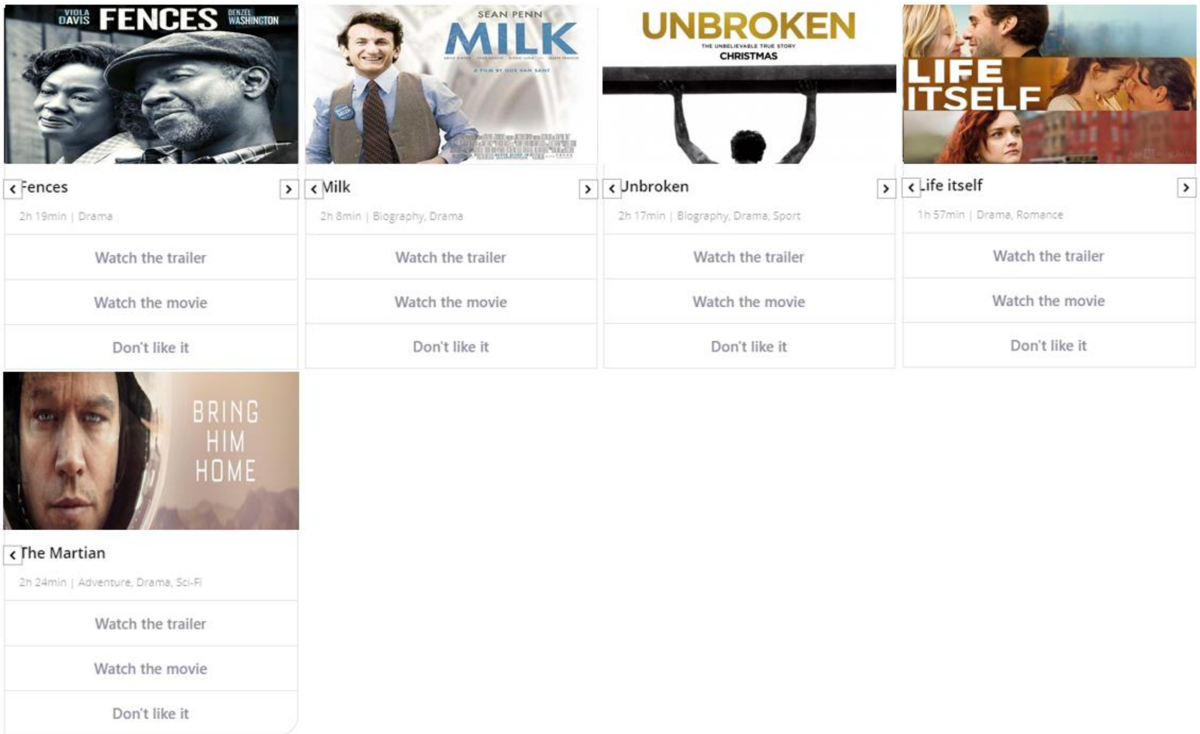
Conversational RS: Hi Maria! Welcome on MovieExplorer! Here is Dobby, your virtual assistant. Ready to give you movie suggestions anytime you ask 😊

Conversational RS: How can I help you?

User: Hi Dobby! Which movie do you suggest me?

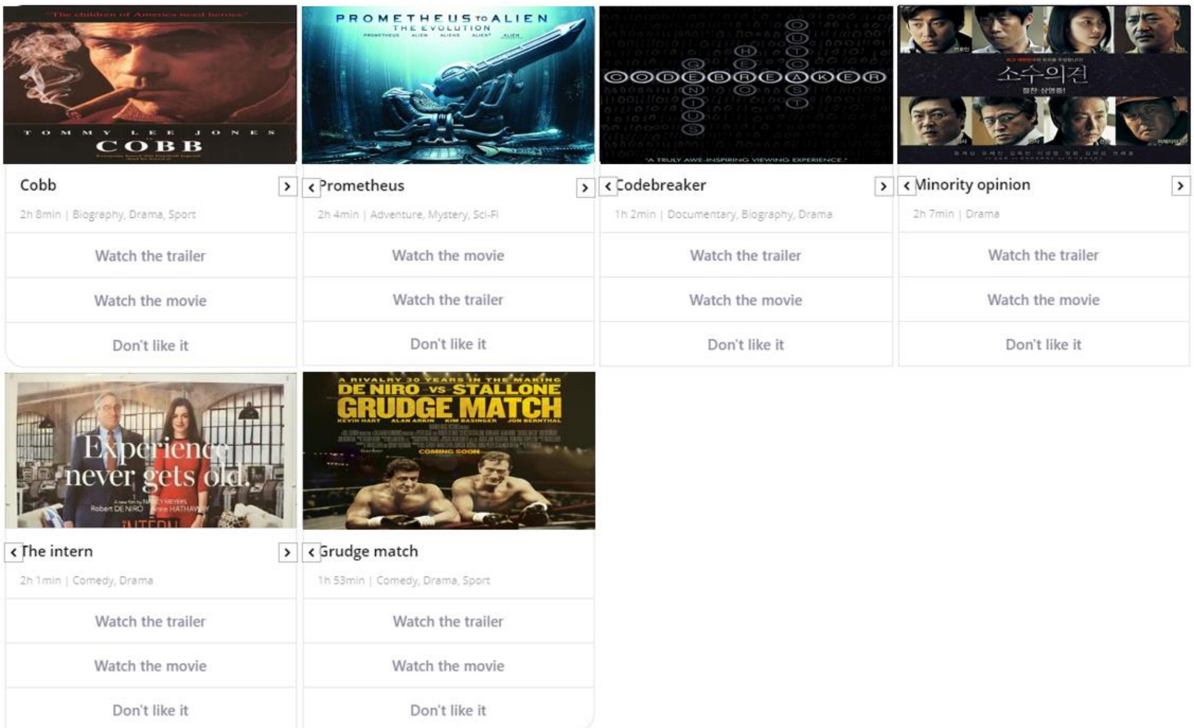
Conversational RS: Here are some movies similar to the one you usually watch. Check it out! [carousel]





User: I would like to watch something different today

Conversational RS: Then what about these suggestions? [carousel]



Conversational RS: Afterwards, if you have time let me know what do you think in this way I can improve. Enjoy! 🍿🎬 [play movie]

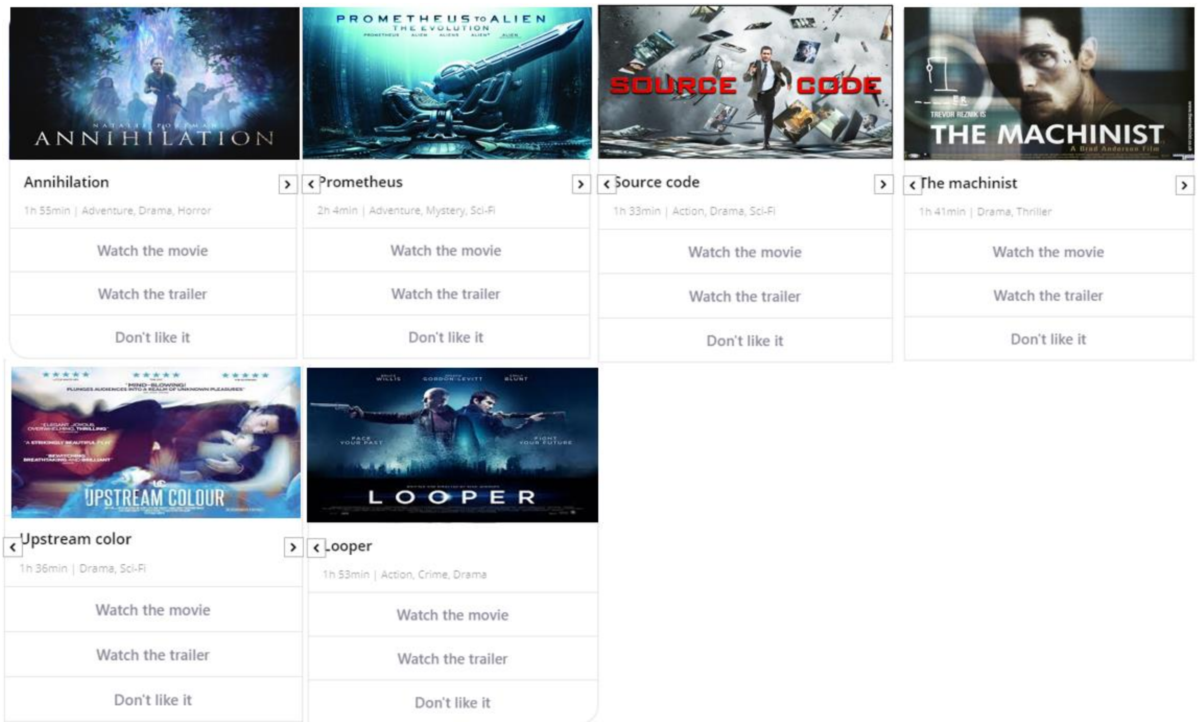
7.2 Sample conversation script high Openness to Experience

Conversational RS: Hi Valerio! Welcome on MovieExplorer! Here is Yoda, your virtual assistant. I am ready to give you movie suggestions anytime 😉

Conversational RS: How can I help you?

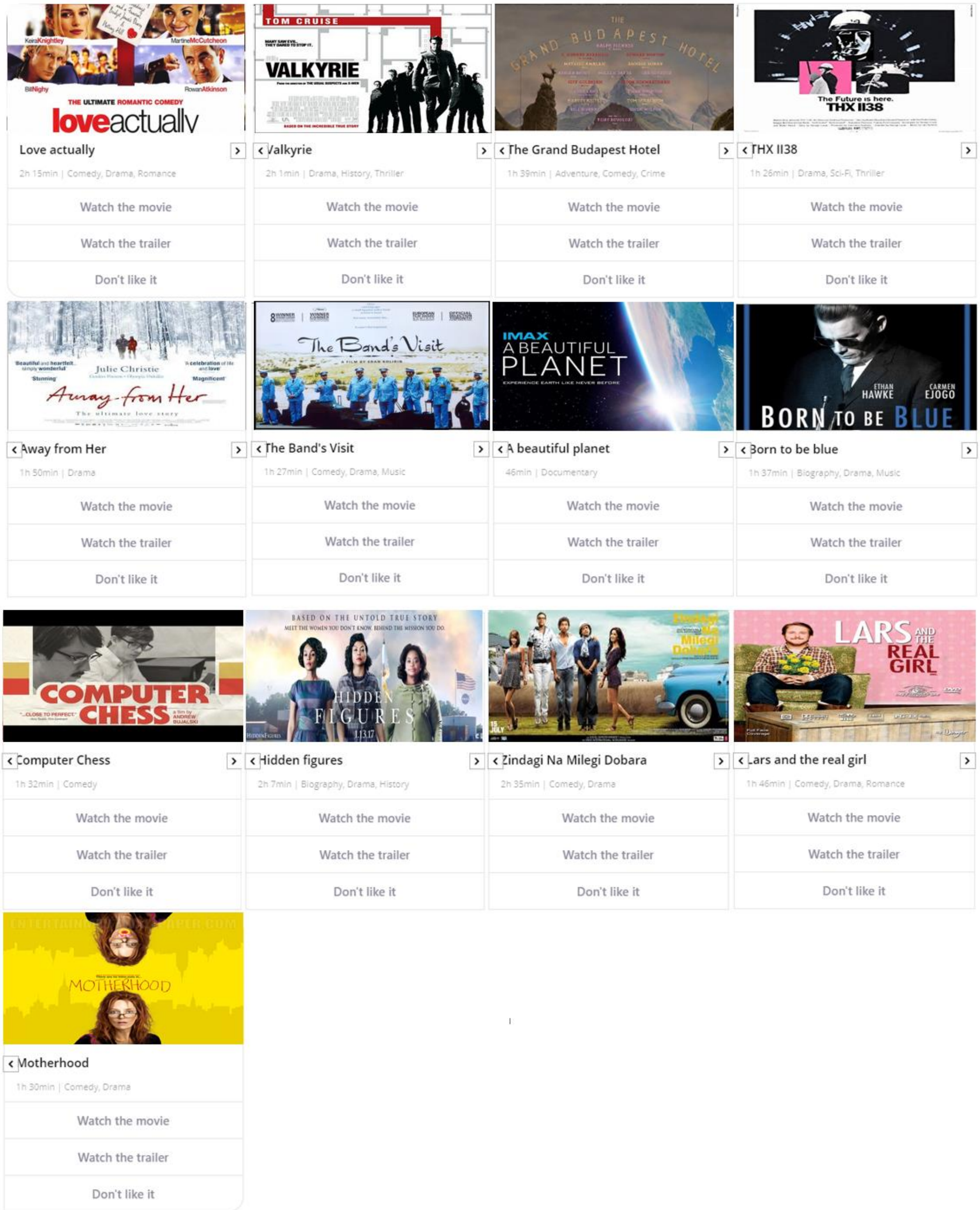
User: Hi Yoda! Which movie do you suggest me?

Conversational RS: Here are some movies I've the feeling you may like because similar to the ones you usually watch. Check it out! [carousel]



User: I would like to watch something different

Conversational RS: Here is something novel and bizarre [carousel]



When you want to tell me your opinions about the movie in this way I can improve my suggestions.

Enjoy! 🍿 🎬 [play movie]

8. References

Abdul-Kader, S. A., & Woods, J. C. (2015). Survey on chatbot design techniques in speech conversation systems. *International Journal of Advanced Computer Science and Applications*, *6*(7).

Abu-Dalbouh, H. (2016). Implementing End-User Privacy through Human Computer Interaction for Improving Quality of Personalized Web. *Computer and Information Science, 9*(1), 75. doi:10.5539/cis.v9n1p75

Adomavicius, G., & Tuzhilin, A. (2005). Personalization technologies: a process-oriented perspective. *Communications of the ACM*, *48*(10), 83-90.

Akhil, P. V., & Joseph, S. (2017). A survey of recommender system types and its classification. *International Journal of Advanced Research in Computer Science*, *8*(9).

Albrechtslund, A. (2007). Ethics and technology design. *Ethics and information technology*, *9*(1), 63-72.

Amatriain, X., & Basilico, J. (2015). Recommender systems in industry: A netflix case study. In *Recommender systems handbook* (pp. 385-419). Springer, Boston, MA.

Anelli, V. W., Basile, P., Bridge, D., Di Noia, T., Lops, P., Musto, C., ... & Zanker, M. (2018, September). Knowledge-aware and conversational recommender systems. In *Proceedings of the 12th ACM Conference on Recommender Systems* (pp. 521-522). ACM.

Attfield, S., Kazai, G., Lalmas, M., & Piwowarski, B. (2011, February). Towards a science of user engagement (position paper). In *WSDM workshop on user modelling for Web applications* (pp. 9-12).

Atzori, M., Boratto, L., & Spano, L. D. (2017). Towards chatbots as recommendation interfaces. In *CEUR WORKSHOP PROCEEDINGS* (Vol. 1945, pp. 28-31). CEUR-WS.

Bakarov, A., Yadrintsev, V., & Sochenkov, I. (2018, May). Anomaly detection for short texts: Identifying whether your chatbot should switch from goal-oriented conversation to chit-chatting. In *International Conference on Digital Transformation and Global Society* (pp. 289-298). Springer, Cham.

Ball, D., Coelho, P. S., & Vilares, M. J. (2006). Service personalization and loyalty. *Journal of services marketing*, *20*(6), 391-403.

Basu, C., Hirsh, H., & Cohen, W. (1998, July). Recommendation as classification: Using social and content-based information in recommendation. In *Aaai/iaai* (pp. 714-720).

Bouvier, P., Sehaba, K., & Lavoué, É. (2014). A trace-based approach to identifying users' engagement and qualifying their engaged-behaviours in interactive systems: application to a social game. *User Modeling and User-Adapted Interaction*, *24*(5), 413-451.

Brandtzaeg, P. B., & Følstad, A. (2017, November). Why people use chatbots. In *International Conference on Internet Science* (pp. 377-392). Springer, Cham.

Brusilovski, P., Kobsa, A., & Nejdl, W. (Eds.). (2007). *The adaptive web: methods and strategies of web personalization* (Vol. 4321). Springer Science & Business Media.

Cantador, I., Fernández-Tobías, I., & Bellogín, A. (2013). Relating personality types with user preferences in multiple entertainment domains. In *CEUR workshop proceedings*. Shlomo Berkovsky.

Chantarotwong, B. (2006). The learning chatbot. *Final year project.[Online]: http://courses. ischool. berkeley. edu/i256/f06/projects/bonniejc. pdf*.

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS quarterly*, *36*(4).

Chen, L., De Gemmis, M., Felfernig, A., Lops, P., Ricci, F., & Semeraro, G. (2013). Human decision making and recommender systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *3*(3), 17.

Chen, L., Wu, W., & He, L. (2013, April). How personality influences users' needs for recommendation diversity?. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems* (pp. 829-834). ACM.

Christakopoulou, K., Radlinski, F., & Hofmann, K. (2016, August). Towards conversational recommender systems. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 815-824). ACM.

Chulyadyo, R., & Leray, P. (2014). A personalized recommender system from probabilistic relational model and users' preferences. *Procedia Computer Science* , *35* , 1063-1072.

Coba, L., Rook, L., Zanker, M., & Symeonidis, P. (2019, March). Decision making strategies differ in the presence of collaborative explanations: Two conjoint studies. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (pp. 291-302). ACM.

Cui, L., Huang, S., Wei, F., Tan, C., Duan, C., & Zhou, M. (2017). Superagent: A customer service chatbot for e-commerce websites. *Proceedings of ACL 2017, System Demonstrations*, 97-102.

Cyr, D., Kindra, G. S., & Dash, S. (2008). Web site design, trust, satisfaction and e-loyalty: the Indian experience. *Online Information Review*, *32*(6), 773-790.

Dahiya, M. (2017). A tool of conversation: Chatbot. *International Journal of Computer Sciences and Engineering*, *5*(5), 158-161.

de Haan, H., Snijder, J., van Nimwegen, C., & Beun, RJ (2018). Chatbot Personality and Customer Satisfaction. Bachelor thesis.

de Oliveira, M. J., Huertas, M. K. Z., & Lin, Z. (2016). Factors driving young users' engagement with Facebook: Evidence from Brazil. *Computers in Human Behavior*, *54*, 54-61.

DeYoung, C. G. (2006). Higher-order factors of the Big Five in a multi-informant sample. *Journal of personality and social psychology*, *91*(6), 1138.

Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The mini-IPIP scales: tiny-yet-effective measures of the Big Five factors of personality. *Psychological assessment*, *18*(2), 192.

Dooms, S., De Pessemier, T., & Martens, L. (2014). Improving IMDb movie recommendations with interactive settings and filters. In *8th ACM Conference on Recommender Systems (Poster-RecSys 2014)* (Vol. 1247).

Driskell, J.E., & Salas, E. (1992). Can you study real teams in contrived settings? the value of small group research to understanding teams. In: R. W. Swezey & E. Salas (Eds.), Teams: Their training and performance pp. 101–126.

Ekstrand, M. D., Riedl, J. T., & Konstan, J. A. (2011). Collaborative filtering recommender systems. *Foundations and Trends® in Human–Computer Interaction*, *4*(2), 81-173.

Fan, H., & Poole, M. S. (2006). What is personalization? Perspectives on the design and implementation of personalization in information systems. *Journal of Organizational Computing and Electronic Commerce*, *16*(3-4), 179-202.

Fast, L. A., & Funder, D. C. (2008). Personality as manifest in word use: Correlations with self-report, acquaintance report, and behavior. Journal of Personality and Social Psychology, 94(2), 334.

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

Flaskerud, J. H. (1979). Use of vignettes to elicit responses toward broad concepts. *Nursing Research*, *28*(4), 210-212.

Følstad, A., & Brandtzæg, P. B. (2017). Chatbots and the new world of HCI. *interactions*, *24*(4), 38-42.

Foltz, P. W., & Dumais, S. T. (1992). Personalized information delivery: An analysis of information filtering methods. *Communications of the ACM*, *35*(12), 51-60.

Ge, M., Delgado-Battenfeld, C., & Jannach, D. (2010, September). Beyond accuracy: evaluating recommender systems by coverage and serendipity. In *Proceedings of the fourth ACM conference on Recommender systems* (pp. 257-260). ACM.

Giese, J. L., & Cote, J. A. (2000). Defining consumer satisfaction. *Academy of marketing science review*, *1*(1), 1-22.

Gnewuch, U., Morana, S., Adam, M. T., & Maedche, A. (2018, June). Faster is Not Always Better: Understanding the Effect of Dynamic Response Delays in Human-Chatbot Interaction. In *ECIS* (p. 113).

Goy, A., Ardissono, L., & Petrone, G. (2007). Personalization in e-commerce applications. In *The adaptive web* (pp. 485-520). Springer, Berlin, Heidelberg.

Gomez-Uribe, C. A., & Hunt, N. (2016). The netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems (TMIS)*, *6*(4), 13.

Graziano, W. G., Habashi, M. M., Sheese, B. E., & Tobin, R. M. (2007). Agreeableness, empathy, and helping: A person × situation perspective. *Journal of Personality and Social Psychology*, 93, 583–599. doi:10.1037/0022- 3514.93.4.583

Hayes, S. C., & Blackledge, J. T. (1998). Single case experimental designs: Clinical research and practice.

Harianto, S. (2018). Personalized Personality Virtual Agents: Assessing the impacts of virtual agent's personality match on user's trust, personal attachment, perceived risk and purchase intention in ecommerce. Master's thesis.

Hassenzahl, M., & Tractinsky, N. (2006). User experience-a research agenda. *Behaviour & information technology*, *25*(2), 91-97.

Hensley, C.B. Selective dissemination of information (SDI): State of the art in May, 1963. In *Proceedings of AFIPS '63* (Spring), 1963, 257–262.

Hill, J., Ford, WR, & Farreras, IG (2015). Real conversations with artificial intelligence: A comparison between human – human online conversations and human – chatbot conversations. *Computers in Human Behavior* , *49* , 245-250.

Hirsh, J. B., & Peterson, J. B. (2009). Personality and language use in self-narratives. *Journal of research in personality*, *43*(3), 524-527.

Horzyk, A., Magierski, S., & Miklaszewski, G. (2009). An Intelligent Internet Shop-Assistant Recognizing a Customer Personality for Improving Man-Machine Interactions. *Recent Advances in intelligent information systems*, 13-26.

Howard, J. A., & Sheth, J. N. (1969). The theory of buyer behavior. *New York*, 63.

Hurley, N., & Zhang, M. (2011). Novelty and diversity in top-n recommendation--analysis and evaluation. *ACM Transactions on Internet Technology (TOIT)*, *10*(4), 14.

Hu, R., & Pu, P. (2009, October). Acceptance issues of personality-based recommender systems. In *Proceedings of the third ACM conference on Recommender systems* (pp. 221-224). ACM.

Hunt, H. K. (1977). CS/D-overview and future research directions. *Conceptualization and measurement of consumer satisfaction and dissatisfaction*, 455-488.

Jannach, D., Resnick, P., Tuzhilin, A., & Zanker, M. (2016). Recommender systems — beyond matrix completion. *Communications of the ACM* , *59* (11), 94-102.

Kang, M., Shin, D. H., & Gong, T. (2016). The role of personalization, engagement, and trust in online communities. *Information Technology & People*, *29*(3), 580-596.

Kaptein, M., & Eckles, D. (2010, June). Selecting effective means to any end: Futures and ethics of persuasion profiling. In *International Conference on Persuasive Technology* (pp. 82-93). Springer, Berlin, Heidelberg.

Kerly, A., Hall, P., & Bull, S. (2007). Bringing chatbots into education: Towards natural language negotiation of open learner models. *Knowledge-Based Systems*, *20*(2), 177–185.

Kleinberg, J. M. (1997, May). Two algorithms for nearest-neighbor search in high dimensions. In *STOC* (Vol. 97, pp. 599-608).

Kucherbaev, P., Psyllidis, A., & Bozzon, A. (2017). Chatbots as conversational recommender systems in urban contexts. *arXiv preprint arXiv:1706.10076*.

Kurilchik, E. (2017). Chatbots as a Digital Marketing Communication Tool: Case Company: Wiredelta.

Law, E. L. C., Roto, V., Hassenzahl, M., Vermeeren, A. P., & Kort, J. (2009, April). Understanding, scoping and defining user experience: a survey approach. In *Proceedings of the SIGCHI conference on human factors in computing systems*(pp. 719-728). ACM.

Li, YM, Wu, CT, & Lai, CY (2013). A social recommender mechanism for e-commerce: Combining similarity, trust, and relationship. *Decision Support Systems* , *55* (3), 740-752.

Liang, TP, Lai, HJ, & Ku, YC (2006). Personalized content recommendation and user satisfaction: Theoretical synthesis and empirical findings. *Journal of Management Information Systems* , *23* (3), 45-70.

Litman, J. A., & Silvia, P. J. (2006). The latent structure of trait curiosity: Evidence for interest and deprivation curiosity dimensions. *Journal of Personality Assessment*, *86*(3), 318-328.

Looney, H. F., Jacobson, M., & Redding, M. J. (2011). *U.S. Patent No. 7,925,549*. Washington, DC: U.S. Patent and Trademark Office.

Lops, P., De Gemmis, M., & Semeraro, G. (2011). Content-based recommender systems: State of the art and trends. In *Recommender systems handbook* (pp. 73-105). Springer, Boston, MA.

Lortie, C. L., & Guitton, M. J. (2011). Judgment of the humanness of an interlocutor is in the eye of the beholder. *PloS one*, *6*(9), e25085.

McCrae, R.R., & John, O.P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, *60*(2), 175–215.

McCrae, R. R., & Costa Jr, P. T. (1995). Trait explanations in personality psychology. *European Journal of Personality*, *9*(4), 231-252.

McCrae, R. R. (1996). Social consequences of experiential openness. *Psychological bulletin*, *120*(3), 323.

McCrae, R. R., & Costa Jr, P. T. (1997). Conceptions and correlates of openness to experience. In *Handbook of personality psychology* (pp. 825-847). Academic Press.

McCrae, R. R., & Costa, P. T. (1999). A five-factor theory of personality. In L. Pervin & O. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 139–153). New York: Guilford Press

McTear, M. F. (2016, November). The rise of the conversational interface: A new kid on the block?. In *International Workshop on Future and Emerging Trends in Language Technology* (pp. 38-49). Springer, Cham.

Mehl, M. R., Pennebaker, J. W., Crow, D. M., Dabbs, J., & Price, J. H. (2001). The Electronically Activated Recorder (EAR): A device for sampling naturalistic daily activities and conversations. *Behavior Research Methods, Instruments, & Computers*, *33*(4), 517-523.

Mehl, M. R., Vazire, S., Ramírez-Esparza, N., Slatcher, R. B., & Pennebaker, J. W. (2007). Are women really more talkative than men?. *Science*, *317*(5834), 82-82.

Moon, Y. (2002). Personalization and personality: Some effects of customizing message style based on consumer personality. *Journal of Consumer Psychology*, *12*(4), 313-325.

Narducci, F., de Gemmis, M., Lops, P., & Semeraro, G. (2018, November). Improving the User Experience with a Conversational Recommender System. In *International Conference of the Italian Association for Artificial Intelligence* (pp. 528-538). Springer, Cham.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of social issues* , *56* (1), 81-103.

Nunes, M. A. S. N., Cerri, S. A., & Blanc, N. (2008, September). Improving recommendations by using personality traits in user profiles. In *International Conferences on Knowledge Management and New Media Technology* (pp. 92-100).

O'Brien, H. L., & Toms, E. G. (2008). What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of the American society for Information Science and Technology*, *59*(6), 938-955

O'reilly, T. (2005). Web 2.0: compact definition.

Opdenakker, R. (2006, September). Advantages and disadvantages of four interview techniques in qualitative research. In *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* (Vol. 7, No. 4).

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. Journal of Experimental Social Psychology, 45, 867–872.

Pearson, P. H. (1970). Relationships between global and specified measures of novelty seeking. Journal of Consulting and Clinical Psychology, 34, 199–204.

Peters, F. (2018). Master thesis: Design and implementation of a chatbot in the context of customer support.

Quijano-Sánchez, L., Bridge, D., Díaz-Agudo, B., & Recio-García, J. A. (2012, September). A case-based solution to the cold-start problem in group recommenders. In *International Conference on Case-Based Reasoning* (pp. 342-356). Springer, Berlin, Heidelberg.

Radziwill, N. M., & Benton, M. C. (2017). Evaluating quality of chatbots and intelligent conversational agents. *arXiv preprint arXiv:1704.04579*.

Rashid, A. M., Karypis, G., & Riedl, J. (2008). Learning preferences of new users in recommender systems: an information theoretic approach. *Acm Sigkdd Explorations Newsletter*, *10*(2), 90-100.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. and Riedl, J. Grouplens: An open architecture for collaborative filtering of netnews. In *Proceedings of CSCW '94* (1994), 175–186.

Ricci, F., Rokach, L., & Shapira, B. (2015). Recommender systems handbook. Second Edition. Springer, Boston, MA.

Rook, L., Sabic, A., & Zanker, M. (2018). Commitment in proactive recommendations: the role of recommendation accuracy, information privacy concerns and personality traits. *Journal of Intelligent Information Systems.*

Rook, L., Sabic, A., & Zanker, M. (2018). Reinforcement Sensitivity and Engagement in Proactive Recommendations: Experimental Evidence. In *Information Systems and Neuroscience* (pp. 9-15). Springer, Cham.

Schafer, J. B., Konstan, J., & Riedl, J. (1999, November). Recommender systems in e-commerce. In *Proceedings of the 1st ACM conference on Electronic commerce* (pp. 158-166). ACM.

Sekaran, U., & Bougie, R. (2016). *Research methods for business: A skill building approach*. John Wiley & Sons.

Shawar, B. A., & Atwell, E. S. (2005). Using corpora in machine-learning chatbot systems. *International journal of corpus linguistics*, *10*(4), 489-516.

Shawar, B. A., & Atwell, E. (2007, January). Chatbots: are they really useful?. In *Ldv forum* (Vol. 22, No. 1, pp. 29-49).

Shawar, B. A., & Atwell, E. (2007, April). Different measurements metrics to evaluate a chatbot system. In *Proceedings of the workshop on bridging the gap: Academic and industrial research in dialog technologies* (pp. 89-96). Association for Computational Linguistics.

Smith, K. A., Dennis, M., Masthoff, J., & Tintarev, N. (2018). A methodology for creating and validating psychological stories for conveying and measuring psychological traits. *User Modeling and User-Adapted Interaction*, 1-46.

Tatai, G., Csordás, A., Kiss, Á., Szaló, A., & Laufer, L. (2003, September). Happy chatbot, happy user. In *International Workshop on Intelligent Virtual Agents* (pp. 5-12). Springer, Berlin, Heidelberg.

Tintarev, N., & Masthoff, J. (2012). Evaluating the effectiveness of explanations for recommender systems. *User Modeling and User-Adapted Interaction*, *22*(4-5), 399-439.

Tintarev, N., Dennis, M., & Masthoff, J. (2013, June). Adapting recommendation diversity to openness to experience: A study of human behaviour. In *International Conference on User Modeling, Adaptation, and Personalization* (pp. 190-202). Springer, Berlin, Heidelberg.

Tintarev, N., & Masthoff, J. (2015). Explaining recommendations: Design and evaluation. In *Recommender systems handbook* (pp. 353-382). Springer, Boston, MA.

Tkalcic, M., & Chen, L. (2015). Personality and recommender systems. In *Recommender systems handbook* (pp. 715-739). Springer, Boston, MA.

Tkalcic, M., Kunaver, M., Tasic, J., & Košir, A. (2009). Personality-based user similarity measure for a collaborative recommender system. In *Proceedings of the 5th Workshop on Emotion in Human-Computer Interaction-Real world challenges* (pp. 30-37).

Turing, A. M., 1950, Computing machinery and intelligence, Mind 50: 433–460

Urbieta, M., Torres, N., Rivero, J. M., Rossi, G., & Dominguez-Mayo, F. J. (2018, May). Improving mockup-based requirement specification with end-user annotations. In *International Conference on Agile Software Development* (pp. 19-34). Springer, Cham.

VanderMeer, D., Dutta, K., Datta, A., Ramamritham, K., & Navanthe, S. B. (2000, October). Enabling scalable online personalization on the Web. In *Electronic Commerce: Proceedings of the 2 nd ACM conference on Electronic commerce* (Vol. 17, No. 20, pp. 185-196).

Vargo, S. L., & Lusch, R. F. (2008). Service-dominant logic: continuing the evolution. *Journal of the Academy of marketing Science*, *36*(1), 1-10.

Vesanen, J. (2007). What is personalization? A conceptual framework. *European Journal of Marketing*, *41*(5/6), 409-418.

Wang, Z., Yu, X., Feng, N., & Wang, Z. (2014). An improved collaborative movie recommendation system using computational intelligence. *Journal of Visual Languages & Computing*, *25*(6), 667-675.

Westbrook, R. A., & Reilly, M. D. (1983). Value-percept disparity: an alternative to the disconfirmation of expectations theory of consumer satisfaction. *ACR North American Advances*.

Xu, A., Liu, Z., Guo, Y., Sinha, V., & Akkiraju, R. (2017, May). A new chatbot for customer service on social media. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (pp. 3506-3510). ACM.

Yarkoni, T. (2010). Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of research in personality* , *44* (3), 363-373.

Yarkoni, T. (2012). Psychoinformatics: New horizons at the interface of the psychological and computing sciences. Current Directions in Psychological Science, 21(6), 391

Yi, Y. (1990). A critical review of consumer satisfaction. *Review of marketing*, *4*(1), 68-123.

Yu, C., Aoki, P. M., & Woodruff, A. (2004). Detecting user engagement in everyday conversations. *arXiv preprint cs/0410027*.

Yueming, S., & Zhang, Y. (2018) Conversational Recommender System. *SIGIR '18 The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*: 235-244

Zanker, M., Rook, L., & Jannach, D. (2019). Measuring the Impact of Online Personalisation: Past, Present and Future. *International Journal of Human-Computer Studies*.

Ziegler, CN, McNee, SM, Konstan, JA, & Lausen, G. (2005, May). Improving recommendation lists through topic diversification. In *Proceedings of the 14th international conference on the World Wide Web* (pp. 22-32). ACM.