JOHANNES KEPLER
UNIVERSITÄT LINZ

Eingereicht von
Bruce Ferwerda

Angefertigt am
Department of Computational Perception

Erstbeurteiler
Markus Schedl

Zweitbeurteiler
Marco de Gemmis

Mitbetreuung
Marko Tkalcic

Improving the User Experience of Music Recommender Systems Through Personality and Cultural Information

Dissertation
zur Erlangung des akademischen Grades
Doktor der Technischen Wissenschaften
im Doktoratsstudium der
Technischen Wissenschaften

**JOHANNES KEPLER
UNIVERSITÄT LINZ**
Altenbergerstraße 69
4040 Linz, Österreich
www.jku.at
DVR 0093696

“To dare is to do ... to fear is to fail.”
— John Goddard

Sworn Declaration

I hereby declare under oath that the submitted Doctoral Dissertation has been written solely by me without any third-party assistance, information other than provided sources or aids have not been used and those used have been fully documented. Sources for literal, paraphrased and cited quotes have been accurately credited.

The submitted document here present is identical to the electronically submitted text document.

Linz, August 1, 2016

Bruce Ferwerda

Abstract (GER)

Personalisierung ist der nächste Schritt zur Verbesserung der User Experience von Empfehlungssystemen. Mit Hilfe von Personalisierung können sich Systeme auf das Verhalten, die Vorlieben und Bedürfnisse der Benutzer anpassen. Um dies zu erreichen, erstellen Systeme Benutzermodelle auf Basis der von ihnen generierten Daten, um sodann Adaptierungen vornehmen zu können. Hierbei stellt sich die Frage, woher genug hochqualitative Daten über einen (neuen) Benutzer gesammelt werden können, um akkurate Modelle zu erstellen. Eine Möglichkeit bieten Fragebögen, sie sind jedoch aufdringlich und verlangen vom Benutzer Zeit und Aufwand, was deren Interaktion mit dem System beeinträchtigt.

Eine mögliche Antwort auf den Mangel an Daten bietet der Trend, Systeme mit sozialen Netzwerken über sogenannte “single sign-on” Schalter für Login und Registrierung zu verbinden. Die Verbindung mit sozialen Netzwerken ist nicht nur praktisch für Benutzer (z.B. durch Zeitersparnis beim Registrierungsprozess), sondern erlaubt auch dem System, weitere Informationen über den Benutzer zu erhalten. Diese sind jedoch nicht immer direkt nützlich für das System. Um sie bestmöglich zu verwerten, werden allgemeine Benutzermodelle benötigt, zu denen das Verhalten, die Vorlieben und Bedrfnisse der Nutzer in Relation gebracht werden können.

In dieser Dissertation werden Persönlichkeit und kultureller Hintergrund von Benutzern zur Benutzermodellierung betrachtet. Diese zwei Konzepte sind dauerhaft und beständig, und beeinflussen das Verhalten, die Vorlieben und die Bedrfnisse in realen Situationen. Der Einfluss dieser Konzepte im technologischen Kontext ist jedoch noch relativ unerforscht. Um diesen Mangel zu beseitigen und Benutzermodelle auf Basis ebendieser Konzepte zu erstellen, beinhaltet diese Dissertation die folgenden wissenschaftlichen Beitrge:

1. Sie untersucht, ob und wie Persönlichkeit und Kultur im Verhältnis zu Verhalten, Vorlieben und Bedürfnissen stehen.
2. Sie beschreibt eine implizite Art persönlichkeits- und kulturrelevante Daten aus sozialen Netzwerken zu erlangen.

Obwohl allgemeine Benutzermodelle in verschiedenen Kontexten verwendbar sind, beschäftigt sich diese Dissertation spezifisch mit Empfehlungssystemen für Musik. Zu diesem Zweck, und Bezug nehmend auf das erste Ziel, wurden mehrere Studien über verschiedene Aspekte des gesamten Musikerlebnisses durchgeführt. Ihre Ergebnisse zeigen, dass Persönlichkeit und kulturelle Aspekte Musikkonsumsverhalten, -vorlieben und -bedürfnisse beeinflussen. Zusätzlich zur Erforschung dieses Zusam-

menhangs wurde eine benutzerzentrierte Evaluierung von echten personalisierten Musikempfehlungen durchgeführt.

Das zweite Ziel ermöglicht das automatisierte Erstellen von persönlichkeits- und kulturbasierten Benutzermodellen mit Hilfe von sozialen Netzwerken wie Facebook, Twitter, und Instagram: 1) Förderung von "Sharing"- und "Posting"-Verhalten auf Facebook, sodass mehr Informationen zur Benutzermodellierung zur Verfügung stehen, 2) Benutzermodellierung auf Basis von limitierten Informationen aus dem Benutzerprofil auf Facebook, 3) das Ausnutzen von Bildeigenschaften von Instagram-Fotos um Persönlichkeitsmerkmale aufgrund der angewandten Fotomanipulationen zu lernen und 4) das Zusammenführen von Informationen aus verschiedenen sozialen Netzwerken, Twitter und Instagram, um Benutzermodelle zu verbessern.

Die Ergebnisse dieser Dissertation beinhalten neue Erkenntnisse darüber, wie die Persönlichkeit und der kulturelle Hintergrund von Benutzern ihr Verhalten, ihre Vorlieben und ihre Bedürfnisse in einem musikalischen Kontext beeinflussen. Die Arbeit zeigt auch, wie Persönlichkeits- und kulturelle Dimensionen aus sozialen Netzwerken gewonnen werden können. Die durchgeführte Forschung zeigt eine neue umfassende Art personalisierte Musikempfehlungssysteme zu erstellen, insbesondere mit begrenzten Daten über das Benutzerverhalten.

Verschiedene Aspekte zur Verbesserung der User Experience wurden berücksichtigt. Sie lassen sich in drei allgemeine Kategorien unterteilen:

- Verstehen der Benutzer und ihres Musikkonsumverhaltens, ihrer -vorlieben und -bedürfnisse.
- Impliziter Erwerb von Persönlichkeitsmerkmalen der Benutzer aus sozialen Netzwerken.
- Benutzerzentrierte Evaluierung von Musikempfehlungssystemen.

Eine allgemeine Beschreibung des vorgestellten personalisierten Musiksystems findet sich in:

- **Ferwerda, B.**, & Schedl, M. (2016) Personality-Based User Modeling for Music Recommender Systems. *In Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases* (Riva del Garda, IT).
- **Ferwerda, B.** (2015) The Soundtrack of My Life: Adjusting the Emotion of Music. *In the 1st Workshop on Collaborating with Intelligent Machines held in conjunction with the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, KR).

- **Ferwerda, B.**, & Schedl, M. (2014) Enhancing Music Recommender Systems with Personality Information and Emotional States: A Proposal. *In Proceedings of the 2nd Workshop on Emotions and Personality in Personalized Services held in conjunction with the 22nd Conference on User Modeling, Adaptation and Personalization* (Aalborg, DK).

Artikel, die den oben vorgestellen Kategorien zugeordnet sind, sind im Folgenden aufgeführt.

Verstehen der Benutzer und ihres Musikkonsumverhaltens, ihrer -vorlieben und -bedürfnisse.

- **Ferwerda, B.**, Tkalcic, M., & Schedl, M. (2016). Exploring Music Diversity Needs Across Countries. *In Proceedings of the 24th Conference on User Modeling, Adaptation and Personalization* (Halifax, NS, CA).
- **Ferwerda, B.**, & Schedl, M. (2016) Investigating the Relationship Between Diversity in Music Consumption Behavior and Cultural Dimensions: A Cross-country Analysis. *In Extended Proceedings of the 24th Conference on User Modeling, Adaptation and Personalization: 1st Workshop on Surprise, Opposition, and Obstruction in Adaptive and Personalized Systems* (Halifax, NS, CA).
- **Ferwerda, B.**, Yang, E., Schedl, M., & Tkalcic, M. (2016) Personality and Taxonomy Preferences, and the Influence of Category Choice Set Size on the User Experience for Music Streaming Services. *Under Review*.
- **Ferwerda, B.**, Yang, E., Schedl, M., & Tkalcic, M. (2015) Personality Traits Predict Music Taxonomy Preferences. *In Extended Abstracts Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, KR).
- **Ferwerda, B.**, Schedl, M., & Tkalcic, M. (2015). Personality & Emotional States: Understanding Users' Music Listening Needs. *In Proceedings of the 23rd Conference on User Modeling, Adaptation and Personalization* (Dublin, IR).
- Tkalcic, M., **Ferwerda, B.**, Hauger, D., & Schedl, M. (2015) Personality Correlates for Digital Concert Program Notes. *In Proceedings of the 23rd Conference on User Modeling, Adaptation and Personalization* (Dublin, IR).

Impliziter Erwerb von Persönlichkeitsmerkmalen der Benutzer aus sozialen Netzwerken.

- **Ferwerda, B.**, Schedl, M., & Tkalcic, M. (2016) Personality Traits and the Relationship with (Non-)Disclosure Behavior on Facebook. *In Companion Proceedings of the 25th International Conference on World Wide Web: 7th International Workshop on Modeling Social Media* (Montreal, QB, CA).
- Skowron, M., **Ferwerda, B.**, Tkalcic, M., & Schedl, M. (2016) Fusing Social Media Cues: Personality Prediction from Twitter and Instagram. *In Companion Proceedings of the 25th International Conference on World Wide Web* (Montreal, QB, CA).
- **Ferwerda, B.**, Schedl, M., & Tkalcic, M. (2016) Using Instagram Picture Features to Predict Users' Personality. *In Proceedings of the 22nd International Conference on MultiMedia Modeling* (Miami, FL, US).
- **Ferwerda, B.**, Schedl, M., & Tkalcic, M. (2015) Predicting Personality Traits with Instagram Pictures. *In Extended Proceedings of the 9th ACM Conference on Recommender Systems: 3rd Workshop on Emotions and Personality in Personalized Services* (Vienna, AT).
- **Ferwerda, B.**, Schedl, M., & Tkalcic, M. (2014) To Post or Not to Post: The Effects of Persuasive Cues and Group Targeting Mechanisms on Posting Behavior. *In Proceedings of the 6th International Conference on Social Computing* (Stanford, CA, US).

Benutzerzentrierte Evaluierung von Musikempfehlungssystemen.

- **Ferwerda, B.**, Graus, M., Vall, A., Tkalcic, M., & Schedl, M. (2017) How Item Discovery Enabled by Diversity Leads to Increased Recommendation List Attractiveness. *In Submission*.
- **Ferwerda, B.**, Graus, M., Vall, A., Tkalcic, M., & Schedl, M. (2016) The Influence of Users' Personality Traits on Satisfaction and Attractiveness of Diversified Recommendation Lists. *In Extended Proceedings of the 10th ACM Conference on Recommender Systems: 4th Workshop on Emotions and Personality in Personalized Systems* (Boston, MA, US).

Abstract (ENG)

Personalization is the next step in order to improve the user experience of recommender systems. With personalization, systems are able to adapt in order to give users an experience that is tailored to their behavior, preferences, and needs. To achieve personalization, systems create user models based on user-generated data, so that a suitable adaptation strategy can be applied. The problem that persists is how to gather enough high-quality data about the (new) user to create accurate models. One way to solve this problem is to use questionnaires. However, this is not desirable since it is obtrusive, takes a lot of effort and time from the user, and thereby disrupts their interaction with the system.

The trend of systems being increasingly connected with social networking sites (SNSs; e.g., single sign-on buttons for login and registration purposes), may offer a better solution to the lack of data problem. The interconnectedness with SNSs is not only convenient for the user (e.g., skipping registration procedures by letting the system connect with their SNSs), but also allows systems to gain additional user information. However, not all the information may be directly useful for the targeted purpose. In order to exploit all the information available, general user models are needed to which users' behavior, preferences, and needs can be related to and inferred from.

In this dissertation, users' personality and their cultural background are considered for user modeling. These two constructs have shown to be enduring and stable, and relate to behavior, preferences, and needs in real-life situations. However, how these two constructs relate in a technological context is still relatively unknown. To address this shortcoming of current research and to effectively create and employ user models in a technological setting, the scientific contributions of this dissertation are twofold:

1. Investigating whether and how personality and culture relate to behavior, preferences, and needs.
2. Implicit acquisition of personality and cultural dimensions from SNS data.

Although the general user models could be employed in any context for personalization, the research conducted in this dissertation specifically focuses on the task of music recommendation. To this end, addressing the first goal, several studies have been conducted on various aspects of the music experience: music browsing behavior, music listening needs, and music recommendation list diversity. The results indicate that personality and cultural aspects do relate to different music consumption behavior, preferences, and needs. Not only the relationships were investigated, but

also a user-centric evaluation was performed, involving real personalized music recommendations.

The second goal provides ways to automatically infer personality and cultural user models from SNSs, such as Facebook, Twitter, and Instagram: 1) promoting sharing and posting behavior on Facebook so that more information becomes available for user model acquisition, 2) user model acquisition from limited user profile information on Facebook, 3) exploiting Instagram picture features to learn personality traits from information on how users modify their pictures on Instagram, and 4) fusing data from several SNSs, Twitter and Instagram, to improve user model acquisition.

The results presented in this dissertation provide new insights in how users' personality and cultural background influence behavior, preferences, and needs in a music context. The presented works also show how personality and cultural dimensions can be obtained from SNS data. The conducted research as a whole provides a novel comprehensive way for creating a personalized music recommender system, especially in cases when users' behavioral data with the system is (still) limited.

Different aspects involving the improvement of the user experience were considered, which can be divided into three general categories:

- Understanding users and their music listening behavior, preferences, and needs.
- Implicit acquisition of users' personal characteristics from social media.
- User-centric evaluation of a music recommender system.

A general description of the proposed personalized music system can be found in:

- **Ferwerda, B.**, & Schedl, M. (2016) Personality-Based User Modeling for Music Recommender Systems. *In Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases* (Riva del Garda, IT).
- **Ferwerda, B.** (2015) The Soundtrack of My Life: Adjusting the Emotion of Music. *In the 1st Workshop on Collaborating with Intelligent Machines held in conjunction with the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, KR).
- **Ferwerda, B.**, & Schedl, M. (2014) Enhancing Music Recommender Systems with Personality Information and Emotional States: A Proposal. *In Proceedings of the 2nd Workshop on Emotions and Personality in Personalized Services held in conjunction with the 22nd*

Conference on User Modeling, Adaptation and Personalization (Aalborg, DK).

The papers depicting each category are highlighted below.

Understanding users and their music listening behavior, preferences, and needs.

- **Ferwerda, B.**, Tkalcic, M., & Schedl, M. (2016). Exploring Music Diversity Needs Across Countries. *In Proceedings of the 24th Conference on User Modeling, Adaptation and Personalization* (Halifax, NS, CA).
- **Ferwerda, B.**, & Schedl, M. (2016) Investigating the Relationship Between Diversity in Music Consumption Behavior and Cultural Dimensions: A Cross-country Analysis. *In Extended Proceedings of the 24th Conference on User Modeling, Adaptation and Personalization: 1st Workshop on Surprise, Opposition, and Obstruction in Adaptive and Personalized Systems* (Halifax, NS, CA).
- **Ferwerda, B.**, Yang, E., Schedl, M., & Tkalcic, M. (2016) Personality and Taxonomy Preferences, and the Influence of Category Choice Set Size on the User Experience for Music Streaming Services. *Under Review*.
- **Ferwerda, B.**, Yang, E., Schedl, M., & Tkalcic, M. (2015) Personality Traits Predict Music Taxonomy Preferences. *In Extended Abstracts Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, KR).
- **Ferwerda, B.**, Schedl, M., & Tkalcic, M. (2015). Personality & Emotional States: Understanding Users' Music Listening Needs. *In Proceedings of the 23rd Conference on User Modeling, Adaptation and Personalization* (Dublin, IR).
- Tkalcic, M., **Ferwerda, B.**, Hauger, D., & Schedl, M. (2015) Personality Correlates for Digital Concert Program Notes. *In Proceedings of the 23rd Conference on User Modeling, Adaptation and Personalization* (Dublin, IR).

Implicit acquisition of users' personal characteristics from social media.

- **Ferwerda, B.**, Schedl, M., & Tkalcic, M. (2016) Personality Traits and the Relationship with (Non-)Disclosure Behavior on Facebook. *In Companion Proceedings of the 25th International Conference on World Wide Web: 7th International Workshop on Modeling Social Media* (Montreal, QB, CA).

- Skowron, M., **Ferwerda, B.**, Tkalcic, M., & Schedl, M. (2016) Fusing Social Media Cues: Personality Prediction from Twitter and Instagram. *In Companion Proceedings of the 25th International Conference on World Wide Web* (Montreal, QB, CA).
- **Ferwerda, B.**, Schedl, M., & Tkalcic, M. (2016) Using Instagram Picture Features to Predict Users' Personality. *In Proceedings of the 22nd International Conference on MultiMedia Modeling* (Miami, FL, US).
- **Ferwerda, B.**, Schedl, M., & Tkalcic, M. (2015) Predicting Personality Traits with Instagram Pictures. *In Extended Proceedings of the 9th ACM Conference on Recommender Systems: 3rd Workshop on Emotions and Personality in Personalized Services* (Vienna, AT).
- **Ferwerda, B.**, Schedl, M., & Tkalcic, M. (2014) To Post or Not to Post: The Effects of Persuasive Cues and Group Targeting Mechanisms on Posting Behavior. *In Proceedings of the 6th International Conference on Social Computing* (Stanford, CA, US).

User-centric evaluation of a music recommender system.

- **Ferwerda, B.**, Graus, M., Vall, A., Tkalcic, M., & Schedl, M. (2017) How Item Discovery Enabled by Diversity Leads to Increased Recommendation List Attractiveness. *In Submission*.
- **Ferwerda, B.**, Graus, M., Vall, A., Tkalcic, M., & Schedl, M. (2016) The Influence of Users' Personality Traits on Satisfaction and Attractiveness of Diversified Recommendation Lists. *In Extended Proceedings of the 10th ACM Conference on Recommender Systems: 4th Workshop on Emotions and Personality in Personalized Systems* (Boston, MA, US).

Curriculum Vitae

Full Name: Bruce Ferwerda

Education

PhD in Computer Science Johannes Kepler University (Linz, AT)	2014-2016
MSc in Human-Technology Interaction Eindhoven University of Technology (Eindhoven, NL)	2010-2012
BEng in E-Technology University of Applied Sciences, HvA (Amsterdam, NL)	2006-2009

Non-Degree Programs

Quantitative Research Methods University of California, Irvine (Irvine, CA, US)	2015-2015
Cognitive Science Yonsei University (Seoul, KR)	2012-2013
Distributed & Ubiquitous Computing Waseda University (Tokyo, JP)	2011-2012
Artificial Intelligence Radboud University (Nijmegen, NL)	2010-2010
IP DeSeRTS (Erasmus Mundus Program) University of Applied Sciences, EVTEK (Espoo, FI)	2010-2010

Experience

Researcher Johannes Kepler University (Linz, AT)	2013-2016
Visiting Researcher host: Alfred Kobsa University of California, Irvine (Irvine, CA, US)	2015-2015
Researcher Yonsei University (Seoul, KR)	2012-2013
UX Researcher & Designer MeasureWorks (Almere, NL)	2012-2013
Visiting Researcher host: Tatsuo Nakajima Waseda University (Tokyo, JP)	2011-2012
Engineer Studio Roosegaarde (Waddinxveen, NL)	2009-2010

Acknowledgments

First of all, I would like to thank my advisor Markus Schedl. The works presented in this dissertation would not have been possible without him. Furthermore, I would like to thank him for giving me the freedom to develop and pursue my own research interest within his project, and for the numerous conference support to get exposure of my work and to be able to create my own academic network.

My special thanks goes out to Marko Tkalcic for the many interesting and fruitful discussions to push the research to the next level. I also want to thank him for the great advisory role in my PhD trajectory even though I am not able to recognize him as one of my official advisors.

Last but not least, I want to thank all of those not directly related to my research (i.e., family, friends, and colleagues). Their moral support has been of great value to fulfill the requirements for my PhD.

Contents

1 Introduction

Recommender systems are increasingly implemented in different areas in order to help users to make (better) choices from an abundance of choice options. Recommender systems are created to provide suggestions that are most likely of interest to the user. The research community on recommender systems made great advances to improve the recommendation algorithms in order to be able to provide the user with more accurate recommendations [46]. Usually by analyzing users' implicit (e.g., clicking or buying behavior) and/or explicit (e.g., rating or weighting of items) feedback on items, a similarity measurement is created based on the captured feedback of the user. Items with the highest predicted similarity rating are then provided to the user as a choice set. It follows that the more accurate the recommendation algorithm is, the more accurate the recommendations will be to the users' interest. A common reasoning is of causality between recommendation accuracy and the user experience; better recommendations will lead to a higher user experience. Therefore, a majority of research has focused on the improvement of the recommendation algorithms to improve the accuracy of the recommendations.

Research has started to look beyond recommendation accuracy and found that other factors beside accuracy have influence on the user experience (e.g., [38]). These factors include other system aspects, such as recommendation diversity (e.g., [4, 23, 58]), but also personal (e.g., product expertise and prior preferences [4, 56]) or situational aspects (e.g., mood and emotion [19]), which all just started to receive attention in the creation of recommender systems. A problem that persist is how to capture these influential factors or how to gather enough behavioral data (this is especially problematic for new users) in order to infer user preferences for a personalized user experience. A simple way to get to know the user is to let them fill in questionnaires. However, this is often undesirable since it is obtrusive, takes a lot of effort and time from the user, and thereby disrupts their interaction with the system.

The implicit acquisition (i.e., without the use of questionnaires) of personal and situational aspects is still challenging. However, the increased connectedness of applications with social networking sites (SNSs) opens new ways to implicitly gather information about the user. For example, single sign-on (SSO) mechanisms [1] allow users to easily login and register through their SNSs, which allow applications to access the basic user information needed, but are often also granted access to other parts of the users' profile [10]. Although new information sources become available, not all the information may be directly useful for the targeted purpose. In order to exploit all the information available, general user models are needed to which users' behavior, preferences, and needs can be related to and inferred from.

General models, such as cultural dimensions and personality traits, have shown to be enduring and stable, and embody people's behavioral patterns, preferences, and needs in real-life situations [36, 37]. However, how these general user models relate in a technological setting is still relatively unknown. The works presented in this dissertation answers two general research questions (RQs):

1. How do personality traits and cultural background relate to behavior, preferences, and needs?
2. How can personality traits [2] be implicitly acquired from SNSs?

To investigate how personality traits and cultural background relate to behavior, preferences, and needs, this dissertation specifically focuses on the music domain. That is, the findings of RQ1 are in particular applicable to music recommender systems. The findings of RQ2 on the other hand are more general, and can be used for any system with an SNS connection. The research that was conducted for both RQs provide a novel comprehensive way for the creation of a personalized music recommender system. Inferring user preferences from these user models are especially useful in situations when it is not (yet) possible to infer preferences from behavioral data within the application.

The dissertation continues with an overview of the related works on the aforementioned aspects, followed with a dissertation outline consisting of a short description of the works that were conducted within this PhD trajectory (§3) with the full papers in the sections following that in §5 (understanding users and their music behavior, preferences, and needs), §6 (implicit acquisition of users' personal characteristics from social media), and §7 (user-centric evaluation of a

[1] Buttons that allow users to register or login with accounts of other applications. For example, social networking services: "Login with your Facebook account."

[2] The focus is laid on implicitly acquiring personality traits as cultural background (i.e., country information) can usually be easily obtained from the user's basic profile information.

music recommender system). This dissertation ends with a general conclusion in §8.

2 Related Works

In this section related work is provided about the key areas of this dissertation. First an overview is given on personalization in recommender systems followed by the related work on the general user models that are used in this dissertation: personality traits and cultural dimensions.

2.1 Personalization in Recommender Systems

Personalization in recommender systems is done on various levels of the system. The most basic personalization is done by providing recommendations to the user that are of most interest to them. By analyzing the user's implicit and/or explicit feedback on items, recommender systems are able to calculate similarity measurements to output a choice set with items consisting of the highest predicted similarity. The generation of item similarity measurements are based on general filtering methods: collaborative filtering (i.e., creating recommendations based on users with a similar taste), content-based filtering (i.e., creating recommendations based on item feedback given in the past), or hybrid (i.e., a combination of collaborative and content-based filtering) [3], and are assessed using metrics of algorithmic accuracy and precision. Ever since research has focused on creating more accurate algorithms so that the recommendations more accurately reflect the user's interest. [3] However, the algorithmic improvements in accuracy showed to not always lead to improvements in the user experience [50]. A recommender system's objective should not only focus on being accurate, but also on being a pleasure to use for the user [42].

By adopting an approach that includes the user experience instead of focusing only on the algorithm, new insights—and sometimes counterintuitive insights in terms of prediction and recommendation

[3] Since this dissertation focuses on the user experience, no detailed overview of past algorithmic research will be provided. For a comprehensive review on the algorithmic side of recommender systems see [1, 5, 11, 29, 30].

accuracy—were gained, which emphasizes the importance of looking at recommender systems from a user experience perspective. For example, a list of items in which each item is highly accurate may become a bad recommendation as a collective [58]. A list consisting of highly accurate items may cause that the items in the recommendation list become too similar and thereby covering a too narrow interest of the user [7]. Having items that are too similar to each other may also have negative psychological effects on the user. It may cause that the user will experience choice overload, which in turn can lead to an increase of choice difficulties [4, 38, 56].

The user experience research in recommender systems does not only focus on algorithmic deviations, but covers other aspects of the system as well, such as presentation and interaction styles. Research has shown that other factors besides the algorithm play a role in the subjective evaluations of users on the recommender system. For example, including explanations about why a certain recommendation is made has shown to positively influence the user experience, such as scrutability (i.e., users will better understand what to change when recommendations are not satisfying), trust (i.e., increased confidence in the system), efficiency and effectiveness (i.e., let users make faster and better decisions), or persuasiveness (i.e., convince users to try a recommendation), which all at the end influences the user satisfaction with the system and its recommendations (see [51] for an overview on the work on providing explanations in recommender systems).

With the findings of how to improve the user experience in recommender systems, the next step in personalization was made. Moreover, research found nuances in their findings showed that individual differences may occur, which indicate that the (general) results on improving the user experience may affect users differently [4, 48, 56]. Bollen et al. [4] and Willemsen et al. [56] showed in their studies with a movie recommender system that besides the positive effects of recommendation diversity on choice difficulties and satisfaction, personal characteristics (e.g., strength of prior preference, domain expertise [4, 56]) influences the found relationships.

The relation between users' personal characteristics and their preferences for recommender systems have not been investigated a lot yet. Although domain expertise and strength of prior preferences have been identified as influential personal characteristics, it remains challenging to capture these personal characteristics without disturbing the user with questionnaires. Another way to overcome this problem is by analyzing behavior data of users with the system. However, two common problems are arising here: 1) gathering enough behavioral data, and 2) not all the behavioral data is useful for or related to the personalization purpose.

One way to overcome the aforementioned problems, is by exploiting information from external sources. Systems are getting increasingly connected with SNSs through SSO mechanisms. Before SNSs release users' profile information to a system, users need to accept a consent form that states which parts of the profile is going to be accessed by the system. Besides accessing users' basic profile information, systems often ask for additional permissions for accessing other parts of users' profile [10]. By granting access to other parts of the profile, systems are able to unobtrusively infer users' preferences. As this data may also not directly be useful for the targeted purpose, and to be able to effectively use all the information available, general user models are needed to which users' behavior, preferences, and needs can be related to. Once the general user model about a user is created, their preferences can be inferred and applied to the system for a personalized user experience. Two general user models (i.e., personality and cultural dimensions) are used in this dissertation, which are discussed in the following sections.

2.2 Personality

The use of users' personality plays a main role in this dissertation. Personality traits are used as a general user model to investigate the relationship with users' their behavior, preferences, and needs for personalization of a music recommender system. Personality has shown to characterize a person's thoughts, feelings, social adjustments, and behaviors, which subsequently influences their expectations, self-perceptions, values, attitudes, and their reactions to others, problems, and stress [39, 57]. Although recent findings have shown that people can alter their personality, it requires a lot of effort and awareness [34]. In general, people's personality is enduring and remains stable over time, and since it characterizes different patterns of a person, it makes it very well suited for user modeling.

To define a person's personality, different models have been created to categorize personality, where the five-factor model (FFM) is most well known and widely used [41]. The FFM consists of five general dimensions that describe personality. Each of the five dimensions relates to a cluster of correlated primary factors. Table 1 shows the general dimensions with the corresponding primary factors.

General Dimensions	Primary Factors
Openness to Experience	Artistic, curious, imaginative, insightful, original, wide interest
Conscientiousness	Efficient, organized, planful, reliable, responsible, thorough
Extraversion	Active, assertive, energetic, enthusiastic, outgoing, talkative
Agreeableness	Appreciative, forgiving, generous, kind, sympathetic, trusting
Neuroticism	Anxious, self-pitying, tense, touchy, unstable, worrying

Table 1: Five-factor model adopted from [35]

2.2.1 Personality and the Relationship with Behavior, Preferences, and Needs

Until now, personality research has mainly focused on identifying individual differences in real-life situations (i.e., social and physical environments; [40]). How personality-based differences in a technological setting manifest is still relatively unknown. However, some of the findings on personality in real-life situations are already suitable to be implemented in domain-specific technologies. For example, Rentfrow and Gosling [45] identified relationships between personality traits and music preferences, such as cheerful music with vocals showed a positive relationship with extraversion while artistic and intricate music showed a positive relationship with openness to experience. Although these relationships were identified as preferences in real-life situations, they could also be implemented in personalized systems (e.g., the results of the example could be implemented in music recommender systems).

There is an increased interest in identifying personality-based individual differences in technological settings and how to implement the findings for personalization. In an exploratory study on health-promoting applications, Halko and Kientz [28] found indications that conscientious users would be most likely to achieve their health goals and more likely to use and influenced by socially-based technologies. Chen et al. [9] showed personality-based diversity preferences in a movie recommender system, such as neurotic users favor a recommendation list with movies from diverse directors. Tkalcic et al. [53] proposes a method to use personality traits to overcome the ”cold-start problem” in recommender systems. [4]

Including personality information in recommender systems has proven its merits. Hu and Pu [33] showed that personality-based recommender systems are more effective in increasing users' loyalty towards the system and decreasing cognitive effort compared to systems without using personality information. Although the general nature of personality makes it applicable across different domains [6], the domain-specific relationships with personality need to be revealed first so that the system can adopt the right personalization strategy.

2.2.2 Implicit Personality Acquisition

Although the advantages of personality-based systems have been shown and the relationships between personality traits with domain-specific personalizations are being investigated, the acquisition of personality traits is still challenging. The most straightforward way to get to know the personality of users is by using a questionnaire. However, this is often time-consuming and interrupts the user's interaction with the system. Inferring personality traits from behavioral data with the system is problematic as well as enough data for inference is often lacking; especially with new users. Another possibility is by using external information sources, which is made possible by the increased connectedness of systems with SNSs. With the use of SSO mechanisms systems are accessing users' basic profile information. Besides the basic profile information additional permissions are often asked to access other parts of the user profile [10]. By granting access, opportunities are provided to unobtrusively infer users' personality traits.

There is an increase body of work focusing on inferring personality traits from different kind of information sources (see [55] for an overview). As for inferring personality traits from SNSs that could be implemented in a personality-based system, research has been mainly focusing on Facebook [2, 8, 27, 43, 47] and Twitter [26, 44].

[4] The cold-start problem is most prevalent in recommender systems and occurs with new users of the application. It refers to that (almost) no information exists yet about the user to make inferences from.

The research on Facebook has been focusing on different aspects of a user profile to infer user's personality from. Celli et al. [8] looked at profile pictures on Facebook and showed, for example, that extraverts tend to have pictures where they are smiling and are taken with other people. Park et al. [43] used the linguistic features of Facebook users to determine their personality, whereas Ross et al. [47] found relationships between personality and Facebook usage. The research on inferring personality from Twitter has analyzed the linguistic features that are used by a user (e.g., negations, articles; [26]) and how users are connected with their social network (e.g., following, followers; [44]).

2.3 Cultural Dimensions

Next to personality traits, cultural dimensions have been used in some of the works in this dissertation. Whereas personality traits focus on the individual, cultural dimensions describe a whole society: how a culture has its effects on values, which influences behavior of members of the society [31]. As well as with personality traits, the domain-specific relationships with cultural dimensions need to be revealed before we are able to use them for personalization. This is immediately also the disadvantage of using cultural dimensions as it is hard to infer these relationships due to the difficulties to perform such large-scale experiments with enough participants around the world. However, once the relationships are known, the advantage of using cultural dimensions over personality traits is that country information is usually standard in a user profile of the system, whereas personality traits need to be inferred indirectly.

There are different cultural models to which behavior, preferences, and needs can be related to (e.g., Hofstede [31], GLOBE [32] and Trompenaar's [54] cultural dimensions). The most comprehensive model is the Hofstede's model of cultural dimensions [31]. Although this model originates from 1968, it is still being actualized. Hofstede's cultural dimensions comprises 97 countries, which are categorized by six dimensions: power distance index, individualism, uncertainty avoidance index, masculinity, long-term orientation, and indulgence. Each of the dimensions are shortly described below:

Power distance: defines the extent to which power is distributed unequally by less powerful members of institutions (e.g., family). High power distance indicates that a hierarchy is clearly established and executed in society. Low power distance indicates that authority is questioned and attempted to distribute power equally.

Individualism: defines the degree of integration of people into societal groups. High individualism is defined by loose social ties. The main emphasis is on the ”I” instead of the ”we,” while opposite for low individualistic cultures.

Masculinity: defines a society's preference for achievement, heroism, assertiveness and material rewards for success (countries scoring high in this dimension). Whereas low masculinity represents a preference for cooperation, modesty, caring for the weak and quality of life.

Uncertainty avoidance: defines a society's tolerance for ambiguity. High scoring countries in this scale are more inclined to opt for stiff codes of behavior, guidelines, laws. Whereas more acceptance of different thoughts and/or ideas are more common for those scoring low in this dimension.

Long-term orientation: is associated with the connection of the past with the current and future actions and/or challenges. Lower scoring countries tend to believe that traditions are honored and kept, and value steadfastness. High scoring countries believe more that adaptation and circumstantial, pragmatic problem-solving are necessary.

Indulgence: defines in general the happiness of a country. Countries scoring high in this dimension are related to a society that allows relatively free gratification of basic and natural human desires related to enjoying life and having fun (e.g., be in control of their own life and emotions). Whereas low scoring countries show more controlled gratification of needs and regulate it by means of strict social norms.

3 Dissertation Outline

The general scientific contribution of the works presented in this dissertation comes in twofold:

- Showing how users' personality traits and cultural dimensions related to behavior, preferences, and needs
- Implicit acquisition of personality from SNSs.

While the works in each of the aforementioned points have their own novel contributions to the respective fields. The works altogether provide a holistic view of how personality and cultural information can be applied and used to a music recommender system to improve the user experience.

Each of the research conducted can be placed in one of the categories of Figure 1. A short description of each work is discussed in the upcoming sections.

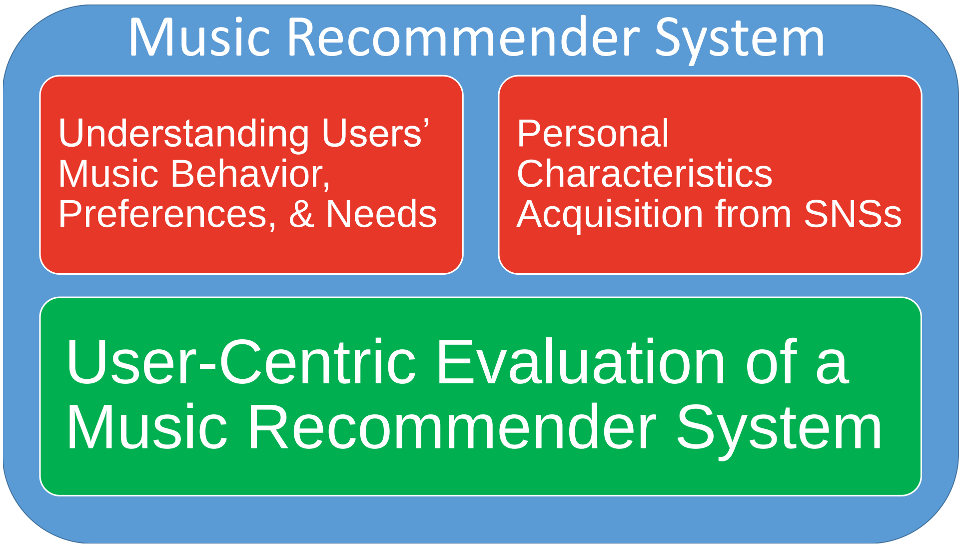

Figure 1: Categorization of the works that have been done.

3.1 Music Systems

Based on the conducted research, several proposals have been made about how to implement the findings. In §4.1 [17] a music recommender system is proposed that besides using personality for personalizing a playlist also takes into account the current mood of the user. The current mood of the users is captured by real-life tracking of the user's Twitter feeds since Tweeting is a real-time activity, which users often use to express their feelings about their activities immediately. This allows for adjusting the music playlist in order to fit the user's current emotional state and their preferred way of emotion regulation.

In §4.2 [12] a system is proposed with which users can easily adjust the expressed emotion of a song by using predefined settings to either fit their emotional state, emotion regulation strategy, or fit a specific context.

The system proposed in §4.3 [14] is the most comprehensive proposal and incorporates the findings of all of the personality-based research in the sections below. It shows how the results can be implemented in different areas of the music recommender system to provide users with a personalized user experience (i.e., user interface, music recommendations, playlist composition).

3.2 Understanding Users' Music Listening Behavior, Preferences, and Needs

The research presented in the following sections reveals relationships between users' personality or cultural dimensions with different behaviors, preferences, and needs.

User Interface
In §5.1 [24] and §5.2 [25] the relationship was explored between personality traits and music browsing strategies of the most popular music taxonomies (i.e., mood, genre, and activity) that are used by systems to organize their music. Furthermore, in §5.2 [25] an extension is presented where choice overload is investigated (i.e., the number of categories after choosing a music taxonomy). For these studies an online music streaming service was simulated where participants were asked to interact with the system. By capturing their navigational behavior in the system and their subjective system evaluation (e.g., satisfaction, attractiveness, preference) through concluding questionnaires, relationships with their personality could be revealed. For example, findings indicate that those scoring high on openness to experience show a high preference for browsing for mu-

sic by mood, while conscientious users show a preference for browsing by activity. The findings of could be used to adapt the user interface to meet the user's preferred way of music browsing.

Listening Needs

The study in §5.3 [19] looked at the relationship between personality traits and emotionally-laden music and how preferences change depending on the emotional state of the user. A controlled experiment was used consisting of the following steps:

1. Assessing participant's initial emotional state
2. Inducing an emotional state (one of the six basic emotions: happy, angry, disgust, fear, sad, surprise) through standardized video clips
3. Assessing participant's emotional state after the video clip (manipulation check)
4. Rating emotionally-laden music (standardized music clips based on the six basic emotions) on the listening likelihood considering the current emotional state

The results show that, in general, users like to listen to music in line with their emotional state. However, individual differences based on personality occur; especially in a negative emotional state (e.g., sadness). We found that when in a negative emotional state, those who scored high on openness to experience, extraversion, and agreeableness tend to cheer themselves up with happy music, while those who scored high on neuroticism tend to prefer to dwell a bit longer in this negative state by listening to sad music. This has important implications for playlist generation. By inferring users' emotional state (e.g., mining user-generated content), the next song can be better targeted toward their needs.

Meta information

In §5.4 [52] the amount of meta information to be presented alongside the music piece was investigated, and explored whether the preferred amount is influenced by the user's personality. The results showed that the following personality traits tend to have a higher preference for more meta information: openness to experience, agreeableness, conscientiousness, and extraversion. This provides implications about the amount of meta information a system should present to the user without them experiencing information overload, which in turn, negatively affects the user experience of the user.

Diversity Needs

In §5.5 [23] and §5.6 [13] the diversity in music listening behavior was explored between countries. A Last.fm dataset was used with listening behavior of 53,309 users of 47 countries. Different diversity measurements were computed on a genre and artist level in order to explore whether differences in listening behavior is culturally embedded. Relationships of diversity in music listening behavior and cultural dimensions were found. For example, music listeners from individualistic countries tend to listen to more diverse music on an artist as well as on a genre level. The results provide implications on how recommendation list diversity should differ based on where the user is from. Individual diversity differences are explored and discussed in §7.1 [16] and §7.2 [15] together with a user-centric evaluation.

3.3 Implicit Acquisition of Users' Personal Characteristics From Social Media

In order to apply the personality-based results for personalization, attention was given to the implicit acquisition of personality traits from SNSs. The most popular SNSs were explored to acquire personality traits from: Facebook, Instagram, and fusing multiple sources (i.e., Twitter and Instagram).

Facebook

A usual prerequisite for implicit personality acquisition is the availability of data. In §6.1 [18] the underlying psychological mechanisms of sharing and posting behavior of Facebook users was investigated. The findings show that sharing and posting behavior often is subject to anticipated regret. This anticipated regret originates from the (wrong) estimation of how the user's audience (i.e., social network) perceive the content. Based on these results, technologies can be developed that analyze the user's social network, which would allow for the creation of proxy measures to counteract on wrong estimations of the content appropriateness by the user.

In §6.2 [21] personality acquisition when data is limited was investigated. The relationship between disclosing or not disclosing of profile sections of a Facebook user profile and personality traits were explored. Results showed that non-disclosure behavior provide indications of a user's personality. The personality predictor that was created showed to be able to approximate the results state-of-the-art methods.

Instagram

Even though Instagram is a popular SNS, it is still unexplored for personality acquisition. In §6.3 [20] and §6.4 [22] a personality predictor was created based on the picture features of user's Instagram pictures (i.e., how filters were applied to modify the characteristics of the pictures). The rational of analyzing Instagram pictures and their relationship with the user's personality comes from that Instagram features different filters that allow users to modify their pictures. The filters facilitate the possibility for users to create and express their identity, which could be related to their personality.

The focus in §6.3 [20] and §6.4 [22] was laid on picture features instead of the type of filter, since the type of filter users use depend on how the original pictures looks like. It could well be that different filters are applied to achieve the same feeling in the end result. By analyzing the picture features, distinct relationships with personality traits were found. For example, very open users tend to apply filters to their pictures so that they express more green tones. A reliable predictor could be created with the relationships found between picture features and personality traits.

Fusing Twitter and Instagram

Since personality acquisition from Twitter has been explored well, the study presented in §6.5 [49] looked at whether combining cues from different SNSs would be able to increase the accuracy of a personality predictor. By using linguistic and meta information from Twitter and linguistic and image information from Instagram, a significant improvement in predicting user's personality was achieved.

3.4 User-Centric Evaluation of a Music Recommender System

As mentioned before: only relying on measurements of algorithmic accuracy and precision is not enough. Applying a user-centered approach may lead to conclusions that deviate from algorithmic evaluations. Therefore, it is important to conduct user-centric evaluation when the goal is to improve the user experience of a system.

The user-centric evaluation in this dissertation could only be on part of the results presented in this dissertation. This was due to the difficulty to recruit enough participants. The user-centric evaluation was done based on music recommendation list diversity. In §7.1 [16] the mechanisms underlying the increased attractiveness due to music list diversity was explored. Findings show that diversifying music recommendation lists is only beneficial if it contributes to the discovery of new music that broadens or deepens the taste of the user.

In §7.2 [15] was investigated how satisfaction and attractiveness with diversified music recommendation lists differ based on the personality of users. The results show that different levels of diversification were preferred based on the user's personality. Conscientious users showed to have a preferences for more diversification in the recommendation list, while agreeable users have a preferences for a mid-level of diversity.

4 Music Systems

Ferwerda, B., & Schedl, M. (2014) Enhancing Music Recommender Systems with Personality Information and Emotional States: A Proposal. *In Proceedings of the 2nd Workshop on Emotions and Personality in Personalized Services held in conjunction with the 22nd Conference on User Modeling, Adaptation and Personalization* (Aalborg, DK).

Ferwerda, B. (2015) The Soundtrack of My Life: Adjusting the Emotion of Music. *In the 1st Workshop on Collaborating with Intelligent Machines held in conjunction with the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, KR).

Ferwerda, B., & Schedl, M. (2016) Personality-Based User Modeling for Music Recommender Systems. *In Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases* (Riva del Garda, IT).

Enhancing Music Recommender Systems with Personality Information and Emotional States: A Proposal

Published in: Extended Proceedings of the 22nd Conference on User Modeling, Adaptation, and Personalization (UMAP 2014): 2nd Workshop on Emotions and Personality in Personalized Systems (EMPIRE 2014; Aalborg, DK)

BRUCE FERWERDA, Johannes Kepler University, Linz (AT)
MARKUS SCHEDL, Johannes Kepler University, Linz (AT)

Abstract

This position paper describes the initial research assumptions to improve music recommendations by including personality and emotional states. By including these psychological factors, we believe that the accuracy of the recommendation can be enhanced. We will give attention to how people use music to regulate their emotional states, and how this regulation is related to their personality. Furthermore, we will focus on how to acquire data from social media (i.e., microblogging sites such as Twitter) to predict the current emotional state of users. Finally, we will discuss how we plan to connect the correct emotionally laden music pieces to support the emotion regulation style of users.

Keywords: Music Recommender Systems, Personality, Emotional States, Emotion Regulation

1 Introduction

Research on recommender systems have shown increased interest to incorporate psychological aspects. Especially the relationship between personality and user preferences has gained a lot of attention. For example, knowledge about the influence of personality traits on music taste [25], and diversity in item recommendations [39] have been exploited to improve the user tailored recommendation. As personality is defined as the individual differences in enduring emotional, interpersonal, experiential, attitudinal and motivational styles [12, 17], one can expect to be able to infer much more based on personality traits to improve the recommendation.

The goal of this project is to improve music recommendations by incorporating additional psychological factors. More specifically, we focus on emotional states and their relationship with personality to infer music taste and preferences. By knowing the user's current emotional state, a system can anticipate its recommendation with an emotionally laden song that is in line with the user's style of emotion regulation (e.g., changing or maintaining their emotional state).

In the following sections we give a brief introduction about what is known about personality and emotional states, and work towards how we are planning to use it to improve recommendations.

2 Personality

Personality has shown to be an enduring factor that influences an individual's behavior [13], interest, and tastes [14, 25]. As personality plays such a prominent role in shaping human preferences, one can expect similar patterns (i.e., behavior, interest, and tastes) to emerge between similar personality traits [2].

Different models have been created to categorize personality, where the five-factor model (FFM) is most well known and widely used [17]. The FFM consists of five general dimensions that describe personality. Each of the five dimensions consist clusters of correlated primary factors. Table 1 shows the general dimensions with the corresponding primary factors.

General dimensions	Primary factors
Openness	artistic, curious, imaginative, insightful, original, wide interest
Conscientiousness	efficient, organized, planful, reliable, responsible, thorough
Extraversion	active, assertive, energetic, enthusiastic, outgoing, talkative
Agreeableness	appreciative, forgiving, generous, kind, sympathetic, trusting
Neuroticism	anxious, self-pitying, tense, touchy, unstable, worrying

Table 1: Five-factor model

There is an emerging interest in how personality relates to user preferences in different domains. This provide valuable information for the development of domain specific recommender systems. Knowing someone's personality can help to infer their preferences [23, 24, 25], and can therefore contribute to a more accurate recommendation. For example, music preferences were found to be correlated with personality traits [25]. Rentfrow and Gosling

[25] categorized music pieces into 4 music-preference dimensions (reflective and complex, intense and rebellious, upbeat and conventional, and energetic and rhythmic), and found correlations with the FFM general dimensions, such as, a relation between energetic and rhythmic music and extraversion and agreeableness.

The prediction of personality parameters is starting to establish by either using implicit acquisition (e.g., personality prediction by extracting data from social media [8, 16, 22]), or explicit acquisition by letting users answering a personality quiz [10]. Although the implicit method is unobtrusive, accuracy is compromised as it dependents on the quality of the source (e.g., frequency of expressing on social media). On the other hand, the explicit method is more accurate, but intrusive and time consuming.

3 Emotional States

We can find emotions in every facet of our life, such as during: decision making, objective and subjective thinking, creativity. To categorize the emotional states we experience, Ekman [5] defined six basic emotional states in which we can categorize experienced emotion: anger, disgust, fear, happiness, sadness, and surprise. Others on the other hand believe that emotions are a mix of dimensions of emotional states [35].

To deal with our emotional states throughout the day, we adapt different strategies. Parkinson and Totterdell [21] defined 162 different strategies (e.g., exercising, music listening, taking a bath). Especially listening to music plays an important role. Research has found that music is the second most strategy used [36, 7]. It can change, create, maintain, or enhance emotions [3]. This suggest that music can play an important supportive role when people dealing with their emotions in daily life.

Just as with personality, there is also an implicit (e.g., blog text) and an explicit way to detect emotion with the same drawbacks. Although the implicit detection has advanced, it goes without saying that automatic capturing of online emotional states remain challenging. As Scherer [28] noted "The inherent fuzziness and the constant evolution of language categories as well as inter-language, inter-cultural, and inter-individual differences make it difficult to define central working concepts".

4 Personality & Emotional States

How we regulate our emotions have been investigated with relation to our personality. Of particular interest are the neuroticism and extraversion dimensions. These dimensions are associated with experiencing negative and positive affect consistently. For example, Tamir [32] found that people scoring high on the neuroticism dimension tend to increase their level of worry. Similarly, people who score low on the extraversion dimension tend to be less motivated to increase their happiness [33].

While most studies are focusing on personality traits in relation to emotion regulation, there is a small area that argues that the emotion regulation style can be explained by one's implicit theory of emotion. In other words whether someone beliefs that emotions are fixed (entity theorist), or more malleable (incremental theorist). Entity theorists experience more negative emotions, that is, less favorable emotion experiences, lower well-being, greater depression, more loneliness, and poorer social adjustments compared to incremental theorists [34].

Music has the ability to induce intense emotions (positive and negative) [40]. Some studies have investigated how, and whether the emotion that consist in music is used by people in their emotion regulation. Thoma et al. [38] categorized different music pieces on valence and arousal, and found that different pieces were preferred depending on the emotionally laden situation. Similarly, Van Goethem and Sloboda [7] found that people use music to support their regulation strategy. For example, music is used to help to distract from the affect or situation, or can help to think about it in a rational way. Despite findings on an individual level (i.e., personality) and how music is used as a regulation strategy, there is still a gap in connecting these two. That is, it is still unknown how music is used to regulate emotions on an individual level.

5 How to improve recommendations?

As music plays a role in emotion regulation of people, and the way how people regulate their emotions seem to be dependent on their personality (or their implicit theories of emotion), the music that people use to support their emotion regulation may also be dependent on their personality.

Whereas personality is usually used to alleviate the cold-start problem in recommender systems (i.e., new users and sparse data sets) [11], or to determine the amount of diversity in the recommendation [39], including emotional states can help to improve music recommendations on the fly.

Currently, music recommender systems anticipate their subsequent recommendations on the music that the user currently listens to. The recommendation is based on similarity by comparing what others with similar taste have listened before (collaborative filtering), or by matching properties (e.g., genre, artist) of the music pieces (content-based filtering). This can result in that recommendations given may fit the user's taste, but may not match the user's actual *needs* at that moment. For example, the systems knows that a user likes Beyoncé. Beyoncé has a range of different emotionally laden songs from up-tempo to ballads. By knowing the user's emotion at a specific moment, the recommender system can anticipate and propose a piece of emotion-laden music that lies within the taste of the user that can support the regulation of the experienced emotion.

5.1 A Scenario

Anna is a 22 year old student. When she listens to music, she often makes use of an online radio. This online radio knows Anna's taste so it can anticipate on the next song to play for Anna. Besides knowing Anna's taste, the system knows that Anna is a little bit neurotic.

On one day Anna is at home, listening to an up-tempo song of Beyoncé. The next song that the radio put in the cue is another up-tempo song, but this time by Katy Perry. Suddenly Anna receives some bad news that makes her sad. She post her feelings on Twitter. The radio system notices this and based on her personality (neuroticsm), it adjust the song in the cue. Instead of playing an up-tempo song, it replaces it with a sad song of Katy Perry. By knowing how Anna likes to regulate her emotions, the system can anticipate the play-list accordingly.

6 Proposal

In the following sections we discuss the initial ideas that we have to improve recommender systems by incorporating the user's personality and current emotional state. We start with describing how we plan to investigate the relationship between personality and emotion regulation through music. After that we discuss the methods for the personality and emotion acquisition from social media, and finally we discuss how we plan to find the emotionally laden songs. For the incorporation of emotion and personality, we assume that system already initiated the user's music taste.

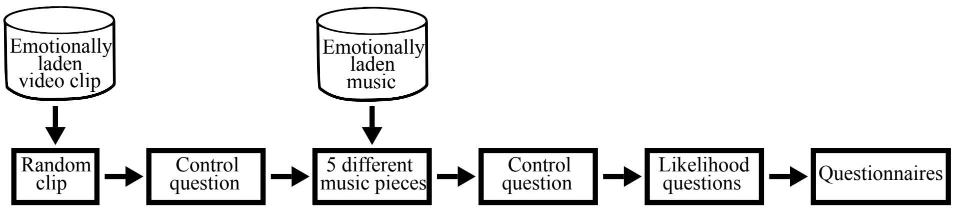

Figure 1: User study work flow

6.1 Step 1: User Study

The first step would be to investigate how people prefer to regulate their emotions and the relationship with their personality. For example, people scoring high on neuroticism tend to increase their level of worry [32]. Therefore, they may not want to listen to music that tries to change their worry state, but want music that is in line with that state instead.

Although there is much research done on the inducing effect of emotional laden music [4, 27], not much is known about how people use music to regulate their emotions. We plan to conduct a online user study using participants on Amazon Mechanical Turk. In this user study we will use the set of film clips (see Figure 1 for the experiment work flow), developed by Hewig et al. [9] to induce one of the basic emotional states. Presenting film clips is one of the methods that is frequently used to induce emotions in psychological experiments. For the user study, we will assign participants randomly to a emotionally laden film clip (anger, disgust, fear, happiness, sadness, or neutral). After showing participants the film clip, we will ask as a control the emotion the film clip induced. In the next step we will present different emotionally laden music fragments (anger, fear, happiness, sadness, and tenderness) and ask participants the likelihood that they would listen to such music when being in the just induced emotional state. The music fragments we will use are categorized by Eerola and Vuoskoski [4] based on the basic emotions they bear. As a control question we will ask in addition what kind of emotion the music pieces induce. To conclude we will ask the FFM questions, implicit theory of emotions questions, and demographics (i.e., age and gender). This will give us information about how music is used in different emotional states and how this is related to personality traits.

With the results of the aforementioned user study we will create a model to predict the emotionally laden music pieces that users would like to listen to when in a certain emotional state. A second user study will be carried out to create the dataset for testing the model. The dataset will contain data about their current emotional states of users and the emotional laden songs they want to listen to. A 10-fold cross-validation method will be used for validation.

Once the model is created and verified, we will know how people prefer to regulate their emotional states with music throughout the day, and how this is related to their personality. In the next step we will move toward the extraction of personality and emotional states from social media.

6.2 Step 2: Personality & Emotional State Acquisition

We will use Twitter as our main source to extract the personality and emotional state parameters from. Tweets are crawled by using the Twitter API. Furthermore, we will limit ourselves to tweets with English as the main language.

Personality Acquisition For the acquisition of personality we will work toward an implicit detection of the parameters, i.e., without the need of a questionnaire. Results of previous research of extracting personality parameters from tweets are promising. Golbeck et al. [8] were able to predict personality parameters from Twitter within 11%-18% of their actual value by looking at the content of users' tweets. Other research done by Quercia et al. [22] were able to estimate personality parameters (RMSE below 0.88 on a [1, 5] scale) by only looking at the users' characteristics (e.g., listeners, popular, highly-read, and influential users). We plan to explore the techniques used by prior research and possibly combining them to improve predictions. Another direction that may be worth taking into account would be to incorporate historical tweets that reflect listening behavior of users. Rentfrow and Gosling [25] found relations between personality and music genres. By looking at historical music tweets of users, we are able to extract the genre of the song which in turn can provide us personality information.

Emotional State Acquisition Although we realize that emotional states are not expressed constantly, we do believe that social media is a platform that is increasingly used by users to express themselves. This includes emotional states depicting personal (e.g., anger, frustrations) to global topics (e.g., politics, sport events) [1, 37].

The acquisition of the emotional states from textual collections of user-generated data on the web has been well established (for an overview of this field see [20]). Results indicate that emotional indicators can be extracted accurately. However, acquisition of these indicators from microblogging sites has been done scarcely. Most of the studies focus on the polarity (positive, negative, or neutral) [20] or try to include the magnitude of the emotion (mild and strong) [30]. Only a few have tried to categorize microblogging text based on existing emotional categorizations [1, 29].

One approach that we bear in mind that to build upon is the use of emotion lexicons. These lexicons consist of terms related to an emotion. Several lexicons have been created based on different emotion categorizations and have been tested on tweets. Such as, Sintsova, Musat, and Pu [29] created a lexicon compatible with the Geneva Emotion Wheel categorization of emotions, Roberts et al. [26] based their emotion lexicon on Ekman's categorization, and Suttles, and Ide [31] on Plutchik's. Especially the work of Roberts et al. [26] would be suitable to build upon as the Ekman's categorization is on the basis of our work. We will be able to complement the predictability of emotions by including metadata that have shown to consist of emotional indicators, for example, hashtags [18, 19], traditional emoticons [6], and emoji [26].

6.3 Step 3: Emotion Classification of Songs

The user study (see §6.1) will give us insights in how emotionally laden songs are used in the emotion regulation process. For the system to be able to anticipate its recommendation, we need to find the right emotional annotated song.

The field of emotion classification in music is still evolving (see for an overview [15]). Currently, different methods are used to annotate music pieces on their emotion: direct human annotation (e.g., surveys, social tags, games), indirect human annotation (e.g., web documents, social tag clouds, lyrics), or content-based analysis (e.g., audio, images, video). As Kim et al. [15] noted "Recognizing musical mood remains a challenging problem primarily due to the inherent ambiguities of human emotions." To find the right emotionally laden song within a collection in this project, we will initially turn to the tags provided by Last.FM website. Last.FM currently provide songs with the tags happy, sad, angry, and relaxed. Based on these tags we can make a first attempt to match emotionally laden songs with the user's way of regulating their emotional state.

7 Conclusion

By including the user's current emotional state, we propose that music recommendations can be improved. Our next efforts will be to investigate how people regulate their emotions with music. That is, what kind of emotionally laden music people are listening when being in a specific emotional state.

Additionally, we will investigate how emotion regulation with music is related to their personality.

For the acquisition of personality and emotion, we will focus on microblogging sites. As social media generates a constant stream of communication, we believe that microblogging sites as Twitter are suitable to extract personality and emotional states of users. Although accurate results are achieved from textual collections of user-generated data on the web, analyzing microblogging sites remains challenging. The amount of text is scarce as the text posted on Twitter is limited to 140 characters. However, the ability to express oneself in a short and fast way lend itself to post content more often. To extract personality and emotional states from Twitter feeds we will initially trust on different existing methods and combine them to improve predictability of the parameters.

With the findings of the aforementioned steps, we can start matching music that fits the user's way of emotion regulation. To find suitable music, we will initially rely on the emotional tags that Last.FM provides.

8 Acknowledgment

This research is supported by the European FP7/2007-2013 program through the PHENICX project under grant agreement no. 601166 and by the Austrian Science Fund (FWF): P25655.

References

[1] Bollen, J., Mao, H., Pepe, A.: Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In: ICWSM (2011)

[2] Burger, J.: Personality. PSY 235 Theories of Personality Series, Cengage Learning (2010)

[3] DeNora, T.: Music as a technology of the self. Poetics 27(1), 31–56 (1999)

[4] Eerola, T., Vuoskoski, J.K.: A comparison of the discrete and dimensional models of emotion in music. Psychology of Music 39(1), 18–49 (2011)

[5] Ekman, P.: An argument for basic emotions. Cog. & Emotion 6, 169–200 (1992)

[6] Go, A., Bhayani, R., Huang, L.: Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford pp. 1–12 (2009)

[7] van Goethem, A., Sloboda, J.: The functions of music for affect regulation. Musicae scientiae 15(2), 208–228 (2011)

[8] Golbeck, J., Robles, C., Edmondson, M., Turner, K.: Predicting personality from twitter. In: Privacy, security, risk and trust (passat), 2011 ieee third international conference on and 2011 ieee third international conference on social computing (socialcom). pp. 149–156. IEEE (2011)

[9] Hewig, J., Hagemann, D., Seifert, J., Gollwitzer, M., Naumann, E., Bartussek, D.: Brief report: A revised film set for the induction of basic emotions. Cognition & Emotion 19(7), 1095–1109 (2005)

[10] Hu, R., Pu, P.: Using personality information in collaborative filtering for new users. Recommender Systems and the Social Web p. 17 (2010)

[11] Hu, R., Pu, P.: Enhancing collaborative filtering systems with personality information. In: Proceedings of the fifth ACM conference on Recommender systems. pp. 197–204. ACM (2011)

[12] John, O.P., Srivastava, S.: The big five trait taxonomy: History, measurement, and theoretical perspectives. Handbook of personality: Theory and research 2, 102–138 (1999)

[13] Jung, C.: Psychological Types. No. v. 6 in Bollingen Series No. 20, Princeton University Press (1976)

[14] Kemp, A.E.: The musical temperament: Psychology and personality of musicians. Oxford University Press (1996)

[15] Kim, Y.E., Schmidt, E.M., Migneco, R., Morton, B.G., Richardson, P., Scott, J., Speck, J.A., Turnbull, D.: Music emotion recognition: A state of the art review. In: Proc. ISMIR. pp. 255–266. Citeseer (2010)

[16] Kosinski, M., Stillwell, D., Graepel, T.: Private traits and attributes are predictable from digital records of human behavior. Proceedings of the National Academy of Sciences 110(15), 5802–5805 (2013)

[17] McCrae, R.R., John, O.P.: An introduction to the five-factor model and its applications. Journal of personality 60(2), 175–215 (1992)

[18] Mohammad, S.M.: # emotional tweets. In: Proceedings of the First Joint Conference on Lexical and Computational Semantics-Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation. pp. 246–255. Association for Computational Linguistics (2012)

[19] Mohammad, S.M., Kiritchenko, S.: Using hashtags to capture fine emotion categories from tweets. Computational Intelligence (2014)

[20] Pang, B., Lee, L.: Opinion mining and sentiment analysis. Foundations and trends in information retrieval 2(1-2), 1–135 (2008)

[21] Parkinson, B., Totterdell, P.: Classifying affect-regulation strategies. Cognition & Emotion 13(3), 277–303 (1999)

[22] Quercia, D., Kosinski, M., Stillwell, D., Crowcroft, J.: Our twitter profiles, our selves: Predicting personality with twitter. In: Privacy, security, risk and trust (passat), 2011 ieee third international conference on and 2011 ieee third international conference on social computing (socialcom). pp. 180–185. IEEE (2011)

[23] Rawlings, D., Ciancarelli, V.: Music preference and the five-factor model of the neo personality inventory. Psychology of Music 25(2), 120–132 (1997)

[24] Rentfrow, P.J., Goldberg, L.R., Zilca, R.: Listening, watching, and reading: The structure and correlates of entertainment preferences. Journal of personality 79(2), 223–258 (2011)

[25] Rentfrow, P.J., Gosling, S.D.: The do re mi's of everyday life: the structure and personality correlates of music preferences. Journal of personality and social psychology 84(6), 1236 (2003)

[26] Roberts, K., Roach, M.A., Johnson, J., Guthrie, J., Harabagiu, S.M.: Empatweet: Annotating and detecting emotions on twitter. In: LREC. pp. 3806–3813 (2012)

[27] Scherer, K.R.: Which emotions can be induced by music? what are the underlying mechanisms? and how can we measure them? Journal of new music research 33(3), 239–251 (2004)

[28] Scherer, K.R.: What are emotions? and how can they be measured? Social science information 44(4), 695–729 (2005)

[29] Sintsovaa, V., Musata, C.: Fine-grained emotion recognition in olympic tweets based on human computation. WASSA 2013 12 (2013)

[30] Strapparava, C., Mihalcea, R.: Learning to identify emotions in text. In: Proceedings of the 2008 ACM symposium on Applied computing. pp. 1556–1560. ACM (2008)

[31] Suttles, J., Ide, N.: Distant supervision for emotion classification with discrete binary values. In: Computational Linguistics and Intelligent Text Processing, pp. 121–136. Springer (2013)

[32] Tamir, M.: Don't worry, be happy? neuroticism, trait-consistent affect regulation, and performance. Journal of personality and social psychology 89(3), 449 (2005)

[33] Tamir, M.: Differential preferences for happiness: Extraversion and trait-consistent emotion regulation. Journal of Personality 77(2), 447–470 (2009)

[34] Tamir, M., John, O.P., Srivastava, S., Gross, J.J.: Implicit theories of emotion: affective and social outcomes across a major life transition. Journal of Personality and Social Psychology 92(4), 731 (2007)

[35] Thayer, R.E.: The biopsychology of mood and arousal. Oxford University Press (1989)

[36] Thayer, R.E., Newman, J.R., McClain, T.M.: Self-regulation of mood: strategies for changing a bad mood, raising energy, and reducing tension. Journal of personality and social psychology 67(5), 910 (1994)

[37] Thelwall, M., Buckley, K., Paltoglou, G.: Sentiment in twitter events. Journal of the American Society for Information Science and Technology 62(2), 406–418 (2011)

[38] Thoma, M.V., Ryf, S., Mohiyeddini, C., Ehlert, U., Nater, U.M.: Emotion regulation through listening to music in everyday situations. Cognition & emotion 26(3), 550–560 (2012)

[39] Wu, W., Chen, L., He, L.: Using personality to adjust diversity in recommender systems. In: Proceedings of the 24th ACM Conference on Hypertext and Social Media. pp. 225–229. ACM (2013)

[40] Zentner, M., Grandjean, D., Scherer, K.R.: Emotions evoked by the sound of music: characterization, classification, and measurement. Emotion 8(4), 494 (2008)

The Soundtrack of My Life: Adjusting the Emotion of Music

Published in: ACM Conference on Human Factors in Computing Systems (CHI 2015): Collaborating with Intelligent Machines, extended proceedings (Seoul, KR)

BRUCE FERWERDA, Johannes Kepler University, Linz (AT)

Abstract

In this position paper for the CHI2015 workshop “Collaborating with Intelligent Machines," we propose an application to easily adjust the emotion of a musical melody by using predefined settings. For the adjustment of the melody we rely on music production rules and musical structural characteristics related to predefined emotions that have been defined by prior work.

Keywords: Music Recommender Systems, Emotional States, Emotion Regulation

1 Proposal

Music plays an import role in our lives. Out of 162 emotion regulation strategies, listening to music has been listed as the second strategy that people most often use [6, 9]. This is not surprising, as music has shown to be able to induce bona fide emotions [7, 12].

Professional composers have shown to be able to express effectively the intended emotion of their music piece to their audience [10]. Although, we intuitively would say that the emotion of a music piece is conveyed by the lyrics, studies indicate that the piece’s emotion mainly lies within its melody. Even when lyrics and melody conflict in their expressed emotion, the melody has been proven dominant in emotion determination [1]. This provides opportunities to adjust the music piece’s emotion to any given situation regardless of the lyrics.

In this position paper, we propose an application that can adjust the emotion of a music piece in a simple manner. To adjust the emotion of a song, we rely on production rules. It has been argued that the experienced emotion of a music piece is determined by a multiplication of features: structural features, performance features, listener features, and contextual features [8]. A short description of each feature is found below.

- Structural features:
  - Acoustic as well as the foundational structures of a piece that makes up the music.
- Performance features:
  - The manner in which a piece is executed by the performer. This includes physical appearance, reputation of the performer.
- Listener features:
  - Social and individual identity of the listener (e.g., personality, musical knowledge).
- Contextual features:
  - Location and occasion where the music piece is played.

The magnitude of the felt emotion increases as more features are positively related. The effect of each feature is compounded by one another. As performance features are not adjustable in the context of the application as we propose, we further focus only on the other features in this position paper.

Previous work investigated the range of different musical structural characteristics to communicate emotions. For example, happy music has been identified with a mean of 4.984 notes per second, while sad music has 1.333 [2]. With this information we propose predefined buttons in the application that can adjust music structural features according to the desired emotion (see Figure 1 for a mock-up UI).

Similarly, buttons could be created for contextual features in the same way as proposed for structural features. Often the desired emotion in a certain context is already known (e.g., sad music at a funeral). Instead of buttons with emotion labels, buttons with predefined context could be created that in turn adjust the musical characteristic of the music piece to fit the context.

In addition, a more personalized approach is possible by depicting the listener features. A possible focus of listener features is personality. Some studies have shown that there are individual differences in how people perceive (e.g., [11]) and prefer music (e.g., [4]). For example, high agreeable or neurotic people tend to have stronger sad feelings [11]. Based on one's personality, a musical emotional adjustment can be strengthened or weakened. For example, sad music has been identified with a mean of 1.333 notes per second, with a lower bound of 1.112 and an upper bound of 1.554. Based on one's personality it can be decided to adjust the music piece more towards the upper -or lower bound of the emotion.

Figure 1: Mock-up of a music player interface that adjust the music according to the chosen emotion.

With this position paper we join the emergence of music applications that include personality and/or affect information (e.g., [3, 5]). Our proposed application supports easy music creation for every situation by adjusting the conveyed emotion of the piece. It can also be useful for off-site collaboration in the creation of music by being able to easily try out different emotional settings.

Acknowledgements

This research is supported by the EU FP7/'13-'16 through the GiantSteps project under grant agreement no. 610591.

References

[1] Ali, S.O., Peynircioğlu, Z.F.: Songs and emotions: are lyrics and melodies equal partners? Psychology of Music 34(4), 511–534 (2006)

[2] Bresin, R., Friberg, A.: Emotion rendering in music: range and characteristic values of seven musical variables. Cortex 47(9), 1068–1081 (2011)

[3] Ferwerda, B., Schedl, M.: Enhancing music recommender systems with personality information and emotional states: A proposal. In: 2nd EMPIRE Workshop. pp. 1613–0073 (2014)

[4] Ferwerda, B., Schedl, M., Tkalcic, M.: Personality & emotional states: Understanding users' music listening needs. UMAP 2015 Extended Proceedings

[5] Ferwerda, B., Yang, E., Schedl, M., Tkalcic, M.: Personality traits predict music taxonomy preferences. In: CHI'15 Extended Abstracts on Human Factors in Computing Systems. ACM (2015)

[6] Parkinson, B., Totterdell, P.: Classifying affect-regulation strategies. Cognition & Emotion 13(3), 277–303 (1999)

[7] Scherer, K.R.: Which emotions can be induced by music? what are the underlying mechanisms? and how can we measure them? Journal of new music research 33(3), 239–251 (2004)

[8] Scherer, K.R., Zentner, M.R.: Emotional effects of music: Production rules. Music and emotion: Theory and research pp. 361–392 (2001)

[9] Thayer, R.E., Newman, J.R., McClain, T.M.: Self-regulation of mood: strategies for changing a bad mood, raising energy, and reducing tension. Journal of personality and social psychology (1994)

[10] Thompson, W.F., Robitaille, B.: Can composers express emotions through music? Empirical Studies of the Arts 10(1), 79–89 (1992)

[11] Vuoskoski, J.K., Eerola, T.: Measuring music-induced emotion a comparison of emotion models, personality biases, and intensity of experiences. Musicae Scientiae 15(2) (2011)

[12] Zentner, M., Grandjean, D., Scherer, K.R.: Emotions evoked by the sound of music: characterization, classification, and measurement. Emotion 8(4), 494 (2008)

Personality-Based User Modeling for Music Recommender Systems

Published in: Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECMLPKDD 2016; Riva del Garda, IT)

BRUCE FERWERDA, Johannes Kepler University, Linz (AT)
MARKUS SCHEDL, Johannes Kepler University, Linz (AT)

Abstract

Applications are getting increasingly interconnected. Although the interconnectedness provide new ways to gather information about the user, not all user information is ready to be directly implemented in order to provide a personalized experience to the user. Therefore, a general model is needed to which users' behavior, preferences, and needs can be connected to. In this paper we present our works on a personality-based music recommender system in which we use users' personality traits as a general model. We identified relationships between users' personality and their behavior, preferences, and needs, and also investigated different ways to infer users' personality traits from user-generated data of social networking sites (i.e., Facebook, Twitter, and Instagram). Our work contributes to new ways to mine and infer personality-based user models, and show how these models can be implemented in a music recommender system to positively contribute to the user experience.

Keywords: Personalization, Music Recommender Systems

1 Introduction

An abundance of information about users is getting available with the increased interconnectedness of applications, which provide new ways to tackle problems that systems, such as recommender systems are facing (e.g., lacking behavioral data to infer preferences, such as with the "cold-start problem"). [1] For example, the implementation of single sign-on (SSO)

[1] The cold-start problem is most prevalent in recommender systems and occurs with new users of the application. It refers to that (almost) no information exists yet about the user to make inferences from.

mechanisms [2] allow users to easily login and register to the application, but also let applications import user information from the connected application, which could be used for personalization.

Although with the interconnectedness of applications new information sources become available, not all the new information is directly applicable to create personalized experiences with. Therefore, a general user model is needed to which users' behavior, preferences, and needs can be connected to in order to create personalized experiences for users. This allows the creation of only one user model that can be used across applications without the need of information that is directly related to a specific behavior, preference, or need of the user [1].

We model users based on their personality to make inferences about their behavior, preferences, and needs. Personality has shown to be a stable and enduring factor, which influences an individual's behavior, interest, and taste. As personality plays such a prominent role in shaping human preferences, one can expect similar patterns (i.e., behavior, interest, and taste) to emerge for people with similar personality traits, which makes it suitable for user modeling. In our works, we rely on the widely used five-factor model (FFM), which categorizes personality into five traits: openness to experience (O), conscientiousness (C), extraversion (E), agreeableness (A), and neuroticism (N) [10].

In the next sections we provide an overview of our works on user modeling, which comes in twofold: 1) understanding the relationship between personality traits of users and their behavior, preferences, and needs, and 2): implicit acquisition of users' personality traits from social media.

2 Understanding the User

In order to create personality-based recommender systems, the relationship with their behavior, preferences, and needs need to be identified first. We conducted several user studies on different aspects of the user experience in music recommender systems in order to identify relationships with users' personality.

Listening needs. In [4] we aimed to understand the music listening needs of users in order to provide better personalized recommendations. We investigated the relationship between personality traits and the preference for different kinds of music, and how these preferences change depending on users' emotional state. Our findings show that, in general, users

[2] Buttons that allow users to register or login with accounts of other applications. For example, social networking services: "Login with your Facebook account."

like to listen to music in line with their emotional state. However, individual differences based on personality occur; especially in a negative emotional state (e.g., sadness). We found that when in a negative emotional state, those who scored high on openness to experience, extraversion, and agreeableness tend to cheer themselves up with happy music, while those who scored high on neuroticism tend to prefer to dwell a bit longer in this negative state by listening to sad music. This has important implications for playlist generation. By inferring users' emotional state (e.g., mining user-generated content), the next song can be better targeted toward their needs.

Meta information. In [14] we investigated the amount of meta information a user would want about the music pieces that is listened to. The results showed that the following personality traits tend to have a higher preference for more meta information: openness to experience, agreeableness, conscientiousness, and extraversion. This provides implications about the amount of meta information a system should present to the user without them experiencing information overload, which in turn, negatively affects the user experience of the user.

User interface. In [8] we simulated an online music streaming service to identify the relationship between personality traits and the way users browse for music. By exploring the most frequently used taxonomies to categorize music (i.e., by genre, activity, mood), we were able to identify distinct music browsing behavior based on users' personality, which could be used to create adaptive user interfaces. For example, findings indicate that those scoring high on openness to experience show a high preference for browsing for music by mood, while conscientious users show a preference for browsing by activity.

3 Acquisition of Users' Personality Traits

Besides identifying relationships between personality traits and users' behavior, preferences, and needs, we also looked into the implicit personality acquisition of users. We specifically focused on personality acquisition from social networking sites (SNSs: e.g., Facebook, Twitter, Instagram), as they are getting increasingly interconnected through SSO buttons. Besides accessing users' basic profile information, applications often ask for additional permissions to access other parts of the users profile [2]. By granting access, applications are able to unobtrusively infer users' personality traits. We report the RMSE on personality trait prediction (i.e., O, C, E, A, N) for each of our work below ($r \epsilon$ [1,5]).

Several works exist that show that it is possible to infer personality traits from user-generated data of SNSs (e.g., Facebook [11], and Twitter [9, 12]). In [5, 7] we add to the work on SNS analyses by inferring personality traits from users' Instagram picture features. We showed that personality traits are related to the way Instagram users modify their pictures with filters, and a reliable personality predictor can be created based on that ($RMSE$: O=.68, C=.66, E=.90, A=.69, N=.95). For example, open users tend to apply filters to their pictures in order to make them look more greenish. In [13] we tried to increase the prediction accuracy by fusing information from different SNSs (i.e., Instagram and Twitter). We show a significant improvement of the prediction accuracy when combining different sources ($RMSE$: O=.51, C=.67, E=.71, A=.50, N=.73).

One problem with the implicit acquisition of personality is that when users are not sharing information, the acquisition fails. We investigated this problem from two different directions: 1) understanding the underlying mechanisms of sharing information, 2) personality acquisition with limited user information.

In [3] we found that the lack of sharing and posting comes from the uncertainty of approval of the users viewing the posts. We were able to increasing sharing and posting by analyzing the user's social network and create proxy measures about how the shared or posted content would be received.

In [6] we looked at whether or not disclosing Facebook profile information reveals personality as well. By solely analyzing whether profile sections were disclosed or not (e.g., occupation, education), disregarding their actual content, we were able to create a personality predictor that is able to approximate the prediction accuracy of methods extensively analyzing content ($RMSE$: O=.73, C=.73, E=.99, A=.73, N=.83). This provide opportunities to still being able to infer users' personality even when they are not disclosing information.

4 Conclusion

This paper gave an overview of our work on creating personalized experiences in music recommender systems. We revealed relationships between personality traits and different user behavior, needs, and preferences to improve the user experience, and showed how personality can be mined and inferred using the increased connectedness between applications and SNSs.

5 Acknowledgment

This research is supported by the Austrian Science Fund (FWF): P25655.

References

[1] Cantador, I., Fernández-tobías, I., Bellogín, A.: Relating Personality Types with User Preferences in Multiple Entertainment Domains. EMPIRE (2013)

[2] Chia, P.H., Yamamoto, Y., Asokan, N.: Is this app safe?: a large scale study on application permissions and risk signals. In: WWW. ACM (2012)

[3] Ferwerda, B., Schedl, M., Tkalcic, M.: To post or not to post: The effects of persuasive cues and group targeting mechanisms on posting behavior. In: SocialCom (2014)

[4] Ferwerda, B., Schedl, M., Tkalcic, M.: Personality & emotional states: Understanding usersâĂŹ music listening needs. In: UMAP (2015)

[5] Ferwerda, B., Schedl, M., Tkalcic, M.: Predicting personality traits with instagram pictures. In: EMPIRE (2015)

[6] Ferwerda, B., Schedl, M., Tkalcic, M.: Personality traits and the relationship with (non-) disclosure behavior on facebook. In: WWW (2016)

[7] Ferwerda, B., Schedl, M., Tkalcic, M.: Using instagram picture features to predict users' personality. In: MMM (2016)

[8] Ferwerda, B., Yang, E., Schedl, M., Tkalcic, M.: Personality traits predict music taxonomy preferences. In: CHI Extended Abstracts (2015)

[9] Golbeck, J., Robles, C., Edmondson, M., Turner, K.: Predicting personality from twitter. In: SocialCom (2011)

[10] McCrae, R.R., John, O.P.: An introduction to the five-factor model and its applications. Journal of Personality (1992)

[11] Park, G., Schwartz, H.A., Eichstaedt, J.C., Kern, M.L., Kosinski, M., Stillwell, D.J., Ungar, L.H., Seligman, M.E.: Automatic personality assessment through social media language. Journal of personality and social psychology 108(6), 934 (2015)

[12] Quercia, D., Kosinski, M., Stillwell, D., Crowcroft, J.: Our twitter profiles, our selves: Predicting personality with twitter. In: SocialCom (2011)

[13] Skowron, M., Tkalčič, M., Ferwerda, B., Schedl, M.: Fusing social media cues: personality prediction from twitter and instagram. In: WWW (2016)

[14] Tkalcic, M., Ferwerda, B., Hauger, D., Schedl, M.: Personality correlates for digital concert program notes. In: UMAP (2015)

5 Understanding Users and Their Music Listening Behavior, Preferences, and Needs

Ferwerda, B., Yang, E., Schedl, M., & Tkalcic, M. (2015) Personality Traits Predict Music Taxonomy Preferences. *In Extended Abstracts Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, KR).

Ferwerda, B., Yang, E., Schedl, M., & Tkalcic, M. (2016) Personality and Taxonomy Preferences, and the Influence of Category Choice Set Size on the User Experience for Music Streaming Services. *Under Review*.

Ferwerda, B., Schedl, M., & Tkalcic, M. (2015). Personality & Emotional States: Understanding Users' Music Listening Needs. *In Proceedings of the 23rd Conference on User Modeling, Adaptation and Personalization* (Dublin, IR).

Tkalcic, M., **Ferwerda, B.**, Hauger, D., & Schedl, M. (2015) Personality Correlates for Digital Concert Program Notes. *In Proceedings of the 23rd Conference on User Modeling, Adaptation and Personalization* (Dublin, IR).

Ferwerda, B., Tkalcic, M., & Schedl, M. (2016). Exploring Music Diversity Needs Across Countries. *In Proceedings of the 24th Conference on User Modeling, Adaptation and Personalization* (Halifax, NS, CA).

Ferwerda, B., & Schedl, M. (2016) Investigating the Relationship Between Diversity in Music Consumption Behavior and Cultural Dimensions: A Cross-country Analysis. *In Extended Proceedings of the 24th Conference on User Modeling, Adaptation and Personalization: 1st Workshop on Surprise, Opposition, and Obstruction in Adaptive and Personalized Systems* (Halifax, NS, CA).

Personality & Emotional States: Understanding Users' Music Listening Needs

Published in: Proceedings of the 23rd Conference on User Modeling, Adaptation, and Personalization (UMAP 2015; Dublin, IE)

BRUCE FERWERDA, Johannes Kepler University, Linz (AT)
MARKUS SCHEDL, Johannes Kepler University, Linz (AT)
MARKO TKALCIC, Johannes Kepler University, Linz (AT)

Abstract

This position paper describes the initial research assumptions to improve music recommendations by including personality and emotional states. By including these psychological factors, we believe that the accuracy of the recommendation can be enhanced. We will give attention to how people use music to regulate their emotional states, and how this regulation is related to their personality. Furthermore, we will focus on how to acquire data from social media (i.e., microblogging sites such as Twitter) to predict the current emotional state of users. Finally, we will discuss how we plan to connect the correct emotionally laden music pieces to support the emotion regulation style of users.

Keywords: Music Recommender Systems, Personality, Emotional States, Emotion Regulation

1 Introduction

We experience emotions in every facet of our life (e.g., during decision making, thinking, creativity), and our behavior is influenced by the emotional state we are in [16]. To regulate our emotional states, we rely on 162 different methods where listening to music is the second most used method [9].

Within an emotion regulation method, we can adopt different strategies, such as, changing, enhancing, or maintaining our emotional states. Previous research has found that the preferred strategy is based on individual differences (e.g., [11]). For example, some people prefer to be cheered up when feeling sad, while others would like to stay in this emotional state a bit longer. Research has shown that composers are able to effectively express the intended emotion of their song to their audience [13], and that music is able to induce bona fide emotions [16]. The ability of music to express and induce different kind of emotions makes it well suited to support emotion regulation.

Current music applications that feature the ability for users to listen to music that fits their emotional state, assume that they want to listen to music in line with how they feel. However, since people adopt different emotion regulation strategies, they may not always desire music which is similar with their emotional state. Hence, in order to recommend the most appropriate music for users and their current emotional state, understanding music listening needs on a general as well as on an individual level is needed.

With this work we seek to expand the understanding of how one's emotional state relates to the preferred (emotionally laden) music. Music is known to regulate emotions, but it is not known *how* emotional states relate to different types of emotionally laden music preferences, nor how preference differences breakdown on an individual level by looking at personality traits.

This leads us to the following research questions:

1. How do emotional states relate to emotionally laden music preferences?
2. How do personality traits relate to emotionally laden music preferences?

An online user study was conducted where participants were asked to rate different emotionally laden music pieces on the listening likelihood based on their current emotional state. Among 359 participants we found that the emotional state is related to the use of emotionally laden music. Furthermore, individual differences were identified based on personality traits.

We continue with the related work, method, findings, and discussion.

2 Related work

Ample research has investigated the effects of music and the importance of it in everyday life. For example, Thompson and Robitaille found that composers are effective in transferring the intended emotion of the music pieces to their audience [13], indicating that people perform well at interpreting music emotions. People are not only good at recognizing emotions in music, but music is also able to induce emotions in such a way that it is used as

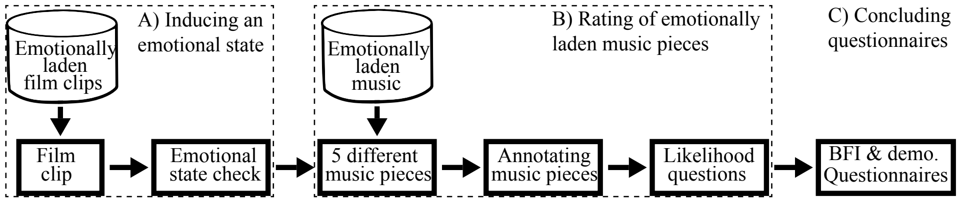

Figure 1: User study work flow. A) We induced and checked participants' emotional state, then B) let them annotate emotionally laden music and asked the likelihood of listening to it, and C) asked to fill in the personality (BFI) questionnaire and demographics.

experimental stimuli (e.g., [16]). Others have investigated how people use music. Parkinson and Totterdell surveyed affect-regulation strategies, and found that music is used as a common means [9]. Although the effects and usages of music has been extensively investigated, it is striking that to our knowledge, no research has focused on *how* emotionally laden music is used to regulate which emotion and *how* differences exist on an individual level. With this work we try to answer these questions.

3 Method

Procedure. We developed an online experiment to get insights into the relationship between emotional states and emotionally laden music (Figure 1). In this experiment, participants were put in an emotional state and were asked to rate different emotionally laden music pieces on the listening likelihood, based on their emotional state. Participants were recruited (N=382) through Amazon Mechanical Turk. Participation was restricted to those located in the United States, and with a very good reputation. Several comprehension-testing questions were used to filter out fake and careless entries. This left us with 359 completed and valid responses. Gender (174 men and 185 women) and age (range 19-68, median 31) information indicated an adequate distribution.

Participants were informed that an emotional state is going to be induced and were given a consent form. The study started with an example to familiarize participants with the study. To induce an emotion, we used the film clips presented in Table 1. The film clip of *Hannah and her Sisters* was always shown in the example to provide a constant (neutral) baseline stimulus [5]. For the actual study, we randomly assigned the remaining film clips. A short synopsis was provided before playback to increase involvement, and improve understanding of the film clips' content [15]. In line with the procedure of Hewig et al., we asked participants at the end of the film

Table 1: The film clips with their length, and emotion.

Film Clip	Length (s)	Emotion
Hannah and her Sisters	92	Neutral
Crimes and Misdemeanors	63	Neutral
All the President's Men	65	Neutral
An Officer and a Gentleman	111	Happy
When Harry met Sally	149	Happy
Witness	91	Anger
My Bodyguard	236	Anger
Silence of the Lambs	202	Fear
Halloween	208	Fear
An Officer and a Gentlemen	101	Sad
The Champ	171	Sad
Maria's Lovers	58	Disgust
Pink Flamingos	29	Disgust

Table 2: The albums with the track number, length, and emotion.

Album (track number)	Length (s)	Emotion
The Rainmaker (3)	18	Happy
Batman (18)	20	Happy
Lethal Weapon 3 (8)	14	Anger
The Rainmaker (7)	15	Anger
Batman Returns (5)	16	Fear
JFK (8)	14	Fear
The English Patient (18)	25	Sad
Running Scared (15)	19	Sad
Shine (10)	20	Tender
Pride & Prejudice (1)	16	Tender

clip to indicate how *they were feeling* (by selecting an emotion from the set as seen in Table 1), and not what they thought the film clip was suppose to express [7]. In the next step, five emotionally laden music pieces, from within and between the emotion categories (Table 2), were randomly presented. Participants were asked to annotate the emotion they thought the music piece was trying to express (by selecting an emotion from the set as seen in Table 2), and the likelihood (5-point Likert scale; never-always) of listening to such emotionally laden music, considering their (reported) emotional state. The study ended with the personality questionnaire and demographics.

Materials. For our stimuli we relied on existing materials that have been tested in prior studies. To induce an emotional state, we used film clips as they are the most powerful emotion elicitation technique in a controlled environment [10]. Hewig et al. designed film clips to induce an emotional state *without sound*, and categorized them based on Ekman's emotion categorization (anger, fear, happy, surprise, disgust, and sad; Table 1) [7]. Using muted stimuli allowed us to control for conflicts with our music pieces in the annotation step of the study. [1]

We used the emotionally laden music pieces created by Eerola and Vuoskoski. They defined music pieces based on the emotional value they bear, based on Ekman's emotion categorization. Film soundtracks were used as they are created with the purpose to mediate powerful emotional cues. Additionally, as they are instrumental, they are relatively neutral in terms of musical preferences and (artist) familiarity (Table 2) [1]. [2]

We assessed personality traits with the widely used 44-item Big Five Inventory (BFI; 5-point Likert scale; disagree strongly - agree strongly [8]), which

[1] As the surprise emotion only lasts seconds [2], we decided not to include this.

[2] Eerola and Vuoskoski [1] replaced disgust with tender, as disgust is rarely expressed by music. Music depicting surprise was omitted due to lack of statistical significance.

describes personality in terms of openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism.

4 Findings

The analyses were done based on participants' reported emotional state after the film clip was shown, not on the intended induced emotion by the film clip. The distribution of the reported emotional states were as follows: happy (n=55), neutral (n=82), anger (n=62), disgust (n=56), fear (n=79), and sad (n=61).

An initial one-way multivariate analysis of variance (MANOVA) was conducted to test the relationship between emotionally laden music pieces and emotional states. A significant MANOVA effect was obtained (Wilks' Lambda = .523, $F(25, 1164.24) =$ 8.89, $p <$.001) with a moderate effect size (η^2=.13). The homogeneity of variance assumption was tested for all the emotionally laden music pieces. Levene's F test showed that the music pieces depicting the emotional states *happy* and *tender* do not meet the requirement of $p >$.05. None of the largest standard deviations of the two pieces were more than four times the size of the corresponding smallest, suggesting that follow-up ANOVAâĂŹs are robust.

Post-hoc tests (Tukey HSD) were performed to examine individual mean difference comparisons across the six emotional states and the five emotionally laden music pieces. The results reported here were compared against a neutral emotional state and were all statistically significant ($p <$.05). Results revealed that in general, participants preferred *happy* and *tender music* when in a *neutral emotional state*. However, in a *angry* or *disgusted state*, participants preferred *angry* or *fearful music*. They also preferred *sad music* when they were feeling *sad*. Additionally, participants indicated a dislike of *happy* and *tender music* when they felt *angry, fearful*, or *disgusted*.

Follow-up ANOVAs were conducted to test for individual differences. Results revealed that in a neutral emotional state, participants who scored high on agreeableness tend to listen more to happy ($F(1, 19.27) =$16.12, $p <$.001) and tender ($F(1, 11.89) =$11.40, $p <$.005) music. When participants felt happy, the ones who scored high on openness tend to listen more to happy music ($F(1, 9.02) =$8.85, $p <$.05). Participants who scored high on neuroticism and felt disgusted tend to listen more to sad music ($F(1, 12.73) =$8.47, $p <$.005). Lastly when participants felt sad, and scored high on extraversion ($F(1, 16.95) =$9.96, $p <$.005), agreeableness ($F(1, 13.29) =$7.81, $p <$.05), or openness ($F(1, 16.29) =$9.57, $p <$.005), they tend to listen more to happy music.

5 Discussion

Our data show that the emotional state influences the (emotionally laden) music people listen to. In a neutral emotional state, happy and tender music is consumed more frequently. Additionally, findings indicate that people in general prefer emotionally laden music that is in line with their emotional state. Angry and fearful music is preferred when feeling angry or disgusted, whereas preference for happy and tender music decreases for these emotional states. Additionally, we found an increase of sad music in a sad state.

Taking personality traits into account, individual differences emerged. One of our findings showed that those who scored high on openness, extraversion, and agreeableness are more inclined to listen to happy music when they are feeling sad. In other words; they are trying to cheer themselves up with happy music. On the other hand, we found that those who are neurotic try to maintain their negative emotional state by listening to more sad songs.

In order to provide personalized music recommendations, we identified important individual differences that deviate from the notion that users desire to listen to music which is in line with their emotional state. By using personality to identify individual differences, we join the emergent interest of personality-based personalized systems. Several solutions have already been proposed to incorporate personality (e.g., [3, 4, 14]). For example, adaptation of the user interface of music recommender systems based on personality traits [4]. Also the extraction of emotion from social media is starting to establish (e.g., Twitter feeds [6]). Given our results there are several implications to consider. Music systems could anticipate the next song in the queue, or provide a list of recommendations, based on the user's current emotional state. This allows the system to better serve the user's music listening needs, and support their emotion regulation strategy.

Although we relied on self-report measures to assess emotions (emotion induction as well as music annotations) through an online platform, over 85% of the responses were in line with the original classifications that have been extensively tested priorly [1, 7]. This suggests that the used methods were effective.

Our results focused on individual differences of music preferences based on general emotional states. However, as Tamir and Ford [12] noted, emotion regulation strategies depend not only on individual differences, but also on the context that people are situated in. They found that people want to experience unpleasant emotions to attain certain instrumental benefits. That is, people want to feel bad when they expect it to give them benefits. For example, in confronting situations. We will address the influence of context in future work.

6 Acknowledgments

This research is supported by the Austrian Science Fund: P25655, and the EU FP7/2013-2016 through the PHENICX project under grant agreement 601166.

References

[1] Eerola, T., Vuoskoski, J.K.: A comparison of the discrete and dimensional models of emotion in music. Psychology of Music (2010)

[2] Ekman, P.: An argument for basic emotions. Cognition & Emotion (1992)

[3] Ferwerda, B., Schedl, M.: Enhancing music recommender systems with personality information and emotional states: A proposal. In: 2nd EMPIRE workshop (2014)

[4] Ferwerda, B., Yang, E., Schedl, M., Tkalcic, M.: Personality traits predict music taxonomy preferences. In: CHI'15 Extended Abstracts. ACM (2015)

[5] Hagemann, D., Naumann, E., Maier, S., Becker, G., Lürken, A., Bartussek, D.: The assessment of affective reactivity using films. Personality and Individual Differences 26(4), 627–639 (1999)

[6] Hasan, M., Rundensteiner, E., Agu, E.: Emotex: Detecting emotions in twitter messages. Academy of Science and Engineering (ASE) (2014)

[7] Hewig, J., Hagemann, D., Seifert, J., Gollwitzer, M., Naumann, E., Bartussek, D.: Brief report: A revised film set for the induction of basic emotions. Cognition & Emotion 19(7), 1095–1109 (2005)

[8] John, O.P., Srivastava, S.: The big five trait taxonomy: History, measurement, and theoretical perspectives. Handbook of personality: Theory and research (1999)

[9] Parkinson, B., Totterdell, P.: Classifying affect-regulation strategies. Cognition & Emotion 13(3), 277–303 (1999)

[10] Schaefer, A., Nils, F., Sanchez, X., Philippot, P.: Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers. Cognition and Emotion 24(7), 1153–1172 (2010)

[11] Tamir, M.: Don't worry, be happy? neuroticism, trait-consistent affect regulation, and performance. Journal of personality and social psychology 89(3), 449 (2005)

[12] Tamir, M., Ford, B.Q.: When feeling bad is expected to be good: emotion regulation and outcome expectancies in social conflicts. Emotion 12, 807 (2012)

[13] Thompson, W.F., Robitaille, B.: Can composers express emotions through music? Empirical Studies of the Arts 10(1), 79–89 (1992)

[14] Tkalcic, M., Ferwerda, B., Hauger, D., Schedl, M.: Personality correlates for digital concert program notes. UMAP 2015, Springer LNCS 9146 (2015)

[15] Tomarken, A.J., Davidson, R.J., Henriques, J.B.: Resting frontal brain asymmetry predicts affective responses to films. J. of personality and social psychology (1990)

[16] Zentner, M., Grandjean, D., Scherer, K.R.: Emotions evoked by the sound of music: characterization, classification, and measurement. Emotion 8(4), 494 (2008)

Personality Traits Predict Music Taxonomy Preferences

Published in: Extended Abstract Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2015; Seoul, KR)

BRUCE FERWERDA, Johannes Kepler University, Linz (AT)
EMILY YANG, Johannes Kepler University, Linz (AT)
MARKUS SCHEDL, Johannes Kepler University, Linz (AT)
MARKO TKALCIC, Johannes Kepler University, Linz (AT)

Abstract

Music streaming services increasingly incorporate additional music taxonomies (i.e., mood, activity, and genre) to provide users different ways to browse through music collections. However, these additional taxonomies can distract the user from reaching their music goal, and influence choice satisfaction. We conducted an online user study with an application called "Tune-A-Find," where we measured participants' music taxonomy choice (mood, activity, and genre). Among 297 participants, we found that the chosen taxonomy is related to personality traits. We found that openness to experience increased the choice for browsing music by mood, while conscientiousness increased the choice for browsing music by activity. In addition, those high in neuroticism were most likely to browse for music by activity or genre. Our findings can support music streaming services to further personalize user interfaces. By knowing the user's personality, the user interface can adapt to the user's preferred way of music browsing.

Keywords: Personality, Music, Recommender Systems, Taxonomy

1 Introduction

With large amounts of music available online, music information retrieval (MIR) plays an important role in navigating through music. Much research has been done on classifying music (assigning textual labels to characterize music) in order for systems to manage music collections and help users find music that they want to listen to (for an overview see [6]). Music streaming services use music classification to organize music in predefined taxonomies for their users to browse through. While "genre" is the most widely used taxonomy, other ways of organizing music have emerged. Popular music streaming services, such as: 8tracks, AccuRadio, Songza, Spotify, [1] incorporate additional taxonomies such as mood or activity to better serve different music browsing needs of its users. These additional taxonomies are in line with findings of how people use music in their lives [12].

While ample research has focused on choice satisfaction with an item, chosen from a set of similar items (e.g., [1]), others have shown that the creation of choice satisfaction already starts at the taxonomy level [10]. Research has demonstrated that additional taxonomies can distract, even when they are not considered relevant for the search of an item [15]. Distraction can raise the search effort because of competing attention [11], and complicate the search for an item [2]. Choosing and making trade-offs between taxonomies becomes challenging when they possess unique or complementary features that are not directly comparable [7]. In the end, it can decrease consumers' preference strength for, and satisfaction with the eventually picked item [10].

The genre, mood, and activity taxonomy used by music streaming services include features that are not directly comparable; they all provide different perspectives to browse for music. Displaying them simultaneously can complicate music browsing, and eventually diminish users' satisfaction with the chosen music and/or the system. Therefore, it is important to understand taxonomy preferences on an individual level. If the system can adapt the interface according to individual preferences by anticipating on taxonomies that may distract, it can contribute to increased user satisfaction. E.g., by emphasizing the user's preferred taxonomy and pushing others to the back, or not showing them at all.

Prior psychology studies have shown that personality is an enduring factor that influences user's behavior, interest, and taste (e.g., [16]; more details in the box below). This started an emerging interest in building personality-based recommender systems (e.g., [18]; more details in the box below). With this work, we contribute to this emerging field by showing that users' music browsing strategy is related to personality traits. For this we use the music browsing taxonomies (mood, activity, and genre) frequently used by music streaming services. Our results can be used to further improve personality-based systems. By knowing the relationship between personality and music taxonomies, we could exploit personality information in order to personalize the user interface and counteract on distracting taxonomies.

[1] http://www.8tracks.com, http://www.accuradio.com, http://www.songza.com, http://www.spotify.com

Personality & user preferences
Different studies have shown relationships between personality and user preferences. In the music domain Rentfrow and Gosling have investigated the relationship between personality traits and genre preferences. E.g., one of their findings show that extraversion is related to preferences for country, pop, religious, and soundtrack music [16]. In this study we take a broader perspective, and focus on the frequently used taxonomies by music streaming services to order their music collections. We show with our results that the relationship between personality and music is not limited to music preferences, but can be extended to music browsing strategies.

Personality-based recommender systems
Information about the relationship between personality traits and user preferences have been given interest to improve recommender systems. Tkalcic et al. for example, proposed a method to overcome the cold-start problem by including personality information to enhance the neighborhood measure [18]. Moreover, Hu and Pu found that personality-based recommender systems are more effective in increasing users' loyalty towards the system and decreasing cognitive effort compared to systems without personality information [9].

This leads us to the following research question:

RQ: *How do personality traits predict taxonomy (mood, activity, genre) preferences in music streaming services?*

We conducted an online user study where we asked participants to interact with a music application. Among 297 participants we found that personality traits are a predictor of taxonomy preferences. We found in line with our hypotheses, that those scoring *high* on *openness to experience* chose *mood*, those scoring *high* on *conscientiousness* or *neuroticism* chose *activity*, and those scoring *high* on *neuroticism* chose *genre*.

2 How taxonomy preferences can be predicted

Our thesis is that taxonomy preferences can be inferred from personality traits. Personality is considered an enduring factor that influences behavior, interest, and taste. Therefore, knowing one's personality can help to infer their preferences (e.g., music genre preferences [16]). Different models have been created to categorize personality, where the five-factor model (FFM) is one of the most well known and widely used. The FFM consists of five general dimensions that describe personality: openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism [13].

In this study, we hypothesize that these dimensions, except for agreeableness, can help us predict users' music taxonomy preference. The agreeableness dimension is related to helpfulness, trusting, and sympathetic [3]. These are factors that we believe are not relevant for predicting taxonomy preferences in this study, and therefore agreeableness will not be further discussed. The dimensions relevant to our hypotheses are discussed below.

2.1 Openness to experience

The openness to experience dimension refers to characteristics such as: active imagination, willingness to consider new ideas, divergent thinking, and intellectual curiosity. Those who score high on this scale tend to be unconventional and independent [3]. High openness to experience has also shown to be related to the openness of feelings. It has shown to relate to aesthetic emotions, as well as to greater awareness, clarity, and intensity of their own emotions at the time [17]. As those scoring high on openness to experience are more aware of, and more capable to judge their own emotions, we expect them to be more likely to chose for the mood taxonomy.

H1: *Those scoring high on openness to experience are more likely to choose for mood.*

2.2 Conscientiousness

Conscientiousness refers to characteristics such as self-discipline. People that score high on the conscientiousness scale tend to be more organized, plan oriented, and determined compared to those scoring low [3]. They also live lives that are overall less emotional, more balanced, more predictable, and will encounter fewer emotionally intense situations (fewer extreme lows, as well as fewer extreme highs). This dimension is considered to be the least emotionally charged, and is correlated with positive and negative emotions [8, 17]. Therefore, we expect this dimension to be more related to activity and genre, as these taxonomies bear more concreteness and are less emotionally charged.

H2: *Those scoring high on conscientiousness are more likely to choose for activity or genre.*

2.3 Extraversion

Extraverts are considered to be very sociable, energetic, optimistic, friendly, and assertive [3]. Findings show that for high extraverts, they might use physical and other energizing activities to distract themselves from negative emotions. However, this is a last resort and is only used when all other paths toward active modification are not available [4]. Although this does not seem to be a situation that would happen often, we think that extraversion may be a predictor for choosing activity.

H3: *Those scoring high on extraversion are more likely to choose for activity.*

2.4 Neuroticism

The neuroticism dimension indicates emotional stability and personal adjustment. High scoring on neuroticism are those that frequently experience emotional distress and wide swings in emotions, while those scoring low on neuroticism tend to be calm, well adjusted, and not prone to extreme emotional reactions [3]. Additionally, those who are highly neurotic do not believe that emotions are malleable, but rather difficult to control and strong in their expressions [8]. Previous research has found that music is used to regulate emotions [16]. As neurotic people do not consider emotions to be easily changed, we believe that they prefer to browse for music by activity or genre.

H4: *Those scoring high on neuroticism are more likely to choose for activity or genre.*

3 Method

To investigate personality and taxonomy preferences, we made use of a fake music application named “Tune-A-Find." This application was created to study user interaction and music preferences in a simulated music streaming service environment. The application consists of a simple interface with three taxonomies (mood, activity, and genre) for participants to browse for music (see Figure 1). A tooltip provided users a description of each taxonomy. We randomized taxonomy order to prevent order effects.

Prior to the start, we told participants that they were going to test a new music streaming service. Additionally, we instructed them to interact with

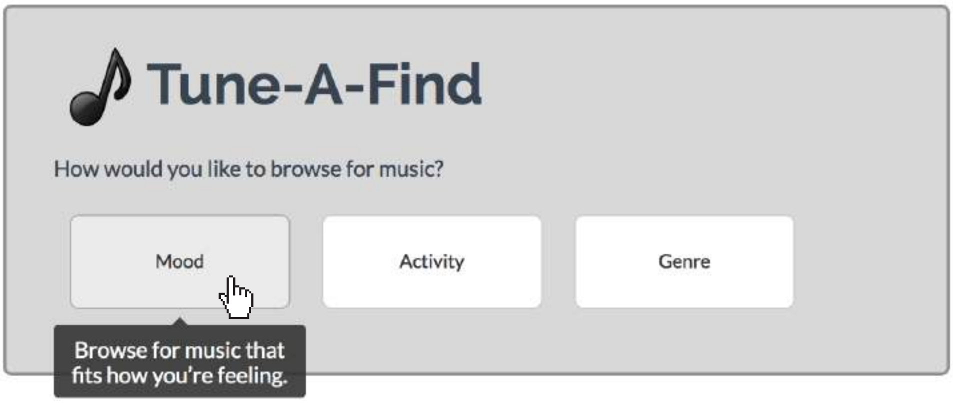

Figure 1: Screenshot of Tune-A-Find with “mood" tooltip. See the box below for all tooltip descriptions.

Tooltip descriptions
Mood: “Browse for music that fits how you’re feeling."
Activity: “Browse for music that fits what you’re doing."
Genre: “Browse for music by music style."

the system in the for them most ideal way in order to limit experience bias.

We collected participants’ first taxonomy choice to investigate their intrinsic taxonomy preferences. At the end of the study, we asked them to fill in the 44-item Big Five Inventory on a 5-point Likert scale (disagree strongly - agree strongly) in order to measure personality.

We recruited 326 participants through Amazon Mechanical Turk. Participation was restricted to those located in the United States, and also to those with very good reputation to avoid careless contributions. Participants were recruited at various times of the day to balance night and day time music application usage. Several comprehension-testing questions were used to filter out fake and careless entries. This left us with 297 completed and valid responses. Age (19 to 68, with a median of 31) and gender (159 males and 138 females) information indicated an adequate distribution.

4 Findings

Using a chi-square test of independence, we tested our hypotheses by looking at the relationship between participants’ five personality dimensions and the chosen music taxonomy (mood, activity, and genre). We discuss our re-

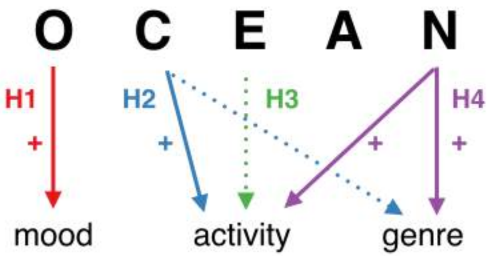

Figure 2: Hypotheses visualization. (O)penness to experience, (C)onscientiousness, (E)xtraversion, (A)greeableness, (N)euroticism. Solid lines represent accepted hypotheses, dotted lines the rejected ones.

sults in the following sections. The phi coefficient (ϕ) was used as the index of effect size. All findings indicate a moderate effect size. A recall of our hypotheses can be found in Figure 2.

4.1 Mood (*n*=68)

Chi-square test results indicated a marginally significant effect between openness to experience and mood χ^2(1, $N = 297$) = 3.117, $p = .07$, $\phi = .202$. This means that those who scored high on openness to experience were more likely to choose for mood than for activity or genre. All other personality dimensions were not significant.

4.2 Activity (*n*=16)

The chi-square test results for activity indicated a significant effect of neuroticism χ^2(1, $N = 297$) = 12.663, $p < .001$, $\phi = .306$. Additionally, we found a marginal significant effect of conscientiousness χ^2(1, $N = 297$) = 3.210, $p = .07$, $\phi = .204$. However, no significant effects were found for extraversion χ^2(1, $N = 297$) = .507, ns or the other dimensions. Results indicate that those who scored high on neuroticism or conscientiousness were more likely to choose activity.

4.3 Genre (*n*=213)

The chi-square test results for genre showed a significant effect of neuroticism χ^2(1, $N = 297$) = 6.583, $p = .01$, $\phi = .249$, but no significant effect was found of conscientiousness χ^2(1, $N = 297$) = 0, ns or the other dimensions.

This implies that those who scored high on neuroticism were more inclined to choose for genre.

5 Discussion

We found that those scoring high on openness to experience are likely to choose for mood (H1). Our hypothesis regarding high conscientiousness was partially supported (H2). High conscientiousness increased preference for activity, but not for genre. This could be explained by the fact that highly conscientious people are characterized as hard-working, task- and goal-oriented (taking to an extreme, they can be workaholics and perfectionists) [14]. Therefore, it can be that those scoring high on this dimension prefer a taxonomy where they can find music that fit their task or goal rather than looking for music based on genre. No significant results were found for our hypothesis between high extraversion and a preference for activity (H3). Extraverts are only likely to use physical and energizing activities as a last resort when all other paths toward modification of negative emotions are not available [4]. This could explain the rejection of our hypothesis, as these kind of situations are unlikely to occur frequently. Lastly, we found support for our hypothesis that those scoring high on neuroticism are more likely to choose for activity or genre (H4).

6 Conclusion

Our data suggest that personality traits are a predictor for music taxonomy preferences used in music streaming services. Given our results, there are several HCI related implications to consider. Personality information has already been proposed to improve recommendations in recommender systems (e.g., [5, 18]). Based on our results, personality information could be further incorporated to personalize user interfaces. E.g., preferred taxonomies could be displayed prominently, while other taxonomies are less emphasized, or music streaming services could recommend music based on the preferred taxonomy.

The current study focused on independent music taxonomies (mood, activity, and genre), but, for instance, our results show that high neuroticism is related to both activity and genre. People may hence not prefer one music taxonomy, but may be interested in combinations (e.g., sad pop music, funky road trip music, or happy cooking music). Furthermore, we will look at category preferences within a taxonomy. Such as, categories within mood (e.g., happy, sad, angry), activity (e.g., sleeping, relaxing, partying), and genre (e.g., pop, rock, jazz). Finally, cultural differences could play a role in taxonomy usage. Future work should address this.

7 Acknowledgements

This research is supported by the Austrian Science Fund (FWF): P25655; and by the EU FP7/2013-2016 through the PHENICX project under grant agreement no.601166.

References

[1] Bollen, D., Knijnenburg, B.P., Willemsen, M.C., Graus, M.: Understanding choice overload in recommender systems. In: RecSys. ACM (2010)

[2] Bravo, M.J., Farid, H.: Object recognition in dense clutter. Perception & psychophysics 68(6), 911–918 (2006)

[3] Burger, J.: Personality. PSY 235 Theories of Personality Series, Cengage Learning (2010), http://books.google.at/books?id=bZY7I2_8yRMC

[4] Carver, C.S., White, T.L.: Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: the bis/bas scales. J. of personality and social psych. (1994)

[5] Ferwerda, B., Schedl, M.: Enhancing music recommender systems with personality information and emotional states: A proposal. In: EMPIRE, Aalborg, Denmark. pp. 1613–0073 (2014)

[6] Fu, Z., Lu, G., Ting, K.M., Zhang, D.: A survey of audio-based music classification and annotation. Multimedia, IEEE Transactions on pp. 303–319 (2011)

[7] Gourville, J.T., Soman, D.: Overchoice and assortment type: When and why variety backfires. Marketing science 24(3), 382–395 (2005)

[8] Gross, J.J., John, O.P.: Mapping the domain of expressivity: multimethod evidence for a hierarchical model. J. of personality and social psych. (1998)

[9] Hu, R., Pu, P.: Acceptance issues of personality-based recommender systems. In: RecSys. pp. 221–224. ACM (2009)

[10] Iyengar, S.S., Wells, R.E., Schwartz, B.: Doing better but feeling worse looking for the âĂIJbestâĂİ job undermines satisfaction. Psychological Science 17(2), 143–150 (2006)

[11] Janiszewski, C.: The influence of display characteristics on visual exploratory search behavior. Journal of Consumer Research (3), 290–301 (1998)

[12] Lonsdale, A.J., North, A.C.: Why do we listen to music? a uses and gratifications analysis. British Journal of Psychology 102(1), 108–134 (2011)

[13] McCrae, R.R., John, O.P.: An introduction to the five-factor model and its applications. Journal of personality 60(2), 175–215 (1992)

[14] Ozer, D.J., Benet-Martinez, V.: Personality and the prediction of consequential outcomes. Annu. Rev. Psychol. 57, 401–421 (2006)

[15] Perruchet, P., Rey, A., Hivert, E., Paction, S.: Do distractors interfere with memory for study pairs in associative recognition? Memory & cognition 34(5), 1046–1054 (2006)

[16] Rentfrow, P.J., Gosling, S.D.: The do re mi's of everyday life: the structure and personality correlates of music preferences. J. of personality and social psych. (2003)

[17] Shiota, M.N., Keltner, D., John, O.P.: Positive emotion dispositions differentially associated with big five personality and attachment style. The Journal of Positive Psychology 1(2), 61–71 (2006)

[18] Tkalcic, M., Kunaver, M., Košir, A., Tasic, J.: Addressing the new user problem with a personality based user similarity measure. In: DEMRA (2011)

Personality and Taxonomy Preferences, and the Influence of Category Choice on the User Experience for Music Streaming Services

Published in: Submitted to Journal

BRUCE FERWERDA, Johannes Kepler University, Linz (AT)
EMILY YANG, Johannes Kepler University, Linz (AT)
MARKUS SCHEDL, Johannes Kepler University, Linz (AT)
MARKO TKALCIC, Johannes Kepler University, Linz (AT)

Abstract

Music streaming services increasingly incorporate different ways for users to browse for music. Next to the commonly used “genre" taxonomy, nowadays additionally taxonomies, such as mood and activities are often used. As additional taxonomies have shown to be able to distract the user in their search, we looked at how to predict taxonomy preferences in order to counteract this. Additionally, we looked at how the number of categories presented within a taxonomy influences the user experience. We conducted an online user study where participants interacted with an application called “Tune-A-Find." We measured taxonomy choice (i.e., mood, activity, or genre), individual differences (e.g., personality traits and music expertise factors), and different user experience factors (i.e., choice difficulty and satisfaction, perceived system usefulness and quality) when presenting either 6- or 24-categories within the picked taxonomy. Among 297 participants, we found that personality traits are related to music taxonomy preferences. Furthermore, our findings show that the number of categories within a taxonomy influences the user experience in different ways and is moderated by music expertise. Our findings can support personalized user interfaces in music streaming services. By knowing the user's personality and expertise, the user interface can adapt to the user's preferred way of music browsing.

Keywords: Personality, Taxonomy, Music, Categorization, Overchoice

1 Introduction

With the rapid increase of technological advances, Internet connectivity became faster and cheaper. Music streaming services are eagerly making use of these developments by putting large music collections online. The availability of online music gives users the possibility to choose from an immense music collection and stream the content directly to their devices at any given time. While it provides users an almost unlimited choice of music, the abundance of music is a cognitive burden. To overcome this, music streaming services try to organize their collections in such a way so that users can easily browse and find what they would like to hear. Different classification methods from music information retrieval (MIR) are used to organize music (see for an overview [4, 31]) in order for users to browse through these large amount of music content.

The “genre" taxonomy has been most commonly used to organize music. However, lately, music streaming services have shown interest in a more user centered approach for organizing music. Popular music streaming services, such as: 8tracks (http://www.8tracks.com), AccuRadio (http://www.accuradio.com), Songza (http://www.songza.com), Spotify (http://www.spotify.com), have started to provide additional taxonomies to better serve users with diverse music browsing needs. Derived from research that investigated how people use music in their everyday life [57], taxonomies such as mood and activity are included for users to find music they want to listen to in different situations or contexts.

By providing different taxonomies, diverse browsing needs of users are served. Prior research has shown that satisfaction with the eventually chosen item already starts at the taxonomy level [37]. However, taxonomies can compete, especially when they possess unique or complementary features that are not directly comparable [34]. Taxonomies that are not relevant for the search goal can distract the user [65]. They can complicate the search process [9], and increase the search effort because of competing attention [44]. This can result in that the taxonomy influences users' preference strength and satisfaction with the eventually chosen item [43]. Therefore, understanding taxonomy preferences on an individual level is important to provide a personalized music experience. For example, music streaming services could emphasize the taxonomy that is important to the user while muting less preferred taxonomies to the background, or not showing them at all.

To predict taxonomy (mood, activity, or genre) preferences, personality traits can play an important role. Personality has shown to be a reliable predictor for human preferences. It has shown to be an enduring factor that influences people's behavior [47], interests, and taste [27, 50, 69]. As personality has shown to be a prominent factor in shaping human preferences (e.g., behavior, interest, and taste), one can expect similar patterns to emerge between similar personality traits [10]. For that, we believe that a preference for a taxonomy (i.e., mood, activity, or genre) can also be derived by the personality characteristics of a user.

In addition to taxonomy preferences, we take a further look into the subsequent step of music browsing; the categories within a taxonomy. We look at how the amount of categories within a music taxonomy influences the user experience (i.e., choice satisfaction, choice difficulty, perceived system quality, and perceived system usefulness). Ample research has shown that presenting more options may not always be a good thing. It can cause overchoice (also referred as “choice overload"), that in turn influences the difficulty to make a choice and satisfaction with the item and decreases choice satisfaction [6, 36, 41, 71, 75, 77]. However, whether overchoice occurs has shown to be influenced by different moderators (e.g., expertise, choice set attractiveness). See for an overview [74]). To our knowledge, overchoice and its moderators in the context of music browsing have not been studied so far. As we will further discuss in the next section, expertise plays an important role in whether overchoice occurs. In this study, we rely on measurements to infer music expertise related to the chosen music taxonomy.

Our work makes several contributions. Firstly, we contribute to the field of personality-based preferences. We show that personality does not only explain music genre preferences [69], but that it extends to the overarching music categorizations (i.e., taxonomies) by showing that personality traits are related to different music browsing strategies (i.e., browsing for music by mood, activity, or genre). Secondly, we contribute to the decision making literature by extending the knowledge about when and how overchoice occurs in the context of music. For this we look at the categories within a taxonomy, and show that music expertise is an important influencing factor on the evaluation of the system and chosen item.

Based on our work, music services could be further personalized. The service could adapt the user interface depending on the user's personality and level of music expertise. This allows for counteracting on decreased user experience by the user interface.

The research questions (RQs) that we try to answer in this work are:

RQ 1: *How do personality traits relate to taxonomy (mood, activity, genre) preferences in music streaming services?*

RQ 2: *How does the size of the choice set influence the user experience (i.e., choice satisfaction, perceived system usefulness, and perceived system quality), and how is this moderated by expertise?*

To investigate these research questions, we conducted a user study (preceded by a preliminary study) in which we simulated a music streaming service application. Among 297 participants we found that personality is related to different music browsing strategies. We found that openness to experience is positively related to the mood taxonomy, conscientiousness and neuroticism are positively related to the activity taxonomy, and neuroticism is positively related to the genre taxonomy.

We also looked at the size of the choice set (i.e., 6 or 24 items) within a chosen music taxonomy in order to investigate the overchoice effect. We found that music expertise plays a moderating role in how the system is evaluated by the user. Within the mood taxonomy, participants with higher music expertise rated the system as more useful and of higher quality when they were facing the choice set with less options to choose from. However, this was the other way around for the genre taxonomy. Here, participants with a higher music expertise rated the system more useful and of a higher quality when facing a bigger choice set.

We investigated two different RQs within one study. Therefore, the remainder of the paper is structured as follows. We first discuss the related work separately for each RQ in Section 2. In Section 2.1 we discuss the related work of RQ 1, and in Section 2.2 the related work of RQ 2. After the related work, we continue with the materials (Section 3) that were used for the user study to answer RQ 1 and 2. In Section 4 we discuss the preliminary study that was necessary to define the content for the user study. Subsequently, we divided Section 5 into Study A and B, where we will treat the hypotheses, findings, and discussion related to RQ 1 and 2. We discuss the limitations and future work in Section 6. Finally we round off the paper by drawing conclusions in Section 7.

2 Related Work

We divided our user study into two parts in order to investigate both RQs, which we call from hereon Study A and Study B. Therefore, we separated the related work into two parts as well. The first part will discuss work that is related to the taxonomies and personality traits (Study A. Section 2.1), and the second part focuses on the overchoice in presenting different number of categories within a taxonomy (Study B. Section 2.2).

2.1 Study A - Taxonomies

In the following sections we discuss how taxonomies influences users' decision making, and how personality is able to predict the preference for a taxonomy.

2.1.1 Taxonomic influence

The effects of overchoice have been well studied. However, most research on overchoice in consumer decision making has investigated choice satisfaction by focusing on choices in isolation (i.e., choices within a taxonomy; e.g., [6, 36, 41, 71, 75, 77]). For example, Iyengar and Lepper [41] investigated overchoice by using an assortment of on a specific set of jams. Bollen et al [6] created movie recommendations by using only the Top-5 and Top-20 movies. Although they found effects of overchoice on choices in isolation, others have shown that the satisfaction with the eventually chosen item already starts at the overarching categorizations; the taxonomies (e.g., [34, 37, 65]). Herpen et al [37] asked their participants to choose a shirt from clothing brochures. They found that taxonomies can distract in the decision making process. Their participants experienced higher decision effort, had more difficulties grasping the selection, and were experiencing more confusion when the brochures consist of complementary clothing taxonomies (e.g., shirts, pants, shoes) than when they consist of substitutes (i.e., only shirts). The complementary taxonomies caused participants to take longer to make a decision. In extension, Perruchet et al [65] found that taxonomies can unintentionally act as distractors when not relevant for the initial search goal or when not being actively looked for at all. When taxonomies are placed next to each other, they start to compete. Gourville and Soman [34] showed that competing of the taxonomies is exacerbated especially when they consist of features that are unique and not directly comparable.

Although different taxonomies in music streaming services serve the same goal of providing users with music that they would want to listen to, they also consist of unique features that are not directly comparable. The taxonomies provide users the possibility to browse for music in different ways. In general, the mood taxonomy provides users with music that is similar to how they feel, the activity taxonomy provides music that fits a specific activity, and the genre taxonomy has music categorized based on a set of stylistic criteria. Given that the features of the taxonomies are not directly comparable, they can distract the user and increase the effort of picking the right music taxonomy to continue the music browsing. In the end, it can influence the satisfaction with the eventually chosen music item.

To minimize the negative influence of competing taxonomies, we try to counteract by identifying the intrinsic music browsing preference of the user. By identifying the user's most preferable music browsing strategy, the system can anticipate the desired user interface. For example, the system can display the preferred music browsing taxonomy or already recommend music that is in line with a user's music browsing strategy (e.g., [26]).

In order to identify the music browsing preference of users, we rely on personality traits. We will discuss prior work related to personality in the next section.

2.1.2 Personality

Personality has shown to be an enduring factor that influences an individual's behavior [46], interest, and tastes [50, 69]. As personality plays such a prominent role in shaping human preferences, one can expect similar patterns (i.e., behavior, interest, and tastes) to emerge between similar personality traits [10]. Different models have been created to categorize personality, where the five-factor model (FFM) is the most well known and widely used [59]. The FFM consists of five general dimensions that describe personality. Each of the five dimensions consist of clusters of correlated primary factors. Table 1 shows the general dimensions with the corresponding primary factors.

General dimensions	**Primary factors**
Openness to experience	Artistic, curious, imaginative, insightful, original, wide interest
Conscientiousness	Efficient, organized, planful, reliable, responsible, thorough
Extraversion	Active, assertive, energetic, enthusiastic, outgoing, talkative
Agreeableness	Appreciative, forgiving, generous, kind, sympathetic, trusting
Neuroticism	Anxious, self-pitying, tense, touchy, unstable, worrying

Table 1: The five-factor model adapted from McCrae and John [59].

There is a growing amount of psychological literature investigating the relationship between personality traits and music consumption (e.g., [27, 67, 69, 70, 82]). For example, music preferences were found to be correlated with personality traits. Rentfrow and Gosling [69] categorized music pieces into 4 music-preference dimensions (reflective and complex, intense and rebellious, upbeat and conventional, and energetic and rhythmic), and found correlations with the five general personality dimensions (i.e., openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism), such as, a relationship between energetic and rhythmic music and extraversion and agreeableness. The psychological work on personality provides valuable information for the development of domain specific recommender systems. Knowing someone's personality can help infer their preferences, and can therefore contribute to a better personalized experience.

There has been an emergent interest in how to use personality in a system (e.g., recommender systems), and several directions have been proposed (e.g., [23, 26, 25, 80, 81]). For example, Tkalcic et al [81] propose a method to overcome the “cold-start problem" [1] by including personality information to enhance the neighborhood measurement. Hu and Pu [39] have shown that personality-based recommender systems are more effective in increasing users’ loyalty towards the system and decreasing cognitive effort compared to systems that do not use personality information.

Also research on personality acquisition is progressing. Research has been looking into ways to implicitly acquire personality traits of users and thereby create opportunities to decrease the usage of extensive and time consuming questionnaires. Possibilities to implicitly acquire personality traits become possible with the increased implementation of mechanisms such as “single sign-on buttons" by applications. Single sign-on buttons allow users to conveniently register as a new user and login to an application by using their social media account (e.g., Facebook). However, except for the user’s basic information that is required for the application, applications often gain access to an abundance of other data of the user [14]. Research on new media has shown that it is able to infer personality traits from content of social media trails (e.g., Facebook [2, 29, 33, 64, 72], Twitter [32, 66], and Instagram [28, 30]).

2.2 Study B - Categories

In this work, we further look into the influence of the number of categories presented within each taxonomy (Section 5.3). With this we join decision-making research on overchoice. Overchoice (or choice overload) refers to the increase of choice difficulty and eventual decrease of satisfaction as the number of choices increase. Iyengar and Lepper [41] were one of the first to define overchoice by testing the attractiveness between a set of 6 or 24 types of jam. Although their result shows that initially participants were more attracted to the larger (24 item) set, those who were exposed to the smaller (6 item) set were more inclined to actually buy a pot of jam (3% and 30% respectively of the participants bought jam). Additionally, assessment of satisfaction showed that those who bought jam from the larger choice set were less satisfied compared to those with a purchase from the smaller choice set.

[1] The cold-start problem is most prevalent in recommender systems, and occurs when there is not enough data (yet) to recommend personalized items to the user. This problem especially occurs for new users.

The overchoice effect has been replicated numerous times in different context, and was shown to affect motivation to choose as well as satisfaction with the chosen item (e.g., [6, 36, 42, 71, 77]). Shah and Wolford [77] found a motivational buying decrease in purchasing black pens when the assortment size increases. When they increased the assortment size of the black pens, participants' motivation to purchase decreased from 70% to 33%. Reutskaja and Hogarth [71] investigated overchoice in the context of gift-boxes prices, and found that satisfaction with the chosen gift-box decreased when the number of gift-boxes to choose from increased. Similar findings were shown by Haynes [36] in the context of the number of lottery prices to choose from. Likewise, Bollen et al [6] demonstrated a decrease in choice satisfaction in a movie recommender system among an increase number of attractive movies to choose from.

Although there is an increased chance of a decrease in choice satisfaction, people somehow still cherish more choice, and studies have shown that shops with a large variety even create a competitive advantage by providing more choices (e.g., [1, 8, 12, 16, 40, 48, 58, 54, 62]). So it seems that even though consumers risk to be more dissatisfied with their choice at the end, they still are attracted to more choices.

Why a larger set is more attractive and yet less satisfactory, can be explained by a number of mechanisms. A larger choice set becomes more attractive because of the summed benefit of each option, and thereby the total benefit of the set increases [20]. However, satisfaction decreases because making the right choice becomes more difficult. The psychological cost increases as a consequence of an increased number of choices. In other words, the summed benefit of a larger choice set is outweighed by the cost of comparing each option in order to make the right decision, increased risk of making a wrong choice, and increased expectations with the chosen item [20, 75, 76]. This results in that a larger choice set has a higher chance of decreased satisfaction or that no choice is made at all. Reutskaja and Hogarth [71] showed that overchoice occurs in an inverted satisfaction U-curve, where at one point the total cost of the choice set grows faster than the total benefit of the set, causing a decrease of satisfaction to occur.

Apart from studies that have shown the overchoice effect, there are also studies that demonstrate an opposing view (e.g., [5, 7, 21, 49, 79]). They found that reducing the variety in retail shops often result in decrease sales or no change at all. Scheibehenne et al [74] performed a meta-analysis of 50 studies voting against and in favor of the overchoice hypothesis, and found that the overall effect size comes close to zero. There seem to be necessary preconditions for a choice set before overchoice occurs [73, 74]. As mentioned before, the attractiveness of the choice set plays an important role. When items of a choice set are comparably attractive, and especially when they additionally consist of incomparable features, the chances of overchoice increases ([17, 24]).

Besides that the choice set is undermined with preconditions; also the decision maker's characteristics play a role. A factor that has been shown to play a significant role is expertise with the choice domain [60, 74]. When people have expertise with the choice content, they are less prone to be overwhelmed by the increasing number of choices, and therefore, overchoice is less likely to occur.

In order to investigate the overchoice within a music taxonomy, we first needed to create a choice set that meets the precondition (i.e., a choice set with attractive items). We conducted a preliminary study where we identified the categories that would be most attractive to our participants (see Section 4). Additionally, the moderator effect of music expertise related to each music taxonomy is further investigated in Section 5.3.

3 Method

In order to investigate music taxonomy preferences for music browsing, and overchoice of categories within a music taxonomy, we created an online experiment where we simulated a music streaming service application. This application allowed us to study both RQ 1 and 2 at the same time. In the following sections we will discuss the experiment in more detail, and the materials used.

3.1 Procedure

For the experiment we simulated a music streaming service application named "Tune-A-Find" (see Figure 1 for the work-flow of the experiment) in order to investigate taxonomy preferences, and overchoice of categories within a taxonomy. Before participants started the experiment, instructions were given stating that they were about to test a new music streaming service. We emphasized that it is important that they interact with the application in the most ideal way for them. This allowed us to minimize experience bias with any of the taxonomies. After participants agreed with the instructions they continued by interacting with Tune-A-Find.

The simulated music streaming service consists of a simple interface with three taxonomies (mood, activity, and genre) for participants to browse for music (see Figure 2 and Section 5.2). A tooltip provided users a descrip-

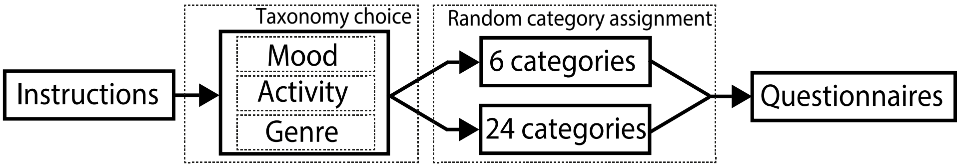

Figure 1: Experiment work-flow. Participants were given instructions about the study, then continued by interacting with the music streaming application (see Section 5.2 for details). After choosing a music taxonomy to continue the music browsing, participants were randomly assigned to either the 6-categories condition or the 24-categories (see Section 5.3 for more details). After picking a category, participants continued to the concluding questionnaires.

tion of each taxonomy. [2] The order of the taxonomies was randomized to prevent order effects. After participants chose a taxonomy to search music by (i.e., mood, activity, or genre), they continued on by choosing a category (i.e., type of mood, type of activity, or type of genre) within the chosen taxonomy.

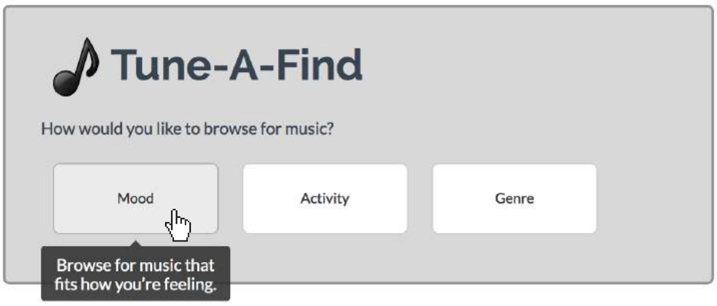

Figure 2: Screen shot of Tune-A-Find with the “Mood" tooltip.

For the categories within a chosen taxonomy, participants were randomly assigned to either the 6- or 24-categories condition (Figure 3 and Section 5.3). The categories within each taxonomy were based on the results of the preliminary study (Section 4). We did not allow participants to go back to pick a different taxonomy. Therefore, we included a “None of the items" option. Category order was randomized with “None of the items" option always placed last. After participants picked a category, they continued with the concluding questionnaires (i.e., user experience, music expertise, personality, and demographics questionnaires). We intentionally did not include real music recommendations in order to avoid response biases in the

[2] Mood tooltip description “Browse for music that fits how you're feeling." Activity tooltip description “Browse for music that fits what you're doing." Genre tooltip description “Browse for music by music style."

questionnaires due to the influence of the algorithm.

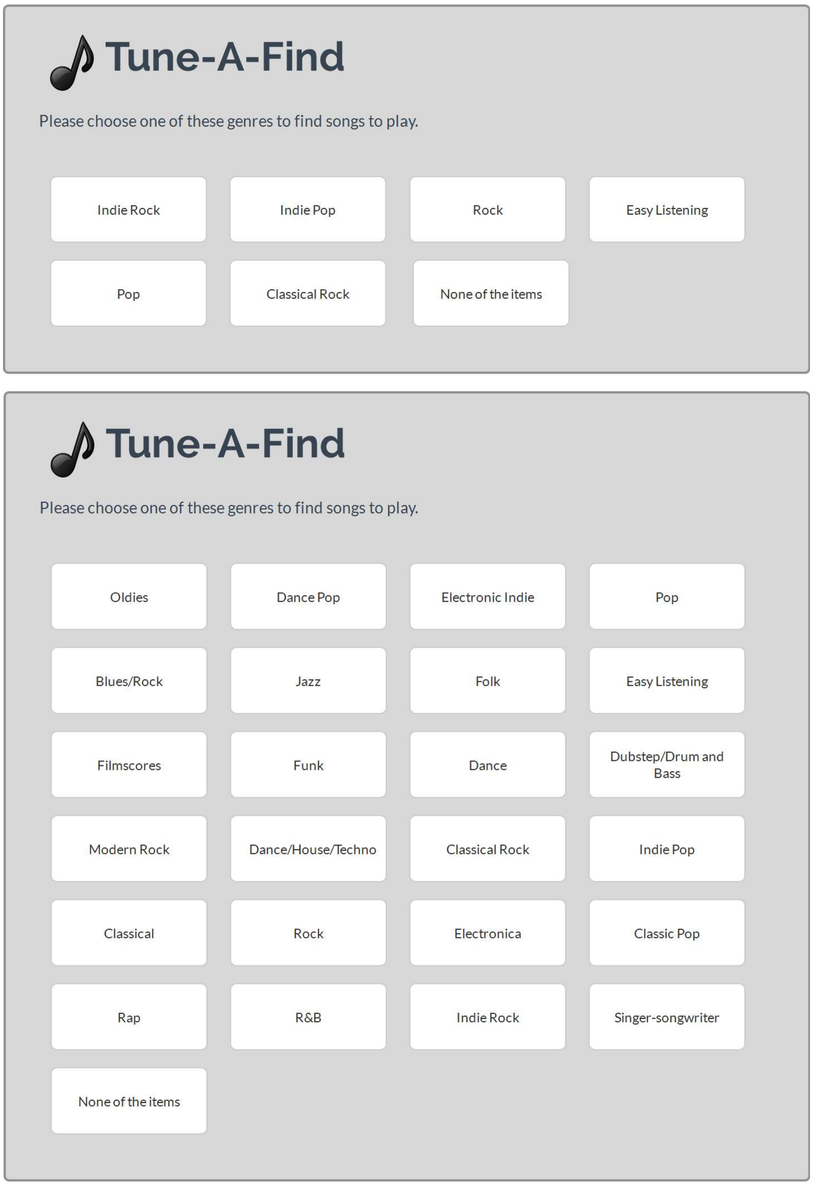

Figure 3: Screen shots of the 6- and 24-categories conditions (top and bottom, respectively) with an extra option of “None of the items."

3.2 Materials

The taxonomies used in Tune-A-Find (i.e., mood, activity, and genre), are based on a close observation of current music streaming services. We found that these labels are increasingly being used (see Table 2).

	Mood	Activity	Genre
8Tracks	X	X	X
AccuRadio		X	X
Earbits			X
Grooveshark			X
Google Play Music			X
Guvera		X	X
Jango		X	X
Last.fm			X
Musicovery	X	X	X
Pandora			X
Slacker		X	X
Songza	X	X	X
Spotify	X	X	X

Table 2: Grasp of the observed music streaming services and the taxonomies they use to organize music.

For the number of categories to present within each taxonomy, we followed the original work of Iyengar and Lepper [41] on overchoice. They observed the occurrence of overchoice between choice sets consisting of 6 and 24-items. We conducted a separate user study to determine which categorical labels (types of mood, types of activity, or types of genres) to include within each taxonomy (see Section 4).

For the concluding questionnaires we made use of existing questionnaires measuring: user experience, music expertise, and personality. To measure user experience factors we adapted the original user experience questionnaire of Knijnenburg et al [52] to fit the music streaming context of our study. The questionnaire depicts different parts of the user experience. It measures participants choice difficulty, choice satisfaction, perceived system usefulness, and perceived system quality (see Appendix B).

In order to measure participants' music expertise, we relied on the Goldsmiths Music Sophistication Index (Gold-MSI [61]). The Gold-MSI questionnaire measures music sophistication based on the following dimensions:

- Active engagement (how much time and money one spends on music)
- Perceptual abilities (cognitive musical ability related to music listening skills)
- Musical training (musical training and practice)
- Signing abilities (skills and activities related to singing)

- Emotions (active behaviors related to emotional responses to music)

In the remainder of this paper, we will talk about “dimension expertise" to refer to the separate dimensions of the Gold-MSI. For this study we adopted parts of Gold-MSI that are related to the taxonomies (i.e., active engagement, perceptual abilities, and emotions. See also Appendix C for the used questions).

In order to measure personality, we relied on the widely used, 44-item Big Five Inventory (5-point Likert scale; disagree strongly - agree strongly. [45]). Finally, standard demographic questions were asked (i.e., age and gender).

4 Preliminary study

In order to determine which categories to use in each taxonomy, we conducted a preliminary study. Prior research has shown that before over-choice occurs, the items in the choice set are subject to preconditions. For example, when the differences between the attractiveness of the items is small, and especially when they consist of incomparable features [17, 24]. We conducted this preliminary study in order to identify the most attractive categories for users in each music taxonomy. In the following sections we outline the method and findings.

4.1 Method

For this preliminary study we recruited 45 participants through Amazon Mechanical Turk, a popular recruitment tool for user-experiments [51]. Only those located in the United States, and with a very good reputation were allowed to participate ($\geq$95% HIT approval rate and $\geq$1000 HITs approved) [3]. We compensated participants with $1 for their participation.

We extracted the categories provided by Songza (see Appendix A for the complete list), as they have a clear separation of categories between taxonomies whereas others (e.g., Spotify) have a mixed taxonomy view. For each taxonomy we asked participants to pick 12 categories that they would most likely use when browsing for music.

[3] HITs (Human Intelligence Tasks) represent the assignments a user has participated in on Amazon Mechanical Turk prior to this study.

4.2 Findings & conclusion

In line with prior work of Iyengar and Lepper [41] on overchoice, and work defining the preconditions of the choice set [17, 24], we picked the top 6 and 24 most attractive (i.e., the categories that participants indicated to use most likely in their music browsing) categories (Table 3). The top 6 and 24 most attractive categories are used for Study B (see Section 5.3) where we investigate overchoice within a music taxonomy (i.e., mood, activity, and genre).

	Mood	**#**	**Activity**	**#**	**Genre**	**#**
1	Energetic	40	Relaxing	30	Pop	29
2	Happy	37	Being Creative	26	Rock	23
3	Soothing	35	Rainy day	24	Rock: Classic Alternative & Punk	21
4	Mellow	34	Staying Up All Night	22	Indie: Indie Rock	20
5	Atmospheric	31	Road Trip	21	Indie: Indie Pop	19
6	Hypnotic	30	Working/Studying (without lyrics)	19	Easy Listening	19
7	Introspective	28	Reading in a Coffee Shop	18	Classical	18
8	Warm	27	Singing in the Shower	18	Blues & Blues Rock	18
9	Motivational	27	Housework	18	Film scores	18
10	Funky	25	Working/Studying (with lyrics)	17	Folk	17
11	Sad	24	Romantic Evening	17	Dance	16
12	Celebratory	24	Gaming	17	R&B	16
13	Nocturnal	23	Energy Boost	17	Pop: Classic Pop	15
14	Aggressive	22	Working Out: Weight Training	17	Rap	15
15	Seductive	22	Unwinding After Work	16	Oldies	14
16	Gloomy	22	Working Out: Cardio	16	Electronica	14
17	Sweet	22	Dance Party: Beach	16	Jazz	14
18	Classy	20	House Party	16	Rock: Modern Rock	13

19	Sexual	19	Barbecuing	15	Indie: Indie Electronic	13
20	Raw	19	Lying Low on the Weekend	15	Dance: House & Techno	13
21	Angsty	18	Sleeping	14	Singer-Songwriter	13
22	Visceral	18	City Cruising	14	Pop: Dance pop	12
23	Spacey	16	Waking Up on the Right Side of the Bed	12	Funk	12
24	Trippy	16	Lying on a Beach	12	Dubstep & Drum 'n Bass	12

Table 3: Top 6- and 24-categories chosen by participants. # represents the number of votes

5 Main studies

In the following subsections, we discuss the main studies where we treat the hypotheses, findings, and discussion for each study separately. Study A depicts the taxonomy preferences (Section 5.2), and Study B addresses the overchoice effect within a chosen taxonomy (Section 5.3). [4]

5.1 Participants

We recruited 326 participants through Amazon Mechanical Turk. Participation was restricted to those located in the United States, and also to those with very good reputation ($\geq$95% HIT approval rate and $\geq$1000 HITs approved) to avoid careless contributions. Participants were recruited at various times of the day to balance night and day time music application usage. Several comprehension-testing questions were used to filter out fake and careless entries. This left us with 297 completed and valid responses. Age (19 to 68, with a median of 31) and gender (159 males and 138 females) information indicated an adequate distribution. Participants were compensated with $2 for their participation.

[4] The studies extend the preliminary results published in [removed for submission] .

5.2 Study A

In Study A, we looked at how taxonomy preferences are related to different personality traits. To investigate this relation we simulated a music streaming service (Figure 2). The application consists of a simple interface with three taxonomies (mood, activity, and genre) for participants to browse for music. A tooltip provided users a description of each taxonomy. The order of the taxonomies were randomized to prevent order effects. Once a taxonomy was picked, participants continued by choosing a category within the chosen taxonomy (this is addressed in Study B in Section 5.3). As we are interested in users' intrinsic taxonomy preferences, participants were not able to go back once a taxonomy was picked. For those who want to choose a different taxonomy, we included an additional option of "None of the items" among the available categories. For those who picked this option, we included an additional question in the concluding questionnaire where they could indicate what they would have picked otherwise in terms of taxonomy (i.e., mood, activity, or genre) as well as the category within a taxonomy.

In order to prevent an experience bias with one of the music taxonomies (i.e., mood, activity, or genre), participants were told during the instructions of the user study that they were going to test a new music streaming service, and therefore it is important that they interact with the system in the most ideal way for them.

As there is no strong evidence from the literature to form hypotheses, we decided to adopt an exploratory approach. We try to draw relationships between our findings to what is known from prior research in the discussion section.

5.2.1 Findings

Using a chi-square test of independence, we explored the relationship between participants' five personality dimensions and the chosen music taxonomy (mood, activity, and genre). We used a median split to divide each personality trait into a low and high measure. The distribution of the music taxonomy choices made by the participants are shown in Table 5. Participants in general chose for genre taxonomy followed by the mood and activity taxonomies. In the following sections we discuss the relationship between personality traits and the music taxonomy chosen by the participants (see Table 4 for an overview).

	χ^2 (*p*)				
	O	**C**	**E**	**A**	**N**
Mood	**3.117**	0.934	0.870	0.044	0.703
	(0.05)	(0.334)	(0.351)	(0.833)	(0.402)
Activity	0.046	**3.210**	0.507	0.406	**12.663**
	(0.830)	**(0.05)**	(0.477)	(0.524)	**(<0.001)**
Genre	3.079	0.000	1.506	0.266	**6.583**
	(0.11)	(0.997)	(0.220)	(0.606)	**(0.01)**

Table 4: Summary of the results for each taxonomy (i.e., mood, activity, and genre) with each personality trait: (O)pennes to experience, (C)onscientiousness, (E)xtraversion, (A)greeableness, and (N)euroticism.

Mood Taxonomy. Results of the chi-square test indicated a positive relationship between openness to experience and mood χ^2(1, $N = 297$) = 3.117, $p = .05$. This means that those who scored high on the openness to experience dimension were more likely to choose for mood than for activity or genre taxonomy. We did not find any significant effects of the other personality traits: conscientiousness χ^2(1, $N = 297$) = .934, $p = .334$, extraversion χ^2(1, $N = 297$) = .870, $p = .351$, agreeableness χ^2(1, $N = 297$) = .044, $p = .833$, and neuroticism χ^2(1, $N = 297$) = .703, $p = .402$.

Activity Taxonomy. When looking at the chi-square test results for the activity taxonomy, we found a positive significant effect of conscientiousness χ^2(1, $N = 297$) = 3.210, $p = .05$. Additionally, we found a positive relationship of neuroticism χ^2(1, $N = 297$) = 12.663, $p < .001$. These results indicate that those who scored high on neuroticism or conscientiousness were more likely to choose the activity taxonomy. We did not find significant effects for openness to experience χ^2(1, $N = 297$) = .046, $p = .830$, extraversion χ^2(1, $N = 297$) = .507, $p = .477$, and agreeableness χ^2(1, $N = 297$) = .406, $p = .524$.

Genre Taxonomy. The chi-square test results for the genre taxonomy indicated a positive significant effect of neuroticism χ^2(1, $N = 297$) = 6.583, $p = .01$, which implies that those who scored high on neuroticism were more inclined to choose for genre than for the other taxonomies.. All the other personality traits were not significant: openness to experience χ^2(1, $N = 297$) = 3.079, $p = .11$, conscientiousness χ^2(1, $N = 297$) = 0, $p = .997$, extraversion χ^2(1, $N = 297$) = 1.506, $p = .220$, and agreeableness χ^2(1, $N = 297$) = .266, $p = .606$.

Additional finding. Additionally, we looked for effects of gender and age. Controlling for gender and age did not result in any significant effects. However, as seen in Table 5, the distribution of gender is interesting and indicates some trends. The distribution of women is higher in mood and activity, while conversely for genre.

Category	#Male (percentage)	#Female (percentage)	#Total
Mood	26 (38%)	42 (62%)	68
Activity	5 (31%)	11 (69%)	16
Genre	128 (60%)	85 (40%)	213

Table 5: Distribution of men and women across the music taxonomy preference.

5.2.2 Discussion

In this study, we investigated whether music taxonomy (mood, activity, and genre) preferences can be inferred from personality traits. We found that there is a relationship between personality traits taxonomy preferences that are used by music streaming services. We visualized our findings in Figure 4.

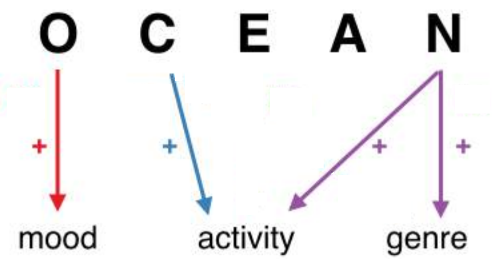

Figure 4: Visualization of our findings: (O)penness to experience, (C)onscientiousness, (E)xtraversion, (A)greeableness, (N)euroticism.

We found a positive relationship between openness to experience and the mood taxonomy. This indicated that those scoring high on openness to experience are likely to choose for music organized by mood. The openness to experience dimension refers to characteristics such as: active imagination, willingness to consider new ideas, divergent thinking, and intellectual curiosity. Those who score high on this scale tend to be unconventional and independent [10]. High openness to experience has also shown to be related to the openness of feelings. It has shown to relate to aesthetic emotions, as well as to greater awareness, clarity, and intensity of their own emotions at the time [11, 15, 78]. Knoll et al [53] found that open individuals show reciprocal behavior towards emotional support. Those scoring high on openness to experience are more aware of, and more capable to judge their own emotions. Therefore, music can play a supportive role for them, hence, they would find greater benefit from browsing for music by mood.

A positive relationship between conscientiousness and the activity taxonomy was found. In other words, high conscientious people show an increased preference for activity, but not for genre. Conscientiousness refers to characteristics, such as, self-discipline. People that score high on the conscientiousness scale tend to be more plan- and goal-oriented, organized, and determined compared to those scoring low [10]. They also live lives that are overall less emotional, more balanced, more predictable, and will encounter fewer emotionally intense situations (fewer extreme lows, as well as highs [35, 83]). This dimension is considered to be the least emotionally charged, and correlated with positive and negative emotions [78]. They are also characterized as hard-working, task- and goal-oriented (taking to an extreme, they can be workaholics and perfectionists [63]). As conscientious people are least emotionally charged, and more plan- and goal-oriented, they would benefit of taxonomies that consist of concrete music categories (e.g., activities) to support their plans and goals instead of more vaguely defined categories (e.g., moods and emotions).

Lastly, we found relationships between neuroticism and the activity and genre taxonomy. This indicates that those scoring high on neuroticism are more likely to choose for activity or genre. The neuroticism dimension indicates emotional stability and personal adjustment. High scoring on neuroticism are those that frequently experience emotional distress and wide swings in emotions, while those scoring low on neuroticism tend to be calm, well adjusted, and not prone to extreme emotional reactions [10]. Additionally, those who are highly neurotic do not believe that emotions are malleable, but rather difficult to control and strong in their expressions [35]. As neurotic people do not consider emotions to be easily changed, they will not benefit much from choosing the mood taxonomy, but more of choosing the activity or genre taxonomies instead.

5.3 Study B

In Study B we looked into how the number of categories presented within a chosen music taxonomy influences the user experience (i.e., category choice satisfaction and difficulty, perceived system quality and usefulness), and how this effect is moderated by the participant's music dimension expertise (i.e., active engagement, emotion, and perceptual abilities). The conditions (6- and 24-categories) of Study B are subsequent to Study A. That is, the data acquisition of Study B originates from the behavior after participants picked a music taxonomy to continue their music browsing (Study A; Section 5.2)

In the following subsections we continue with hypotheses building, findings, and discussion.

5.3.1 Hypotheses

The overchoice effect has been extensively investigated (e.g., [6, 36, 42, 71, 77]). Overchoice is not always bound to occur; the choice set needs to satisfy preconditions. We covered the choice set preconditions in Section 4. However, overchoice does not only depend on choice set characteristics, but the users' characteristics play a role as well. A significant moderator for overchoice is the expertise of the user [12, 13, 55, 60, 74]. In line with findings showing expertise as a moderator for overchoice, we therefore hypothesize that also in the context of this study, expertise plays a role. In order to measure expertise, we rely on the different dimensions (i.e., active engagement, emotion, and perceptual abilities) of the Gold-MSI (as discussed in Section 3.2). The active engagement dimension depicts general music expertise (e.g., how much time and money one spends on music listening), while the dimensions emotion and perceptual abilities depict expertise related to the individual music taxonomies (mood and genre taxonomy respectively). For example, the emotion dimension is related to how often someone might choose music that will send shivers down their spine or how often music can evoke memories of past people and places, thereby mapping to the mood taxonomy. The perceptual ability dimension is related to how well someone can compare two pieces of music or how well someone can identify genres of music, thereby mapping to the genre taxonomy. As the active engagement dimension depicts general behavior, we believe that it has a positive effect in both the mood and genre music taxonomy.

Furthermore, we do not only investigate the effects of overchoice on choice satisfaction, but assess other parts of the user experience as well. Besides satisfaction, we also include choice difficulty, perceived system quality, and perceived system usefulness. Unless otherwise specified, we will refer to these factors as the user experience.

We hypothesize:

H1: *The number of categories within any of the taxonomies will have a positive effect on the user experience for dimension experts in active engagement, but not for non-experts.*

The dimensions emotion and perceptual abilities are more specifically oriented towards the mood and genre music taxonomy. Therefore we hypothesize:

H2: *The number of categories within the mood taxonomy will have a positive effect on the user experience for dimension experts in emotion, but not* *for non-experts*.

H3: *The number of categories within the genre taxonomy will have a positive effect on the user experience for dimension experts in perceptual abilities, but not for non-experts*.

We do not hypothesize effects of overchoice within the activity taxonomy because it depicts specific activities, and is unrelated to any kind of expertise or ambiguity. In other words, an activity that suits a situation is within the choices or not, regardless of the amount presented.

5.3.2 Findings

A multivariate analysis of variance (MANOVA) was conducted to test for user experience (i.e., perceived system usefulness, perceived system quality, choice difficulty, and choice satisfaction) differences between 6- and 24-categories. With the MANOVA we tested only for differences between the number of categories within each music taxonomy without controlling for expertise. Table 6 and 7 show the categories that the participants chose, and the total distribution across the music taxonomies respectively. Results show that for the mood, activity, and genre taxonomies, participants did not experience any significant difference whether it was the smaller choice set or the bigger choice set that they were choosing from. User experience factors were not significantly different when choosing from a choice set of 6- or 24-categories as well.

	Mood	**#**	**Activity**	**#**	**Genre**	**#**
0	None	2	None	3	None	20
1	Energetic	8	Relaxing	0	Pop	16
2	Happy	6	Being creative	2	Rock	30
3	Soothing	5	Rainy day	0	Classical Rock	21
4	Mellow	10	Staying up all night	4	Indie Rock	11
5	Atmospheric	3	Road trip	0	Indie Pop	2
6	Hypnotic	2	Working/studying without lyrics	3	Easy Listening	7
Total		36		12		107

Table 6: Distribution of chosen categories within each taxonomy (6-categories condition).

	Mood	#	Activity	#	Genre	#
0	None	0	None	0	None	5
1	Energetic	2	Relaxing	0	Pop	9
2	Happy	4	Being creative	0	Rock	9
3	Soothing	4	Rainy day	0	Classical Rock	8
4	Mellow	3	Staying up all night	0	Indie Rock	5
5	Atmospheric	1	Road trip	0	Indie Pop	2
6	Hypnotic	0	Working/studying without lyrics	0	Easy Listening	4
7	Introspective	1	Reading in a coffee shop	0	Classical	0
8	Warm	1	Singing in the shower	0	Blues/Rock	3
9	Motivational	1	Housework	2	Film scores	4
10	Funky	1	Working/studying with lyrics	0	Folk	5
11	Sad	1	Romantic evening	0	Dance	1
12	Celebratory	0	Gaming	0	R&B	6
13	Nocturnal	1	Energy boost	0	Classic Pop	4
14	Aggressive	0	Working out: weight training	0	Rap	6
15	Seductive	1	Unwinding after work	0	Oldies	7
16	Gloomy	1	Working out: cardio	2	Electronica	7
17	Sweet	0	Beach party	0	Jazz	5
18	Classy	2	House party	0	Modern Rock	8
19	Sexual	3	Barbequing	0	Electronic Indie	2
20	Raw	1	Lying low on the weekend	0	Dance/House/Techno	2
21	Angsty	1	Sleeping	0	Singer-songwriter	1
22	Visceral	1	City cruising	0	Dance Pop	1
23	Spacey	2	Waking up on the right side of the bed	0	Funk	1
24	Trippy	0	Lying on a beach	0	Dubstep/Drum and Bass	1
Total		32		4		106

Table 7: Distribution of chosen categories within each taxonomy (24-categories condition).

In order to investigate the effects of expertise, we conducted a moderated multiple regression (MMR) analysis. We used the dimensions of the Gold-MSI (i.e., active engagement, perceptual abilities, and emotions) to assess participants' expertise level, and added these as a moderator to the analyses. This allowed us to investigate how expertise influences the overchoice effect and the user experience factors (i.e., perceived system usefulness, perceived system quality, choice difficulty, and choice satisfaction).

The analyses were conducted in two steps (the results of all steps can be found in Appendix D). In the first step we tested for main effects. This allowed us to see the general effects of expertise on the user experience factors within each music taxonomy, regardless of the number of categories. The second step involved the moderators (i.e., emotion, perceptual abilities, and active engagement dimension expertise). By including the moderators, we were able to look at how expertise influences overchoice, and in turn the user experience.

We separately discuss the significant findings of each music taxonomy on the user experience factors (i.e., perceived system usefulness, perceived system quality, choice difficulty, and choice satisfaction) below. In each of the following result sections, we first start with the significant main effects (i.e., the effect of expertise on the user experience without taking into account the different number of categories). After that we continue with the significant moderator effects (i.e., the effect of expertise on overchoice and the user experience).

Mood Taxonomy. When looking at the results of perceived system usefulness, we found a significant main effect of emotion expertise (t(1, 63)=1.939, p=0.05; full results in Appendix D.1). This indicates that in general participants that are emotion experts found the system more useful than non-experts. For perceived system quality, we found a significant main effect of active engagement expertise (t(1, 63)= -2.379, p=0.02), as well as emotion expertise (t(1, 63)=2.285, p=0.02; full results in Appendix D.2). This means that active engaged participants indicated that they perceived the system of lower quality while participations with emotion expertise rated the system of higher quality. Furthermore, we found a main effect on choice satisfaction of emotion expertise (t(1, 63) = 1.764, p=0.08), indicating that those who use music for emotional activities are in general more satisfied with their category label choice.

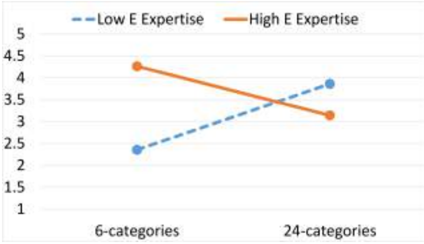

Figure 5: Moderator effect of emotion (E) expertise on **perceived system usefulness** (higher means more useful) within the mood taxonomy.

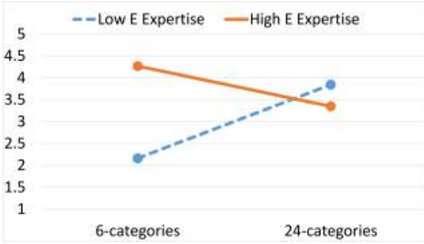

Figure 6: Moderator effect of emotion (E) expertise on **perceived system quality** (higher means higher quality) within the mood taxonomy.

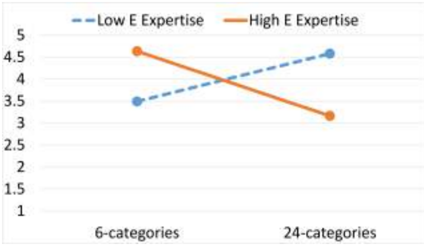

Figure 7: Moderator effect of emotion (E) expertise on **choice difficulty** (higher means easier) within the mood taxonomy.

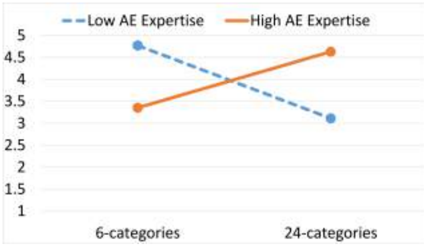

Figure 8: Moderator effect of active engagement (AE) expertise on **choice difficulty** (higher means easier) within the mood taxonomy.

When looking at differences between the number of categories while controlling for the expertise dimensions, we found the following moderator effects on the different factors of the user experience. For the perceived system usefulness, we found a significant moderator effect of emotion expertise (t(1, 63)=-2.147, p=0.03). The results of the moderator effect indicate that emotion experts perceived the system as less useful when given more choices, while non-emotion experts perceived the system as more useful when given more choices (Figure 5). When looking at the perceived system quality, we found a moderator effect of emotion expertise (t(1, 63)=-1.834, p=0.07), indicating that emotion experts perceived the system of less quality when given more choices (Figure 6). Lastly, we identified moderator effects on choice difficulty by emotion expertise and active engagement expertise (full results in Appendix D.3). Emotion experts show a decrease in choice difficulty when given less choices (t(1, 63)=-1.754, p=0.08; Figure 7), whereas active engagement experts show a decrease of choice difficulty when given more choices (t(1, 63)=2.385, p=0.02; Figure 8). No significant effects were found on choice satisfaction.

Activity Taxonomy. As expected, no main or moderator effects were found for the categories within the activity taxonomy.

Genre Taxonomy. No main effects were found of the different expertise dimensions on the user experience. However, moderator effects were observed on the user experience factors when looking at the differences between the number of categories. A significant moderator effect was found on perceived system usefulness when controlling for perceptual abilities expertise (t(1, 197)=2.260, p=0.02; full results in Appendix D.4). Participants with expertise in perceptual abilities rated the system as more useful when given more choices. On the other hand, those with low perceptual abilities rated the system as more useful when given less choices (Figure 9). For perceived system quality, we found a moderator effect of perceptual abilities expertise (t(1, 197)=1.838, p=0.06; full results in Appendix D.5). The results show that perceptual experts rated the system of higher quality when given more choices, while it hardly made a difference for non-experts (Figure 10). No significant effects were found on choice satisfaction or choice difficulty by expertise in perceptual abilities, nor did we find any effects on the user experience factors by active engagement.

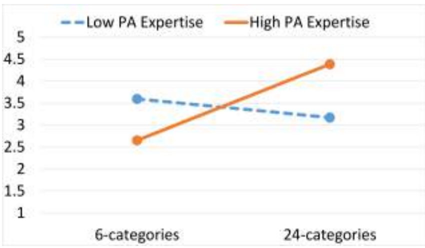

Figure 9: Moderator effect of perceptual abilities (PA) expertise on **perceived system usefulness** (higher means more useful) within the genre taxonomy.

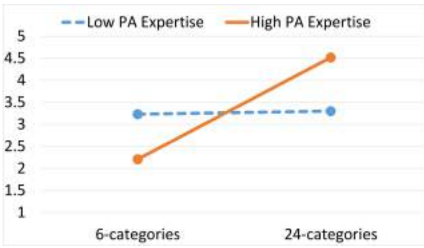

Figure 10: Moderator effect of perceptual abilities (PA) expertise on **perceived system quality** (higher means higher quality) within the genre taxonomy.

5.3.3 Discussion

Our results show that expertise plays a role in whether overchoice occurs or not. With regards to H1, we only found partial support. We hypothesized that general music expertise (active engagement), would play a role in whether overchoice occurs. However, we only found an overchoice effect in the mood taxonomy on choice difficulty. Those who were more expert indicated to find it more difficult to choose a category when they were given less choice, whereas non-experts indicated to experience more difficulties when given more choice. As this was the only effect found, the effect of expertise seem to be very specific, and cannot take any general form.

Remarkable is the effect of emotion expertise within the mood taxonomy. Here, the emotion expertise seem to adopt an opposite effect of overchoice. Therefore, we need to reject our hypothesis (H2). Instead of an increase in the user experience factors when given more choice, emotion experts show a decrease. In other words, they perceived the system as more useful and of higher quality, and indicated to have less difficulties to pick a category, when provided less choice than when given more choice. Whereas, non-experts indicated the opposite effect and were experiencing a higher user experience when given more choices. A possible explanation for this could be that emotional experts are in general more emotionally aroused and therefore prefer less choice because it takes less cognitive effort. This is in line with findings that show that emotional arousal can have an adverse effect on decision making because of reduced cognitive processing [22, 56]. In other words, information processing decreases as a result of emotional arousal. Making a choice from a bigger choice set would then take more effort to assess every option. Also, especially for those who rely more on the emotional triggers of music, making a bad choice will have bigger consequences than making a good choice [3]. Hence, as the choice sets within each music taxonomy were designed to be most attractive, choice difficulty within the mood taxonomy is exacerbated for the more experienced ones.

The effect of expertise in the genre taxonomy is partially in line with our hypothesis (H3). Prior research suggests that expertise is a moderator for overchoice [12, 13, 55, 60, 74]. Those who indicated to be experts in perceptual abilities rated the system of higher quality, and more useful, when more choices were provided.

Striking is that we did not observe a clear overchoice effect on the choice that was made (i.e., choice difficulty and choice satisfaction), but only on the evaluation of the system (i.e., perceived system usefulness and perceived system quality). Evaluating the necessary preconditions for overchoice to occur state that the user needs to have a lack of familiarity with the items, and should not have a clear prior preference for an item [41]. However, not meeting these preconditions should lead to preferring more choice [12, 13], whereas our results show no differences. Others argue that overchoice can only occur when all options are attractive. So, there should be no dominant option and the proportion of non-dominant options should be large [18, 19, 38, 68]. Otherwise, making a decision would be easy, regardless of the size of the choice set. In this study, we tried to control for that by creating choice sets with the most attractive items (see the preliminary study in Section 4). Also, by looking at the distribution of the choices made by the participants (see Table 6 and 7), there is no category that excessively stands out of being chosen. The most plausible explanation for why we did not observe the hypothesized effects comes from Hutchinson [40]. He argued that overchoice seldom occurs among animals, because they seem to have adapted to the different sizes of choice sets that naturally occur in their environment. Although this hypothesis has not been verified on humans so far, it would explain best why the overchoice effect on the choices made (i.e., choice difficulty and choice satisfaction) was not found in our study. The sizes of the choice sets we used are not uncommon for music streaming services. We picked the size of our largest choice set (24 categories) to be in line with the original work of overchoice by Iyengar and Lepper [41]. However, this was just a subset of what would be presented to actual users of such a service. It could be that our participants are accustomed to the sizes of the choice sets we presented, as normally they would need to deal with even larger choice sets.

Although we did not experience the overchoice effect on the category items, it does not mean that our choice sets did not have any effects. We did find effects on the factors evaluating the system (i.e., perceived system usefulness and perceived system quality). These are important factors that help to form users' general perspective of the system as a whole.

5.4 Implications

The results of these studies support the creation of personalized user interfaces by taking into account the user's personality and expertise (a proposed user model can be found in Figure 11). With applications getting more and more connected and sharing resources (e.g., applications connect with social networking sites, such as, Facebook, Twitter, or Instagram), the automatic extraction of personality and expertise becomes more available. A possible scenario could be:

A user has the music application connected to his Facebook account. Based on his Facebook profile, the application inferred that he is someone open to new experiences. Therefore, the music application adjusts the user interface by emphasizing the mood taxonomy to let the user continue browsing for music. By analyzing his profile (e.g., he filled in artists and bands that he likes) and postings (e.g., posting often that he goes to concerts), the system may infer that he is actively engaged with music. Based on this, the system decides to provide him more categories to choose from within the mood taxonomy.

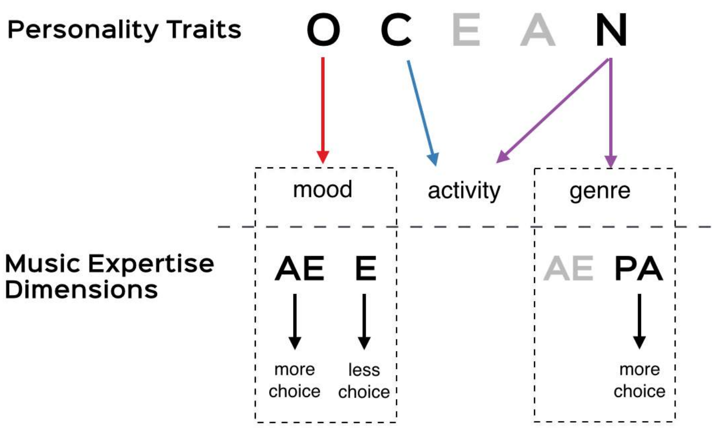

Figure 11: Proposed user model. Personality traits: (O)penness to experience, (C)onscientiousness, (E)xtraversion, (A)greeableness, (N)euroticism. Music expertise dimensions: active engagement (AE), perceptual abilities (PA), emotions (E).

In the last couple of years, it has been demonstrated that personality information can be extracted from social networking sites (SNSs) like Facebook (e.g., [2, 29, 33, 64, 72]), Twitter (e.g., [32, 66]), and Instagram (e.g., [28, 30]). Their findings show that SNSs offer reliable cues to infer different personality traits. Being able to extract personality traits from SNSs caters the possibility for (music) applications to adjust their user interface based on our results. For example, when someone appears to be open to new experiences, the mood taxonomy could be emphasized while other taxonomies could be placed more in the background of the interface. In addition, music recommendations could be given based on the mood taxonomy (e.g., music with similar mood expression).

We believe that also the expertise dimensions (i.e., active engagement, perceptual abilities, and emotions) that we used in Study B, can be inferred from the same increased connectedness with SNSs. For example, active engagement can be inferred by extracting information on concert attendance (e.g., Facebook events, SongKick; `http://www.songkick.com`) as well as purchase behavior (e.g., iTunes store, Amazon; `http://www.amazon.com`). The “About" section, or the posted activities and status updates in SNSs can provide cues to infer perceptual abilities. Analyzing postings of a SNS user could give an indication about the emotion expertise dimension (e.g., postings about induced feelings when listening to a song). Also, there seems to be some relationship between factors of the emotion expertise dimension and the openness to experience personality factor. This could serve as an additional indicator. Music applications could anticipate the choice set based on the expertise dimension of the user.

6 Limitations & Future work

There are several limitations in this study that should be addressed in future work. Our sample focused only on participants situated in the United States. It is possible that cultural differences play a role in taxonomy usage and category preferences. Future work should address this.

We tested the relationship between personality traits and independent music taxonomies (i.e., mood, activity, and genre). One of our results show a relationship between neuroticism and the activity and genre taxonomy.On the other hand, it could well be that people prefer combinations of different taxonomies (e.g., sad pop music, funky road trip music, or happy cooking music).

In the studies we conducted, we intentionally did not include real music recommendations as we believed this could interfere with rightfully answering our research questions. Since this study only simulated the decision making stage of using a music streaming service and did not play any actual music, it may have limiting effects on the holistic user experience.

7 Conclusion

The goal of this work was to investigate whether music browsing strategies are related to personality traits, by looking at the decision making of picking a music taxonomy (mood, activity, or genre) to browse for music. Additionally, we looked at the occurrence of overchoice with the number of categories within the music taxonomies, and how this effect is moderated by expertise.

We found that users' choice of a taxonomy (mood, activity, or genre) to browse for music, is related to their personality. Furthermore, our results, show that overchoice is moderated by expertise. We found that the effects of overchoice is counteracted by expertise in the genre taxonomy. However, having more expertise/experience does not always make choosing easier. In our case, emotion experts (e.g., those who easily identify with emotions in music) had more difficulties making a decision with an increased choice set. Although expertise may take the role as a proxy measure for cognitive processing, by assuming that expertise and experience with the topic makes processing information about the topic easier, this does not always seem true. It seems that in some cases, expertise or more experience can create adverse effects.

Finally, while the majority of prior research focuses on the influence of overchoice on choice satisfaction and/or choice difficulty, we show with our results that overchoice does not necessarily limit its influence to these two factors. Our results show that even when choice satisfaction or difficulty are not affected by the overchoice effect, it may still influence other aspects of the user experience (e.g., system usefulness, and system quality). These other factors of the user experience should not be neglected, and could play an important role in the recurring use of the system by users.

8 Acknowledgements

This research is supported by the Austrian Science Fund (FWF): P25655; and by the EU FP7/2013-2016 through the PHENICX project under grant agreement no.601166.

Appendix A - Categories

Music categories extracted from Songza (`http://www.songza.com`)

Nr	Mood	Activity	Genre
1	Visceral	Housework	Blues & Blues Rock
2	Mellow	Drinking at a Dive Bart	Bluegrass
3	Celebratory	Hanging Out in the Man Cave	Children's
4	Warm	Road Trip	Christian
5	Motivational	Skateboarding	Christian: Gospel
6	Angsty	Working/Studying (without lyrics)	Christmas
7	Trashy	Getting High	Classical
8	Seductive	Going for a Bike Ride	Classical: Crossover
9	Hypnotic	Unwinding After Work	Classical: Vocal
10	Rowdy	Staying up all Night	Country
11	Aggressive	Dance Party: Beach	Country: Contemporary Country
12	Sweet	Singing in the Shower	Dance
13	Soothing	Energy Boost	Dance: Disco & Nu Disco
14	Introspective	Slow Dancing	Dance: House & Techno
15	Raw	Sitting on a Back Porch	Dancehall
16	Gloomy	Reading in a Coffee Shop	Dubstep & Drum 'n Bass
17	Atmospheric	Working Out: Cardio	Easy Listening
18	Nocturnal	Breaking Up	Electronica
19	Cold	Making Out	Film Scores
20	Spacey	Dinner Party: Formal	Folk
21	Earthy	Workout Cool Down	Funk
22	Lush	Getting Lucky	Hawaiian
23	Sexual	Driving in the Left Lane	Indie: Indie Electronic
24	Classy	Working Out: Weight Training	Indie: Indie Folk & Americana
25	Trippy	Lying on a Beach	Indie: Indie Pop
26	Energetic	House Party	Indie: Indie Rock
27	Sprightly	Barbecuing	International/World
28	Funky	Cocktail Party	International: African
29	Campy	Romantic Evening	International: Asian
30	Happy	Rainy Day	International: Brazilian
31	Sad	Waking Up on the Right Side of the Bed	International: Jamaican
32	Cocky	Working/Studying (with lyrics)	International: Mediterranean

33		Dance Party: Sweaty	Jazz
34		Lounging in a Cool Hotel	Jazz: Vocal Jazz
35		Lying Low on the Weekend	Latin
36		Yoga	Latin: Cuban
37		Pleasing Crowd	Latin: Puerto Rican
38		Coding	Latin: Salsa
39		Pool Party	Latin: Tropical
40		Sleeping	Nature Sounds & Soundscapes
41		Dinner Party: Casual	Oldies
42		Gaming	Pop
43		Relaxing	Pop: Classic Pop
44		City Cruising	Pop: Dance Pop
45		Coming Down After a Party	Pop: Soft Pop
46		Stripping	R&B
47		Shopping at a Vintage Store	R&B: Classic R&B
48		Ballroom Dancing	R&B: Contemporary R&B
49		Dirt Road Driving	R&B: Soul
50		Walking Through a City	Rap
51		Dance Party: Fun & Funky	Rap: Classic Mainstream Rap
52		Being Creative	Rap: Old School Rap
53		Cooking with Friends	Rap: Today's Mainstream Rap
54		Girls Night Out	Rap: Underground & Alternative Rap
55		Getting Married	Reggae & Ska
56		Grinding at a Nightclub	Reggaeton
57			Rock
58			Rock: Classic Alternative & Punk
59			Rock: Contemporary Alternative
60			Rock: Emo/Pop-Punk
61			Rock: Hard Rock
62			Rock: Metal
63			Rock: Modern Rock
64			Rock: Rockabilly
65			Singer-Songwriter
66			Showtunes

Appendix B - User Experience

Below the questions depicting the user experience (adapted from [52]).

B.1 Choice Satisfaction

5-point Likert scale: disagree strongly - agree strongly.

Nr	Question
1	I don't like the item I chose (negated).
2	I am enthusiastic about the item I chose.

B.2 Perceived Choice Difficulty

5-point Likert scale: very difficult - very easy.

Nr	Question
1	How difficult was it to choose an item from the list?

B.3 Perceived System Usefulness

5-point Likert scale: disagree strongly - agree strongly.

Nr	Question
1	With this way of finding music, I can make better choices.
2	I don't find this way of finding music useful (negated).
3	I would use this way of finding music more often if it was possible.

B.4 Perceived System Quality

5-point Likert scale: disagree strongly - agree strongly.

Nr	Question
1	I found good items in the list.
2	The list did not consist any of my preferred items (negated).

Appendix C - Music sophistication

Below the questions belonging to corresponding parts of the Gold-MSI (5-point Likert scale: disagree strongly - agree strongly. Adopted from [61]).

C.1 Music Emotions

Nr	Question
1	I sometimes choose music that can trigger shivers down my spine.
2	Pieces of music rarely evoke emotions for me.
3	I often pick certain music to motivate or excite me.
4	Music can evoke my memories of past people and places.

C.2 Active Engagement

Nr	Question
1	I spend a lot of my free time doing music-related activities.
2	I often read or search the Internet for things related to music.
3	I don't spend much of my disposable income on music.
4	Music is kind of an addiction for me - I couldn't live without it.
5	I keep track of new music that I come across (e.g., new artists or recordings).

C.3 Perceptual Abilities

Nr	Question
1	I am able to judge whether someone is a good singer or not.
2	I find it difficult to spot mistakes in a performance of a song even if I know the tune.
3	I can tell when people sing or play out of time with the beat.
4	I can tell when people sing or play out of tune.
5	When I hear a music I can usually identify its genre.

Appendix D - Results

Below the results of the moderated multiple regression for the mood and genre taxonomy. Step 1 depicts the analyses for the main effects (i.e., general effect of expertise on the user experience without taking into account the different number of categories), and Step 2 depicts the moderator effects (i.e., the effects of expertise on the overchoice effect and the user experience)

D.1 Mood Taxonomy & Perceived System Usefulness

		b	SE b	β
Step 1	Constant	3.404	0.529	***
	6- or 24-item	0.096	0.172	0.071
	Active engagement	-0.112	0.127	-0.162
	Emotions	0.297	0.153	0.340ˆ
	Perceptual abilities	-0.059	0.127	-0.074
Step 2	Constant	2.144	0.755	**
	6- or 24-item	2.446	1.036	1.807*
	Active engagement	-0.28	0.174	-0.404
	Emotions	0.7	0.241	0.802**
	Perceptual abilities	-0.036	0.152	-0.045
	Item x Active engagement	0.366	0.258	1.016
	Item x Emotions	-0.657	0.306	-2.138*
	Item x Perceptual abilities	-0.203	0.271	-0.606

Note. R^2=0.063 for step 1, R^2=0.160 for step 2. ˆp<0.1, *p<.05, **p<.01, ***p<.001.

D.2 Mood Taxonomy & Perceived System Quality

		b	SE b	β
Step 1	Constant	3.404	0.61	***
	6- or 24-item	0.191	0.198	0.119
	Active engagement	-0.347	0.146	-0.422*
	Emotions	0.403	0.176	0.390*
	Perceptual abilities	-0.001	0.147	-0.001
Step 2	Constant	1.864	0.874	*
	6- or 24-item	2.903	1.198	1.808
	Active engagement	-0.396	0.201	-0.482
	Emotions	0.804	0.278	0.777
	Perceptual abilities	-0.008	0.176	-0.008
	Item x Active engagement	0.092	0.299	0.214
	Item x Emotions	-0.65	0.354	-1.782ˆ
	Item x Perceptual abilities	-0.06	0.313	-0.152

Note. R^2=0.116 for step 1, R^2=0.200 for step 2. ˆp<0.1, *p<.05, **p<.01, ***p<.001.

D.3 Mood Taxonomy & Choice Difficulty

		b	SE b	β
Step 1	Constant	4.906	0.634	***
	6- or 24-item	-0.096	0.205	-0.060
	Active engagement	0.025	0.152	-0.067
	Emotions	-0.069	0.183	-0.067
	Perceptual abilities	-0.114	0.153	-0.121
Step 2	Constant	3.967	0.900	***
	6- or 24-item	1.925	1.234	1.212
	Active engagement	-0.293	0.208	-0.360
	Emotions	0.320	0.287	0.313
	Perceptual abilities	-0.027	0.182	-0.029
	Item x Active engagement	0.734	0.308	1.735
	Item x Emotions	-0.640	0.365	-1.775ˆ
	Item x Perceptual abilities	-0.462	0.323	-1.178*

Note. R^2=0.024 for step 1, R^2=0.133 for step 2. ˆp<0.1, *p<.05, **p<.01, ***p<.001.

D.4 Genre Taxonomy & Perceived System Usefulness

		b	SE b	β
Step 1	Constant	3.45	0.499	***
	6- or 24-item	0.327	0.134	0.174*
	Active engagement	-0.056	0.094	-0.054
	Emotions	0.043	0.12	0.031
	Perceptual abilities	0.068	0.119	0.047
Step 2	Constant	4.148	0.717	***
	6- or 24-item	-0.863	0.975	-0.458
	Active engagement	-0.037	0.136	-0.036
	Emotions	0.158	0.164	0.115
	Perceptual abilities	-0.242	0.182	-0.168
	Item x Active engagement	-0.004	0.188	-0.007
	Item x Emotions	-0.232	0.239	-0.511
	Item x Perceptual abilities	0.54	0.239	1.172*

Note. R^2=0.032 for step 1, R^2=0.060 for step 2. ˆp<0.1, *p<.05, **p<.01, ***p<.001.

D.5 Genre Taxonomy & Perceived System Quality

		b	SE b	β
Step 1	Constant	3.313	0.631	***
	6- or 24-item	0.595	0.17	0.246
	Active engagement	-0.039	0.119	-0.029
	Emotions	0.027	0.152	0.015
	Perceptual abilities	0.049	0.15	0.027
Step 2	Constant	3.933	0.913	***
	6- or 24-item	-0.468	1.242	-0.194
	Active engagement	0.073	0.173	0.054
	Emotions	0.1	0.209	0.056
	Perceptual abilities	-0.276	0.232	-0.149
	Item x Active engagement	-0.18	0.239	-0.271
	Item x Emotions	-0.135	0.304	-0.232
	Item x Perceptual abilities	0.56	0.305	0.945ˆ

Note. R^2=0.060 for step 1, R^2=0.077 for step 2. ˆp<0.1., *p<.05, **p<.01, ***p<.001.

References

[1] Arnold SJ, Oum TH, Tigert DJ (1983) Determinant attributes in retail patronage: seasonal, temporal, regional, and international comparisons. Journal of Marketing Research pp 149–157

[2] Back MD, Stopfer JM, Vazire S, Gaddis S, Schmukle SC, Egloff B, Gosling SD (2010) Facebook profiles reflect actual personality, not self-idealization. Psychological science

[3] Baumeister RF, Vohs KD, DeWall CN, Zhang L (2007) How emotion shapes behavior: Feedback, anticipation, and reflection, rather than direct causation. Personality and Social Psychology Review 11(2):167–203

[4] Bertin-Mahieux T, Eck D, Mandel M (2011) Automatic tagging of audio: The state-of-the-art. Machine audition: Principles, algorithms and systems pp 334–352

[5] Boatwright P, Nunes JC (2001) Reducing assortment: An attribute-based approach. Journal of marketing 65(3):50–63

[6] Bollen D, Knijnenburg BP, Willemsen MC, Graus M (2010) Understanding choice overload in recommender systems. In: RecSys, ACM, pp 63–70

[7] Borle S, Boatwright P, Kadane JB, Nunes JC, Galit S (2005) The effect of product assortment changes on customer retention. Marketing science 24(4):616–622

[8] Bown NJ, Read D, Summers B (2003) The lure of choice. Journal of Behavioral Decision Making 16(4):297–308

[9] Bravo MJ, Farid H (2006) Object recognition in dense clutter. Perception & psychophysics 68(6):911–918

[10] Burger J (2010) Personality. PSY 235 Theories of Personality Series, Cengage Learning, URL http://books.google.at/books?id=bZY7I2_8yRMC

[11] Buss AH (1980) Self-consciousness and social anxiety. Freeman

[12] Chernev A (2003) Product assortment and individual decision processes. Journal of Personality and Social Psychology 85(1):151

[13] Chernev A (2003) When more is less and less is more: The role of ideal point availability and assortment in consumer choice. Journal of consumer Research 30(2):170–183

[14] Chia PH, Yamamoto Y, Asokan N (2012) Is this app safe?: a large scale study on application permissions and risk signals. In: WWW, ACM

[15] Costa PT, McCrae RR (1992) Neo pi-r professional manual

[16] Craig CS, Ghosh A, McLafferty S (1984) Models of the retail location process-a review. journal of retailing 60(1):5–36

[17] Dhar R (1996) The effect of decision strategy on deciding to defer choice. Journal of Behavioral Decision Making 9(4):265–281

[18] Dhar R (1997) Consumer preference for a no-choice option. Journal of Consumer Research 24(2):215–231

[19] Dhar R, Nowlis SM (1999) The effect of time pressure on consumer choice deferral. Journal of Consumer Research 25(4):369–384

[20] Diehl K, Poynor C (2010) Great expectations?! assortment size, expectations, and satisfaction. Journal of Marketing Research 47(2):312–322

[21] Dreze X, Hoch SJ, Purk ME (1995) Shelf management and space elasticity. Journal of Retailing 70(4):301–326

[22] Easterbrook JA (1959) The effect of emotion on cue utilization and the organization of behavior. Psychological review 66(3):183

[23] Elahi M, Braunhofer M, Ricci F, Tkalcic M (2013) Personality-based active learning for collaborative filtering recommender systems. In: AI* IA 2013: Advances in Artificial Intelligence, Springer, pp 360–371

[24] Fasolo B, Hertwig R, Huber M, Ludwig M (2009) Size, entropy, and density: What is the difference that makes the difference between small and large real-world assortments? Psychology & Marketing 26(3):254–279

[25] Ferwerda B (2015) The soundtrack of my life: Adjusting emotion of music. In: Proceedings of the 1st Workshop Collaborating with Intelligent Machines, Seoul, South Korea

[26] Ferwerda B, Schedl M (2014) Enhancing music recommender systems with personality information and emotional states: A proposal. In: Proceedings of the 2nd Workshop Emotions and Personality in Personalized Services (EMPIRE 2014), Aalborg, Denmark, pp 1613–0073

[27] Ferwerda B, Schedl M, Tkalcic M (2015) Personality & emotional states: Understanding usersâĂŹ music listening needs. UMAP

[28] Ferwerda B, Schedl M, Tkalcic M (2015) Predicting personality traits with instagram pictures. In: Proceedings of the 3rd Workshop on Emotions and Personality in Personalized Systems 2015, ACM, New York, NY, USA, EMPIRE '15, pp 7–10, DOI 10.1145/2809643.2809644, URL http://doi.acm.org/10.1145/2809643.2809644

[29] Ferwerda B, Schedl M, Tkalcic M (2016) Personality traits and the relationship with (non-)disclosure behavior on facebook. In: Proceedings of the International Workshop on Modeling Social Media, ACM

[30] Ferwerda B, Schedl M, Tkalcic M (2016) Using instagram picture features to predict usersâĂŹ personality. In: MultiMedia Modeling, Springer, pp 850–861

[31] Fu Z, Lu G, Ting KM, Zhang D (2011) A survey of audio-based music classification and annotation. Multimedia, IEEE Transactions on pp 303–319

[32] Golbeck J, Robles C, Edmondson M, Turner K (2011) Predicting personality from twitter. In: Privacy, security, risk and trust (passat), 2011 ieee third international conference on and 2011 ieee third international conference on social computing (socialcom), IEEE, pp 149–156

[33] Gosling SD, Gaddis S, Vazire S, et al (2007) Personality impressions based on facebook profiles. ICWSM 7:1–4

[34] Gourville JT, Soman D (2005) Overchoice and assortment type: When and why variety backfires. Marketing science 24(3):382–395

[35] Gross JJ, John OP (1998) Mapping the domain of expressivity: multimethod evidence for a hierarchical model. Journal of personality and social psychology p 170

[36] Haynes GA (2009) Testing the boundaries of the choice overload phenomenon: The effect of number of options and time pressure on decision difficulty and satisfaction. Psychology & Marketing 26(3):204–212

[37] Herpen Ev, Diehl K, Poynor C (2008) Arranged to distraction: how categorizing products with complements versus substitutes alters the experience of product choice. Advance of Consumer Resarch 35:80–83

[38] Hsee CK, Leclerc F (1998) Will products look more attractive when presented separately or together? Journal of Consumer Research 25(2):175–186

[39] Hu R, Pu P (2009) Acceptance issues of personality-based recommender systems. In: RecSys, ACM, pp 221–224

[40] Hutchinson J (2005) Is more choice always desirable? evidence and arguments from leks, food selection, and environmental enrichment. Biological Reviews 80(1):73–92

[41] Iyengar SS, Lepper MR (2000) When choice is demotivating: Can one desire too much of a good thing? Journal of personality and social psychology 79(6):995

[42] Iyengar SS, Huberman G, Jiang W (2004) How much choice is too much? contributions to 401 (k) retirement plans. Pension design and structure: New lessons from behavioral finance pp 83–95

[43] Iyengar SS, Wells RE, Schwartz B (2006) Doing better but feeling worse looking for the âĂIJbestâĂİ job undermines satisfaction. Psychological Science 17(2):143–150

[44] Janiszewski C (1998) The influence of display characteristics on visual exploratory search behavior. Journal of Consumer Research 25(3):290–301

[45] John OP, Donahue EM, Kentle RL (1991) The big five inventory: Versions 4a and 54, institute of personality and social research. University of California, Berkeley, CA

[46] Jung C (1976) Psychological Types. No. v. 6 in Bollingen Series No. 20, Princeton University Press

[47] Jung CG (1923) Psychological types. New York Retrieved September 15, 2014, from `http://psychclassics.yorku.ca/Jung/types.htm`

[48] Kahn BE (1995) Consumer variety-seeking among goods and services: An integrative review. Journal of Retailing and Consumer Services 2(3):139–148

[49] Kahn BE, Wansink B (2004) The influence of assortment structure on perceived variety and consumption quantities. Journal of Consumer Research 30(4):519–533

[50] Kemp AE (1996) The musical temperament: Psychology and personality of musicians. Oxford University Press

[51] Kittur A, Chi EH, Suh B (2008) Crowdsourcing user studies with mechanical turk. In: Proceedings of the SIGCHI conference on human factors in computing systems, ACM, pp 453–456

[52] Knijnenburg BP, Willemsen MC, Gantner Z, Soncu H, Newell C (2012) Explaining the user experience of recommender systems. User Modeling and User-Adapted Interaction 22(4-5):441–504

[53] Knoll N, Burkert S, Schwarzer R (2006) Reciprocal support provision: Personality as a moderator? European journal of personality 20(3):217–236

[54] Koelemeijer K, Oppewal H (1999) Assessing the effects of assortment and ambience: a choice experimental approach. Journal of Retailing 75(3):319–345

[55] Langner T, Krengel M (2013) The mere categorization effect for complex products: The moderating role of expertise and affect. Journal of Business Research 66(7):924–932

[56] Leith KP, Baumeister RF (1996) Why do bad moods increase self-defeating behavior? emotion, risk tasking, and self-regulation. Journal of personality and social psychology 71(6):1250

[57] Lonsdale AJ, North AC (2011) Why do we listen to music? a uses and gratifications analysis. British Journal of Psychology 102(1):108–134

[58] Mazursky D, Jacoby J (1986) Exploring the development of store images. Journal of Retailing 62(2):145–165

[59] McCrae RR, John OP (1992) An introduction to the five-factor model and its applications. Journal of personality 60(2):175–215

[60] Mogilner C, Rudnick T, Iyengar SS (2008) The mere categorization effect: How the presence of categories increases choosersâĂŹ perceptions of assortment variety and outcome satisfaction. Journal of Consumer Research 35(2):202–215

[61] Müllensiefen D, Gingras B, Musil J, Stewart L (2014) The musicality of non-musicians: an index for assessing musical sophistication in the general population. PloS one 9(2):e89,642

[62] Oppewal H, Koelemeijer K (2005) More choice is better: Effects of assortment size and composition on assortment evaluation. International Journal of Research in Marketing 22(1):45–60

[63] Ozer DJ, Benet-Martinez V (2006) Personality and the prediction of consequential outcomes. Annu Rev Psychol 57:401–421

[64] Park G, Schwartz HA, Eichstaedt JC, Kern ML, Kosinski M, Stillwell DJ, Ungar LH, Seligman ME (2014) Automatic personality assessment through social media language.

[65] Perruchet P, Rey A, Hivert E, Paction S (2006) Do distractors interfere with memory for study pairs in associative recognition? Memory & cognition

[66] Quercia D, Kosinski M, Stillwell D, Crowcroft J (2011) Our twitter profiles, our selves: Predicting personality with twitter. In: Privacy, security, risk and trust (passat), 2011 ieee third international conference on and 2011 ieee third international conference on social computing (socialcom), IEEE, pp 180–185

[67] Rawlings D, Ciancarelli V (1997) Music preference and the five-factor model of the neo personality inventory. Psychology of Music 25(2):120–132

[68] Redelmeier DA, Shafir E (1995) Medical decision making in situations that offer multiple alternatives. Jama 273(4):302–305

[69] Rentfrow PJ, Gosling SD (2003) The do re mi's of everyday life: the structure and personality correlates of music preferences. Journal of personality and social psychology 84(6):1236

[70] Rentfrow PJ, Goldberg LR, Zilca R (2011) Listening, watching, and reading: The structure and correlates of entertainment preferences. Journal of personality 79(2):223–258

[71] Reutskaja E, Hogarth RM (2009) Satisfaction in choice as a function of the number of alternatives: When âĂIJgoods satiateâĂİ. Psychology & Marketing 26(3):197–203

[72] Ross C, Orr ES, Sisic M, Arseneault JM, Simmering MG, Orr RR (2009) Personality and motivations associated with facebook use. Computers in Human Behavior 25(2):578–586

[73] Scheibehenne B, Greifeneder R, Todd PM (2009) What moderates the too-much-choice effect? Psychology & Marketing 26(3):229–253

[74] Scheibehenne B, Greifeneder R, Todd PM (2010) Can there ever be too many options? a meta-analytic review of choice overload. Journal of Consumer Research 37(3):409–425

[75] Schwartz B (2004) The paradox of choice. Ecco New York

[76] Schwartz B (2004) The tyranny of choice. SCIENTIFIC AMERICAN-AMERICAN EDITION- 290(4):70–75

[77] Shah AM, Wolford G (2007) Buying behavior as a function of parametric variation of number of choices. Psychological Science 18(5):369–370

[78] Shiota MN, Keltner D, John OP (2006) Positive emotion dispositions differentially associated with big five personality and attachment style. The Journal of Positive Psychology 1(2):61–71

[79] Sloot LM, Fok D, Verhoef PC (2006) The short-and long-term impact of an assortment reduction on category sales. Journal of Marketing Research 43(4):536–548

[80] 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

[81] Tkalcic M, Kunaver M, Košir A, Tasic J (2011) Addressing the new user problem with a personality based user similarity measure. In: First International Workshop on DEMRA 2011, p 106

[82] Tkalcic M, Ferwerda B, Hauger D, Schedl M (2015) Personality Correlates for Digital Concert Program Notes. UMAP pp 1–6

[83] Watson D, Clark LA (1997) Extraversion and its positive emotional core.

Personality Correlates for Digital Concert Program Notes

Published in: Proceedings of the 23rd Conference on User Modeling, Adaptation, and Personalization (UMAP 2015; Dublin, IE)

MARKO TKALCIC, Johannes Kepler University, Linz (AT)
BRUCE FERWERDA, Johannes Kepler University, Linz (AT)
DAVID HAUGER, Johannes Kepler University, Linz (AT)
MARKUS SCHEDL, Johannes Kepler University, Linz (AT)

Abstract

In classical music concerts, the concert program notes are distributed to the audience in order to provide background information on the composer, piece and performer. So far, these have been printed documents composed mostly of text. With some delay, mobile devices are making their way also in the world of classical concerts, hence offering additional options for digital program notes comprising not only text but also images, video and audio. Furthermore, these digital program notes can be personalized. In this paper, we present the results of a user study that relates personal characteristics (personality and background musical knowledge) to preferences for digital program notes.

Keywords: Classical Music, Digital Program Notes, Personality

1 Introduction

Classical music has long resisted the temptation of introducing new devices, such as smartphones or tablets into the live concert experiences. However, things started to change recently. Two efforts are being made in this direction, one within the FP7 Phenicx project[1] and the other by the Philadelphia Orchestra[2], both of them developing a mobile application for supporting the live classical concert experience. Both of these applications feature the program notes in a digital form. Within the Phenicx project we are developing a system that would support personalized digital program notes based on the user profile. The personalization is going to be done of the level of the preferred multimedia type in the notes (music, text or audio) and its length. In order to develop the personalization method we first need

[1] http://phenicx.upf.edu/
[2] https://www.philorch.org/introducing-livenote%E2%84%A2-nights#/

to identify the relationships between the users' personal characteristics (in terms of personality and musical background) and the preferences for the supporting multimedia material in the digital program notes. In this paper, we present the results of a user study that was carried out in order to identify these relationships. To be more concrete, we addressed the following research questions

1. Do the user's preferences for the digital program notes correlate with the user's personality
2. Do the user's preferences for the digital program notes correlate with the user's musical background

2 Related work

Personality and learning styles. The purpose of the program notes is to provide additional information about the concert in terms of the historical background on the piece, the composer and the performer. Similar educational processes, outside the classical music live experience domain, have been investigated in terms of personalization. Research based on learning styles, as a subset of personality, has been done [6] and models for personalizing course materials have been proposed [3].

Personality and user preferences. Personality has been gaining popularity in research on personalized services. It has been used to alleviate the cold-start problem in recommender systems [4, 15]. Research has been carried out to understand how personality relates to user preferences in multiple domains [1]. Special attention has been given to music, where strong relations have been found (i) between personality and genre preferences [14] (ii) between personality and the way people use music (e.g. rationally versus for mood regulation [2]) and (iii) between personality and the way users organize music [5]. However, to the best of our knowledge, there has been no research done on the relations between classical music and personality, except for the aforementioned genre preferences.

Implicit acquisition of personality. For a long time, the only way of assessing the user personality has been through extensive questionnaires, such as the Big-Five Inventory (BFI) [9]. There has been an increasing interest for detecting unobtrusively the user personality from various sources. Related work has been done especially on using social media (e.g. facebook or twitter) to infer the personality in the Five Factor Model (FFM) personality space [7, 13, 10]. These methods can be useful in scenarios when a user logs in an online system with her/his social media account.

3 Usage Scenario

Within the Phenicx project, we are developing technologies for improving the classical concert experience. One of the use cases we focus on is the *digital program notes* scenario, where *automatically generated, tailor-made multimedia editorials are offered to prospective audience members, introducing the concert that will take place with the purpose of informing, educating and pre-engaging users, supporting concert anticipation* [11]. The concrete scenario we foresee is a user (the attendee of a classical concert) having a mobile device (e.g. a smartphone or a tablet device) with a mobile application that supports the scenario. The user logs in the application with her/his social media account (e.g. twitter or facebook). Using some of the aforementioned methods for inferring the user's personality from the social media stream, the application can use this information to tailor the digital program notes to the user's needs. The strategy for personalizing the digital program notes is going to be based on the outcomes of the study we present in this paper.

4 Materials and Methods

We conducted a user study to collect the data needed. We gathered the participants through Amazon Mechanical Turk. In total we had 165 participants aged from 20 to 76 (mean 37, stdev 11), 93 females and 72 males. The participants first filled in a set of background questionnaires: a questionnaire about their musical background, a questionnaire about their musical preferences in general (according to 18 genres) and the Ten-items Personality Inventory (TIPI) questionnaire [8]. The participants were then shown various supporting multimedia content that could fit in an interactive digital program note. For the chosen classical music concert (the piece *La Mer* by the French composer *Claude Debussy* performed by the *Royal Concertgebouw Orchestra*) we prepared multimedia material (text, images and audio) for the composer, piece and performer. Furthermore, each pair (e.g. composer-text) was presented in a long and short version (for images and audio clips *short* meant one item and *long* meant several items). In total we presented the participants 14 combinations of content. For each content combination, the participants gave three ratings: (i) amount of consumption of the content (on the scale *none/some/half/most/all*), (ii) interestingness of the content (on a scale from 1 to 5) and (ii) novelty of the content (on a scale from 1 to 5).

5 Results

As can be seen from Tab. 1, personality traits were correlated with the subjects' musical background. Openness and extraversion were positively significantly correlated to three variables, agreeableness had a negative significant correlation with two variables while conscientiousness had a positive correlation only with the *attending classical concerts* variable. Neuroticism was not significantly correlated with any of the musical background variables.

personality trait	musical background variable	correlation	p value
openness	listening to classical music	0.26	0.0007
	studying an instrument	0.16	0.03
	attending classical concerts	0.19	0.01
extraversion	playing an instrument	0.21	0.008
	attending classical concerts	0.3	0.00009
	attending non-classical concerts	0.24	0.002
agreeableness	listening to non-classical music	-0.22	0.005
	studying an instrument	-0.23	0.003
conscientiousness	attending classical concerts	0.17	0.02

Table 1: Correlations between personality traits and musical background variables.

Personality also appears to be correlated with how much the subjects like classical music as a genre (see Tab. 2). Openness, extraversion and agreeableness have a positive significant correlation with liking classical music, while conscientiousness and neuroticism are not significantly correlated.

personality trait	correlation	p-value
openness	0.17	0.03
extraversion	0.21	0.006
agreeableness	0.33	0.00001
conscientiousness	0.11	0.16
neuroticism	-0.14	0.06

Table 2: Correlation between personality traits and the preferences for the classical music genre.

We observed several significant correlations among variables that can be

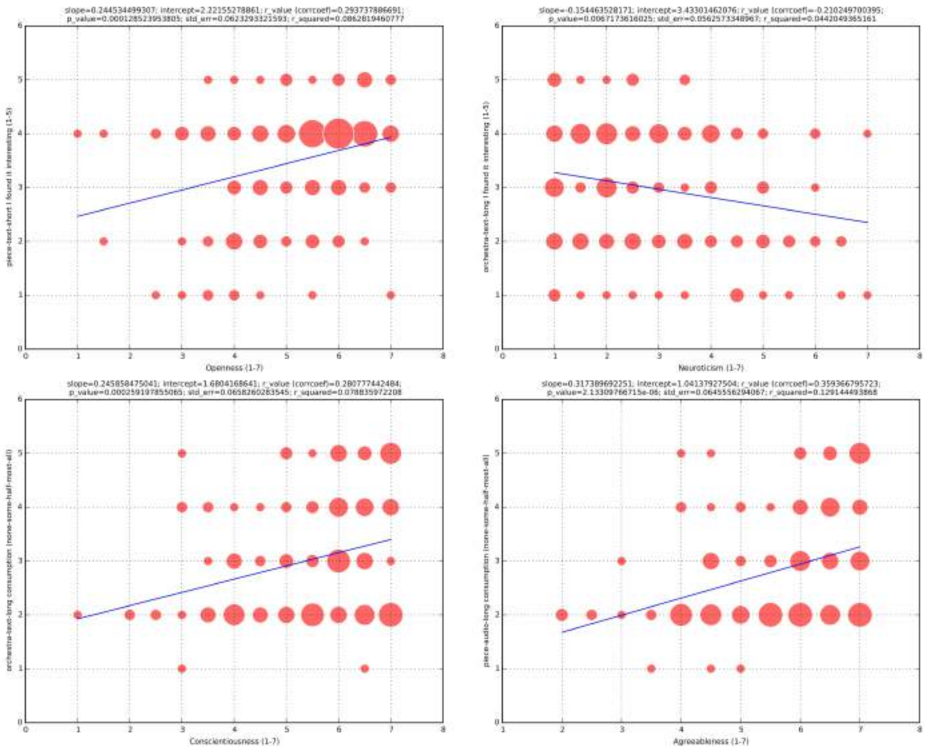

Figure 1: Bubble plots with correlations between personality and content preferences. In the presented four cases all independent variables (x axis) are on the scale from 1 to 7, while the dependent (y axis) are on the scale from 1 to 5. The top-left figure is the relation between openness and the preference for the short description of the piece (slope $b_1 = 0.24$, correlation $r = 0.29$, p-value $p = 0.0001$). The top-right figure is the relation between neuroticism and the preference for a long description about the orchestra ($b_0 = 0.15, r = -0.21, p = 0.007$). The bottom-left figure is the relation between conscientiousness and the degree of consumption of the long orchestra description ($b_0 = 0.25, r = 0.28, p = 0.003$). The bottom-right figure is the relation between agreeableness and the degree of consumption of the long audio ($b_0 = 0.32, r = 0.36, p = 2 \cdot 10^{-6}$)

.

useful in a personalized system. Personality did correlate with with the answers provided by the subjects on the consumption, interestingness and novelty of the multimedia material. For example, people who score high on openness, agreeableness, conscientiousness and extraversion tend to show a positive correlation with consumption, interestingness and novelty, hence meaning that they prefer to consume more of the material, find it interesting and novel. On the other hand, subjects who score high on neuroticism tend to consume less of the material, find it less interesting or novel. Our data show no correlation or negative correlations for neuroticism and the observed answers (see Fig. 1). Similarly, the music background variables showed generally positive correlations with consumption, interestingness and novelty of the multimedia material.

6 Conclusions and Future Work

The analysis of our data showed that there are many significant correlations between personal characteristics (personality and musical background) and preferences for the supporting multimedia content. However, further analyses should be carried out to determine how these personal characteristics relate with various types of material (text/images/audio) or length. The most important step we need to take is to use machine learning techniques to predict the preferred type of digital program notes based on the users' personalities. We plan to use logistic regression for individual binary variables and multinomial logistic regression for individual Likert scale ratings. Furthermore, we plan to use a rule-based algorithm for inferring the preferred modality and length.

We are currently in the process of designing a new user study where we will have better control over the text variance. The text will be curated by musicologists and we will use different levels of complexity (in terms of vocabulary). Furthermore, we will use the Goldsmiths Music Sophistication Index (MSI) [12] to assess the musical background of the users (the MSI instrument was published after we already carried out the study presented here).

Acknowledgement

This research is supported by the European FP7/2007-2013 programme through the PHENICX project under grant agreement no. 601166 and by the Austrian Science Fund (FWF): P25655.

References

[1] Cantador, I., Fernández-tobías, I., Bellogín, A.: Relating Personality Types with User Preferences in Multiple Entertainment Domains. EMPIRE 1st Workshop on "Emotions and Personality in Personalized Services", 10. June 2013, Rome (2013)

[2] Chamorro-Premuzic, T., Furnham, A.: Personality and music: can traits explain how people use music in everyday life? British journal of psychology (London, England : 1953) 98(Pt 2), 175–85 (May 2007), http://www.ncbi.nlm.nih.gov/pubmed/17456267

[3] El-Bishouty, M.M., Chang, T.W., Graf, S., Chen, N.S.: Smart e-course recommender based on learning styles. Journal of Computers in Education 1(1), 99–111 (Mar 2014), http://link.springer.com/10.1007/s40692-014-0003-0

[4] Elahi, M., Braunhofer, M., Ricci, F., Tkalcic, M.: Personality-based active learning for collaborative filtering recommender systems. AI*IA 2013: Advances in Artificial Intelligence pp. 360–371 (2013), http://link.springer.com/chapter/10.1007/978-3-319-03524-6_31

[5] Ferwerda, B., Yang, E., Schedl, M., Tkalčič, M.: Personality Traits Predict Music Taxonomy Preference. In: CHI 2015 Works-in-Progress (2015)

[6] Furnham, A.: The big five versus the big four: the relationship between the Myers-Briggs Type Indicator (MBTI) and NEO-PI five factor model of personality. Personality and Individual Differences 21(2), 303–307 (Aug 1996), http://linkinghub.elsevier.com/retrieve/pii/0191886996000335

[7] Golbeck, J., Robles, C., Edmondson, M., Turner, K.: Predicting Personality from Twitter. In: 2011 SocialCom and PASSAT conference Proceedings. pp. 149–156. IEEE (Oct 2011), http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6113107

[8] Gosling, S.D., Rentfrow, P.J., Swann, W.B.: A very brief measure of the Big-Five personality domains. Journal of Research in Personality 37(6), 504–528 (Dec 2003), http://linkinghub.elsevier.com/retrieve/pii/S0092656603000461

[9] John, O.O.P., Srivastava, S.: The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In: Pervin, L.A., John, O.P. (eds.) Handbook of personality: Theory and research, vol. 2, pp. 102–138. Guilford Press, New York, second edn. (1999), http://books.google.com/books?hl=en\&lr=\&id=b0yalwi1HDMC\&oi=fnd\&pg=PA102\&dq=The+Big-Five+Trait+Taxonomy:+History,+Measurement,+and+Theoretical+Perspectives\&ots=758AL9TvOn\&sig=EdcojL2Ze7jxAWySrhj_Jhbyr9ohttp://books.google.com/books?hl=en\&lr=\&id=b0

[10] Kosinski, M., Stillwell, D., Graepel, T.: Private traits and attributes are predictable from digital records of human behavior. Proceedings of the

National Academy of Sciences pp. 2–5 (Mar 2013), http://www.pnas.org/cgi/doi/10.1073/pnas.1218772110

[11] Liem, C.C.S., Sterren, R.V.D.: Innovating the Classical Music Experience in the PHENICX Project : Use Cases and Initial User Feedback. 1st International Workshop on Interactive Content Consumption (WS-ICC) at EuroITV 2013 (2013)

[12] Müllensiefen, D., Gingras, B., Musil, J., Stewart, L.: The musicality of non-musicians: an index for assessing musical sophistication in the general population. PloS one 9(2), e89642 (Jan 2014), http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3935919\&tool=pmcentrez\&rendertype=abstract

[13] Quercia, D., Kosinski, M., Stillwell, D., Crowcroft, J.: Our Twitter Profiles, Our Selves: Predicting Personality with Twitter. In: 2011 SocialCom and PASSAT conference Proceedings. pp. 180–185. IEEE (Oct 2011), http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6113111http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6113111

[14] Rentfrow, P.J., Gosling, S.D.: The do re mi's of everyday life: The structure and personality correlates of music preferences. Journal of Personality and Social Psychology 84(6), 1236–1256 (2003), http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-3514.84.6.1236

[15] Tkalcic, M., Kunaver, M., Košir, A., Tasic, J.: Addressing the new user problem with a personality based user similarity measure. Proceedings of the DEMRA 2011 and the UMMS 2011 Workshops, Girona, Spain (2011), http://ceur-ws.org/Vol-740/DEMRA_UMMS_2011_proceedings.pdf\#page=106

Exploring Music Diversity Needs Across Countries

Published in: Proceedings of the 24th Conference on User Modeling, Adaptation, and Personalization (UMAP 2016; Halifax, NS, CA)

BRUCE FERWERDA, Johannes Kepler University, Linz (AT)
ANDREU VALL, Johannes Kepler University, Linz (AT)
MARKO TKALCIC, Free University of Bolzano, Bolzano (IT)
MARKUS SCHEDL, Johannes Kepler University, Linz (AT)

Abstract

Providing diversity in recommendations has shown to positively influence the user's subjective evaluations such as satisfaction. However, it is often unknown how much diversity a recommendation set needs to consist of. In this work, we explored how music users of Last.fm apply diversity in their listening behavior. We analyzed a dataset with the music listening history of 53,309 Last.fm users capturing their *total* listening events until August 2014. We complemented this dataset with The Echo Nest features and Hofstede's cultural dimensions to explore how music diversity is applied across countries. Between 47 countries, we found distinct relationships between the cultural dimensions and music diversity variables. These results suggest that different country-based diversity measurements should be considered when applied to a recommendation set in order to maximize the user's subjective evaluations. The country-based relationships also provide opportunities for recommender systems to personalize experiences when user data is limited by being able to rely on the user's demographics.

Keywords: Music Recommendations, Diversity, Cultural Differences

1 Introduction

Providing recommendation diversity to users has become an important feature for recommender systems. Recommending items with high accuracy often result in a set of recommendations that are too similar to each other, and thereby not covering the full spectrum of the user's interest. Giving in on accuracy by introducing diversity can positively influence the user experience, such as user satisfaction [1].

The amount of diversity that should be provided remains a debatable topic. Prior research has identified that individual characteristics (e.g., expertise) play a role in how much diversity is desired by the user [1]. The problem that persist is that often self-report measures are used to identify these influential individual characteristics. Whereas registration and login processes are becoming easier (e.g., single sign-on buttons), asking additional questions may become a bothersome process for (new) users.

The implicit acquisition (i.e., without the use of questionnaires) of individual differences remains challenging. Especially for new users there is not enough behavioral data yet to make inferences. Country information may be a useful alternative as it already consists in a standard user profile and behavior has shown te be culturally embedded. We looked at music listening behavior of Last.fm [1] users from 47 countries and explored how they apply diversity to their playlists. We computed a diversity measure based on the unique listening events (i.e., of artists and genres) of users per country. By using The Echo Nest features we were also able to look at artist characteristics (i.e., how known, familiar, or popular the artists are that have been listened to by users). This provide insights on cultural dependent diversity patterns.

2 Related Work

Recommender systems intend to create a personalized set of items that are most relevant to the user. However, highly relevant items often appear too similar to each other. A set of items showing too much similarities (e.g., highly relevant items) can cause choice overload [9]. In [1] was shown that diversity has a positive effect on the attractiveness of the recommendation set, the difficulty to make a choice, and eventually on the choice satisfaction. Additionally, individual differences were found. E.g., increased expertise has positive effects on perceived item variety and attractiveness.

Besides individual characteristics, research has shown that culture consists of useful cues as well. General behavior and preferences have shown to be rooted and embodied in culture [8]; looking at behavior on a country level may provide useful information for the desired recommendation diversity.

[1] http://www.last.fm

Table 1: Correlation results.

	PDI	IDV	MAS	UAI	LTO	IND
Artist	.279*	-.373**	.155	-.020	-.259*	.080
Genre	.329*	-.265*	.074	-.051	-.108	-.113
Hot.	-.131	-.0.39	-.135	-.359*	-.641**	.557**
Fam.	-.100	-.229	.009	-.255	-.677**	.520**
Disc.	-.367**	-.294*	-.311*	-.366*	-.274*	.517**

Note. *p<.05, **p<.01

3 Method

A Last.fm dataset was used with 53,309 users of 47 countries and their *total* listening history until August 2014. [2] The dataset consists of users' listening (i.e., user ID, time-stamp, artist name, and track name) and profile information (i.e., gender, age, country). A diversity measure was created by aggregating each user's listening history by artist and genre to identify the unique instances of each respectively. [3] E.g., a history of 90 events originating from two artists/genres means a diversity of two. Each diversity measure was normalized ($r \epsilon$ [0,1]) due to the unequal number of users between countries. The dataset was complemented with Hofstede's cultural dimensions (i.e., power distance, individualism, masculinity, uncertainty avoidance, long-term orientation, and indulgence [7]) and The Echo Nest features (i.e., genre, hotness, familiarity, and discovery).

4 Results

Correlation analyses were used to explore the relationship between cultural dimensions and diversity, and report Pearson's correlation ($r \epsilon$ [-1,1]) to indicate the relationship.

Table 1 shows the correlation results between the cultural dimensions and listening characteristics. A negative correlation represents the degree of diversity, whereas a positive correlation indicates homogeneity. Positive correlations were found between a culture's power distance and artist (r=.279, p=.043) and genre (r=.329, p=.027) diversity. This indicates that users in countries scoring high on this dimension tend to apply less diversity by artist as well as by genre. Negative correlations were found between the

[2]Available at `http://www.cp.jku.at/datasets/LFM-1b/`

[3]Genre was obtained through The Echo Nest. To maintain a manageable dataset we decided not to focus on a track level.

individualism dimensions and artist (r=-.373, p=.012) and genre (r=-.265, p=.048) diversity, which indicates that users in individualistic countries tend to apply music diversity on an artist as well as genre level. Finally, a negative correlation was found between long-term orientation and artist diversity (r=-.259, p=.048). Users in countries scoring high on this dimension tend to listen to more diverse artists.

The Echo Nest features allowed us to gain more insights of the diversity characteristics of the artists. A negative correlation indicates that users of a country involve artists that score low on the respective feature of The Echo Nest when scoring high in the correlated cultural dimension (Table 1).

Hotness. Hotness was found to be negatively correlated with uncertainty avoidance (r=-.359, p=.015) and long-term orientation (r=-.641, p=.000), while a positive correlation was found with indulgence (r=.557, p=.000).

Familiarity. Familiarity was found to be negatively correlated with long-term orientation (r=-.677, p=.000), but positively correlated with indulgence (r=.520, p=.000).

Discovery. Discovery showed a negative correlation with five out of six cultural dimensions: power distance (r=-.367, p=.013), individualism (r=-.294, p=.050), masculinity (r=-.311, p=.038), uncertainty avoidance (r=-.366, p=.013), and long-term orientation (r=-.274, p=.042). A positive correlation was found with indulgence (r=.517, p=.000).

5 Conclusion & Implications

We show with our results that different diversity patterns exist and that they are related to cultural dimensions. When looking at the relationship between listening characteristics and cultural dimensions, distinct correlations were found.

Approaching diversity on a country level enables the creation of proxy measures for personalization when data is limited. Users' personality has gained interest to make inferences for personalization (e.g., [2, 6, 11]) on an individual level. The interconnectedness of applications and social media may be exploited to implicitly acquire personality (e.g., Facebook [4], Twitter [10], Instagram [3, 5]). However, a social media connection is still needed. Our results could be used to infer diversity needs based on country information, which is often available through the user's account.

6 Acknowledgment

This research is supported by the Austrian Science Fund (FWF): P25655.

References

[1] Bollen, D., Knijnenburg, B.P., Willemsen, M.C., Graus, M.: Understanding choice overload in recommender systems. In: RecSys. ACM (2010)

[2] Ferwerda, B., Schedl, M.: Enhancing music recommender systems with personality information and emotional states: A proposal. In: EMPIRE (2014)

[3] Ferwerda, B., Schedl, M., Tkalcic, M.: Predicting personality traits with instagram pictures. In: 3rd Workshop on EMPIRE (2015)

[4] Ferwerda, B., Schedl, M., Tkalcic, M.: Personality traits and the relationship with (non-) disclosure behavior on facebook. WWW (2016)

[5] Ferwerda, B., Schedl, M., Tkalcic, M.: Using instagram picture features to predict users' personality. In: MultiMedia Modeling (2016)

[6] Ferwerda, B., Yang, E., Schedl, M., Tkalcic, M.: Personality traits predict music taxonomy preferences. In: CHI'15 Ext. Abstracts (2015)

[7] Hofstede, G., Hofstede, G.J., Minkov, M.: Cultures and organizations: Software of the mind (1991)

[8] Kitayama, S., Park, H.: Cultural shaping of self, emotion, and well-being: How does it work? Soc. and Pers. Psych. Compass (2007)

[9] Scheibehenne, B., Greifeneder, R., Todd, P.M.: Can there ever be too many options? a meta-analytic review of choice overload. J. of Cons. Res. (2010)

[10] Skowron, M., Ferwerda, B., Tkalcic, M., Schedl, M.: Fusing social media cues: Personality prediction from twitter and instagram. WWW (2016)

[11] Tkalcic, M., Ferwerda, B., Hauger, D., Schedl, M.: Personality Correlates for Digital Concert Program Notes. UMAP pp. 1–6 (2015)

Investigating the Relationship Between Diversity in Music Consumption Behavior and Cultural Dimensions: A Cross-Country Analysis

Published in: Extended Proceedings of the 24th Conference on User Modeling, Adaptation, and Personalization (UMAP 2016): 1st Workshop on Surprise, Opposition, and Obstruction in Adaptive and Personalized Systems (SOAP 2016; Halifax, NS, CA)

BRUCE FERWERDA, Johannes Kepler University, Linz (AT)
MARKUS SCHEDL, Johannes Kepler University, Linz (AT)

Abstract

Diversity in recommendation lists or sets has shown to be an important feature in recommender systems as it can counteract on negative effects such as choice difficulty and choice overload. However, how much diversity a recommendation list needs to provide is not clearly defined. By analyzing music listening behavior of listeners in 47 countries, we show that diversity needs may be cultural dependent. For our analyses, we exploited a Last.fm dataset containing almost 1.1 billion single listening events. We investigated several diversity measures to identify how users in different countries apply music diversity to their listening behavior. By analyzing $53{,}309$ Last.fm users, we found distinct diversity behavior related to several cultural dimensions of Hofstede. We show with our results that different diversity needs exist between cultures, and should be taken into account when applying diversity to a recommendation list.

Keywords: Diversity, Cultural Differences, Music Recommender Systems

1 Introduction

By tradition, recommender systems are created to most accurately provide recommendations in line with the user's taste (i.e., output options with the highest predicted ratings). The assumption of this approach is that the higher the recommendation accuracy, the higher the attractiveness of the items for the user. However, it has been shown that by doing this two subsequent effects may occur, which are caused by recommendations that are too attractive: (i) choice difficulties [23], and (ii) choice overload [2]. One way to counteract on the negative effects of too attractive items is to introduce recommendation diversity.

The amount of diversity that a set or list of recommendations should provide has been given limited attention. Prior research has shown that personal characteristics, such as expertise, play a role in the desired amount of recommendation diversity [2]. As the shaping of behavior and preferences has shown to be influenced by culture [13], identifying diversity on a country level may already provide cues about the desired recommendation diversity.

Providing a truly personalized experience to the user is still challenging in today's recommender systems. Often there is simply not enough data available (yet) about the user. A way to solve this problem is to use questionnaires in order to get to know the user. However, this is not desirable since it is obtrusive, takes a lot of effort and time from the user, and thereby disrupts their interaction with the system. Since country information is often available through the user's profile information, identifying diversity needs on a country level could be exploited to provide users with a personalized experience. Quantifying these diversity needs and studying their relationship to cultural dimensions is the focus of the study at hand.

This study is a follow-up investigation of the one presented in [8]. A shortcoming of that study was the quite simple definition of diversity, solely based on absolute numbers of whether an artist is listened to or not, neglecting the frequency the artist is listened to. Furthermore, [8] uses genres taken from the Echonest,[1] which are very broad, thus rendering impossible fine-grained modeling and analysis. Addressing these two shortcomings, the paper at hand (i) investigates several volume- and entropy-based diversity formulations and (ii) models a population's diversity via a dictionary containing more than $2{,}000$ genre names.

2 Related Work

Recommender systems intend to create a personalized set of items that are most relevant to the user. However, highly relevant items often appear too similar to each other, resulting in recommendations that may be perceived as boring to the users. A set of items showing too much similarity (e.g., too many highly relevant items) can, in turn, cause choice overload [19]. Bollen et al. [2] and Willemsen et al. [23] investigated the influence of diversity on movie recommendations and found that diversity has a positive effect on the attractiveness of the recommendation set, the difficulty to make

[1] http://the.echonest.com

a choice, and eventually on the choice satisfaction. Bollen et al. [2] additionally identified individual differences. For example, they showed that increased expertise has positive effects on perceived item variety and attractiveness.

Besides individual user characteristics, research has shown that cultural aspects can provide useful cues too. General behavior and preferences have shown to be rooted and embodied in culture [13], hence looking at behavior on a country level may provide useful information for the desired recommendation diversity. In a comprehensive study, Hofstede et al. [11] describe national cultures among six dimensions: power distance, individualism, masculinity, uncertainty avoidance, long-term orientation, and indulgence. These dimensions describe the effects of a society's culture on the values and behaviors of its members, which we use to explain diversity differences between countries in this study. The data that describes Hofstede's cultural dimensions has been collected since 1967 and is being refined ever since.

3 Method

The study at hand investigates on a standardized corpus of music listening events and listener demographics on the one hand, and Hofstede's cultural dimensions on the other, the relationship between diversity of music tastes (on the genre level) and cultural aspects in a population (on the country level). In the following, we first provide details of the used dataset, its enrichment by genre terms, and the definitions for diversity that we study. Subsequently, we summarize Hofstede's cultural dimensions, against which we compare our diversity scores via correlation analysis.

3.1 Music Dataset

We used the LFM-1b dataset [18] to model diversity and perform our experiments.[2] It is a dataset containing almost 1.1 billion single listening events by more than $120{,}000$ users of Last.fm[3] and covers over 3 million unique artists. Due to its user-generated nature, however, the data is quite noisy, e.g., metadata frequently contains typos. We therefore had to perform simple data cleaning first. Assuming that wrong artist names, which can be the result of misspellings, typos, hacking, and vandalism, etc., do not frequently occur in the dataset, we discarded all artist names that occur in the

[2] `http://www.cp.jku.at/datasets/LFM-1b`
[3] `http://www.last.fm`

listening events of less than 10 users. This cleaning resulted in a dataset of $585{,}095$ artists.

3.2 Modeling Diversity

Geographically, we model diversity on the country level. In order to obtain relevant results, we only consider countries with at least 100 users in the LFM-1b dataset. For detailed numbers of users, please consider Table 2.

In terms of listening behavior, we gauge diversity via scores derived from genre data of users' listening events. To this end, we first retrieve all artists' top tags provided by Last.fm via their API.[4] The resulting tags obviously contain many terms other than genre names, for which reason we index them by a dictionary of $1{,}998$ genre and style terms extracted from Freebase.[5] We further restrict the considered tags to those that have a tag weight of at least 10 according to Last.fm's weighting scheme.[6] This eventually provides us with a set of genre tags for each artist. Statistics of the 50 most frequently occurring genres for selected countries of the dataset are provided in Table 1 for the U.S.A., Japan, and Finland. As can be seen, while there are quite a few genres that are popular among Last.fm users in all three countries (e.g., Rock, Alternative, and Metal), country-specific differences are evidenced too. For instance, J-Pop is a genre very popular in Japan, but not among the top 50 genres in any of the other countries listed here. In contrast, 3 out of the top 10 most popular genres in Finland relate to Metal.

[4] We use the API endpoint `http://www.last.fm/api/show/artist.getTopTags`.

[5] `http://www.freebase.org`

[6] Last.fm employs a non-disclosed approach to weight artist tags based on the number of users who assign the tag to the artist. While details are not provided, these weights are normalized to [0,100]. Our filtering thus discards tags infrequently used to describe the artist under consideration.

Table 1: Relative amount of listening events (playcounts PC in percent) of the 50 most frequent genres and styles for the U.S.A., Japan, and Finland.

U.S.A.		Japan		Finland	
Genre tag	PC	Genre tag	PC	Genre tag	PC
Rock	12.51	Rock	16.01	Rock	11.31
Alternative	9.63	Alternative	8.37	Metal	11.15
Alternative rock	5.86	J-pop	5.77	Alternative	7.30
Metal	4.77	Pop	4.56	Alternative rock	4.56
Pop	3.62	Metal	4.55	Hard rock	4.28
Indie	3.59	Alternative rock	4.26	Heavy metal	3.44
Hard rock	3.12	Indie	3.63	Death metal	2.74
Indie rock	3.09	Electronic	2.29	Classic rock	2.61
Classic rock	2.92	Hard rock	2.24	Pop	2.21
Electronic	2.33	Classic rock	2.23	Indie	2.13
Dance	2.21	Visual Kei	2.03	Electronic	2.00
Psychedelic	1.84	Indie rock	2.02	Indie rock	1.75
Blues	1.77	Heavy metal	1.68	Dance	1.71
Hip-Hop	1.72	Dance	1.66	Progressive rock	1.67
Punk	1.61	Punk	1.53	Nu metal	1.57
Heavy metal	1.49	Psychedelic	1.45	Progressive	1.50
Singer-songwriter	1.34	Anime	1.43	Power metal	1.46
Progressive	1.25	Electronica	1.43	Punk	1.45
Electronica	1.24	Blues	1.18	Alternative metal	1.32
Progressive rock	1.16	Japanese rock	1.17	Psychedelic	1.18
New Wave	1.08	Progressive rock	1.06	Hip-Hop	1.10
Punk rock	1.03	Pop punk	0.91	Electronica	0.90
Nu metal	0.99	Nu metal	0.86	Speed metal	0.89
Alternative metal	0.85	Progressive	0.86	Blues	0.84
Rap	0.83	New Wave	0.84	Punk rock	0.82
Post-punk	0.79	Punk rock	0.83	Viking metal	0.75
Synthpop	0.77	Singer-songwriter	0.75	Progressive metal	0.71
Pop punk	0.77	Death metal	0.75	New Wave	0.70
Rnb	0.75	Synthpop	0.67	Melodic death metal	0.69
Psychedelic rock	0.72	Hip-Hop	0.63	Thrash	0.68
Emo	0.68	Experimental	0.59	Visual Kei	0.66
Experimental	0.68	Jazz	0.59	Groove metal	0.65
Death metal	0.68	Ambient	0.59	Pop punk	0.64
Electro	0.67	Power metal	0.58	Psychedelic rock	0.64
Garage rock	0.67	Electropop	0.57	Hardcore	0.62
Blues-rock	0.66	Electro	0.52	Thrash metal	0.62
House	0.64	Alternative metal	0.52	Industrial	0.60
Techno	0.62	Post-punk	0.51	Singer-songwriter	0.59
Ambient	0.60	Speed metal	0.50	Ambient	0.58
Glam rock	0.57	Pop rock	0.47	Experimental	0.53
Folk	0.53	Instrumental	0.47	Synthpop	0.50
Indie pop	0.52	Emo	0.46	Glam rock	0.49
Art rock	0.41	House	0.44	Emo	0.49
Hardcore	0.41	Blues-rock	0.43	Symphonic metal	0.48
Funk	0.40	Funk	0.42	Metalcore	0.46
Instrumental	0.40	Glam rock	0.41	Instrumental	0.46
Speed metal	0.39	Techno	0.39	Electro	0.44
Soul	0.37	Hardcore	0.38	Technical death metal	0.44
Folk rock	0.37	Fusion	0.38	Rapcore	0.43
Industrial rock	0.36	Soul	0.38	Blues-rock	0.42

Based on the users' demographics, as provided in the LFM-1b dataset, and the artist-related genre information, obtained as described above, we define the following volume- and entropy-based diversity measures, computed per country and reported in Table 2:

Table 2: Diversity scores for the top 47 countries. The columns indicate: country, total number of users, Hofstede's cultural dimensions ($r \in$ [0,100]): power distance index(PDI), individualism (IDV), masculinity (MAS), uncertainty avoidance index (UAI), long-term orientation (LTO), and indulgence (IND), absolute volume of unique genre occurrences (*Vol. abs.*), relative volume of unique genre occurrences (*Vol. rel.*), absolute (*Vol.* $> 1‰$ *abs.*) and relative (*Vol.* $> 1‰$ *rel.*) volume of genre occurrences with relative listening volumes exceeding one per mille, genre entropy, mean (*Vol.* μ) and standard deviation (*Vol.* σ) of listening distributions over genres.

Country	# User	PDI	IDV	MAS	UAI	LTO	IND	Vol. abs.	Vol.rel. (%)	Vol. $>$ $1‰$abs.	Vol. $> 1‰$rel. (%)	Entropy	Vol. μ	Vol. σ
U.S.A.	10255	40	91	62	46	26	68	1111	55.55	132	6.6	0.647383	0.0009	.004580
Russia	5024	93	39	36	95	81	20	1097	54.85	141	7.05	0.665395	0.000912	.004312
Germany	4578	35	67	66	65	83	40	1100	55	138	6.9	0.662084	0.000909	.004346
Great Britain	4534	35	89	66	35	51	69	1103	55.15	132	6.6	0.64291	0.000907	.004702
Poland	4408	68	60	64	93	38	29	1077	53.85	132	6.6	0.647125	0.000929	.004696
Brazil	3886	69	38	49	76	44	59	1053	52.65	119	5.95	0.626137	0.00095	.005271
Finland	1409	33	63	26	59	38	57	1042	52.1	131	6.55	0.656833	0.00096	.004694
Netherlands	1375	38	80	14	53	67	68	1081	54.05	142	7.1	0.658458	0.000925	.004473
Spain	1243	57	51	42	86	48	44	1043	52.15	136	6.8	0.657332	0.000959	.004723
Sweden	1231	31	71	5	29	53	78	1062	53.1	124	6.2	0.649503	0.000942	.004678
Ukraine	1143	N/A	N/A	N/A	N/A	86	14	1029	51.45	139	6.95	0.665125	0.000972	.004543
Canada	1077	39	80	52	48	36	68	1056	52.8	132	6.6	0.65296	0.000947	.004637
France	1055	68	71	43	86	63	48	1045	52.25	140	7	0.66765	0.000957	.004357
Australia	976	38	90	61	51	21	71	1036	51.8	125	6.25	0.64334	0.000965	.004912
Italy	974	50	76	70	75	61	30	1031	51.55	120	6	0.645742	0.00097	.004942
Japan	806	54	46	95	92	88	42	1024	51.2	126	6.3	0.648062	0.000977	.004929
Norway	750	31	69	8	50	35	55	1028	51.4	129	6.45	0.657356	0.000973	.004700
Mexico	705	81	30	69	82	24	97	1011	50.55	137	6.85	0.655207	0.000989	.004930
Czech Republic	632	57	58	57	74	70	29	983	49.15	133	6.65	0.6687	0.001017	.004593
Belarus	558	N/A	N/A	N/A	N/A	81	15	979	48.95	140	7	0.672649	0.001021	.004543
Belgium	513	65	75	54	94	82	57	1008	50.4	142	7.1	0.66945	0.000992	.004547
Indonesia	484	78	14	46	48	62	38	842	42.1	118	5.9	0.644635	0.001188	.005790
Turkey	479	66	37	45	85	46	49	980	49	119	5.95	0.654673	0.00102	.004732
Chile	425	63	23	28	86	31	68	918	45.9	127	6.35	0.653122	0.001089	.005312
Croatia	372	73	33	40	80	58	33	940	47	129	6.45	0.665861	0.001064	.004904
Portugal	291	63	27	31	104	28	33	918	45.9	136	6.8	0.664801	0.001089	.005023
Argentina	282	49	46	56	86	20	62	927	46.35	119	5.95	0.639404	0.001079	.005586
Switzerland	277	34	68	70	58	74	66	970	48.5	132	6.6	0.66451	0.001031	.004768
Austria	276	11	55	79	70	60	63	932	46.6	140	7	0.671787	0.001073	.004804
Denmark	272	18	74	16	23	35	70	950	47.5	136	6.8	0.664297	0.001053	.004858
Hungary	272	46	80	88	82	58	31	901	45.05	137	6.85	0.687505	0.00111	.004544
Serbia	253	86	25	43	92	52	28	910	45.5	141	7.05	0.677889	0.001099	.004746
Romania	237	90	30	42	90	52	20	951	47.55	137	6.85	0.676884	0.001052	.004409
Bulgeria	236	70	30	40	85	69	16	926	46.3	143	7.15	0.681036	0.00108	.004766
Ireland	220	28	70	68	35	24	65	906	45.3	125	6.25	0.652082	0.001104	.005270
Lithuania	202	42	60	19	65	82	16	892	44.6	138	6.9	0.672913	0.001121	.004969
Slovakia	192	104	52	110	51	77	28	878	43.9	136	6.8	0.684491	0.001139	.004614
Greece	175	60	35	57	112	45	50	907	45.35	134	6.7	0.688293	0.001103	.004447
Latvia	165	44	70	9	63	69	13	904	45.2	134	6.7	0.675491	0.001106	.004787
New Zealand	164	22	79	58	49	33	75	865	43.25	134	6.7	0.672034	0.001156	.005161
China	162	80	20	66	30	87	24	847	42.35	129	6.45	0.671203	0.001181	.004991
Columbia	159	67	13	64	80	13	83	885	44.25	123	6.15	0.65439	0.00113	.005477
Iran	135	58	41	43	59	14	40	782	39.1	117	5.85	0.65695	0.001279	.005473
India	122	77	48	56	40	51	26	794	39.7	127	6.35	0.665461	0.001259	.005578
Venezuela	118	81	12	73	76	16	100	816	40.8	123	6.15	0.654646	0.001225	.005830
Estonia	107	40	60	30	60	82	16	823	41.15	125	6.25	0.672622	0.001215	.005148
Israel	100	13	54	47	81	38	N/A	830	41.5	133	6.65	0.674123	0.001205	.005378

Overall volume of genre occurrences We count the number of genre tags that appear at least once in at least one user's listening history of the respective country's user base and define it as the absolute volume of genre occurrences (indicated as *Vol. abs.* in Table 2). The relative volume is computed as the fraction of the absolute one and the number of genre tags in the dictionary (*Vol. rel.* in Table 2).

Relative listening volume exceeding one per mille We first compute the total playcounts, i.e. number of listening events of each artist over all users in the country under investigation. Based on the artist–genre mapping, we subsequently calculate these playcounts per genre by aggregating the playcounts of all artists that are tagged by that genre. This absolute genre playcount is then normalized by the total playcount of a country, yielding an estimate of genre g's relative popularity in country c. Formally, the computation of this relative popularity $pop_c(g)$ is given in Equation 1, where G is the set of genres, U_c is the set of users in country c, A_c^g is the set of artists listened to in country c and tagged as genre g, and $pc(u, a)$ denotes the number of listening events (playcounts) of user u to artist a.[7]

$$pop_c(g) = \frac{\sum_{u \in U_c} \sum_{a \in A_c^g} pc(u, a)}{\sum_{g \in G} \sum_{u \in U_c} \sum_{a \in A_c^g} pc(u, a)} \tag{1}$$

To define diversity, we finally count the number of genres whose relative popularity exceeds one per mille of the total listening events. Again, we use this score as absolute measure and we divide it by the number of genre tags to yield a relative estimate (*Vol.* $> 1‰$ *abs.* and *Vol.* $> 1‰$ *rel.* in Table 2).

Entropy Based on the genre-specific playcounts, computed as described in the previous paragraph, we use the normalized genre entropy as diversity measure. Formally, our adapted entropy measure is defined in Equation 2, where G is the set of all genres and $p_c(g)$ is the probability for genre g in country c. We approximate this probability as the relative frequency of genre g's playcounts among all playcounts in country c. The normalization term in the denominator ensures that the resulting diversity scores fall into the range [0,1].

$$H_c(G) = \frac{-\sum_{g \in G} p_c(g) \cdot \log_2 p_c(g)}{\log_2 |G|} \tag{2}$$

[7]We use the term le instead of pc in the formula to avoid confusions with p_c in Equation 2.

Statistics over relative genre playcounts In addition to the volume-based diversity measures and to entropy, we investigate basic statistics of the relative genre playcounts. In particular, we compute mean and standard deviation of the elements $p_c(g)$ with $g \in G$. The corresponding scores are denoted *Vol.* μ and *Vol.* σ, respectively, in Table 2.

3.3 Modeling Cultural Dimensions

The most comprehensive framework for national cultures is considered to be Hofstede et al.'s cultural dimensions. They defined six dimensions to identify cultures [11]:

Power distance index (PDI) Defines the extent to which power is distributed unequally by less powerful members of institutions (e.g., family). High power distance indicates that a hierarchy is clearly established and executed in society. Low power distance indicates that authority is questioned and attempted to distribute power equally.

Individualism (IDV) Defines the degree of integration of people into societal groups. High individualism is defined by loose social ties. The main emphasis is on the "I" instead of the "we," while opposite for low individualistic cultures.

Masculinity (MAS) Defines a society's preference for achievement, heroism, assertiveness and material rewards for success (countries scoring high in this dimension). Whereas low masculinity represents a preference for cooperation, modesty, caring for the weak and quality of life.

Uncertainty avoidance index (UAI) Defines a society's tolerance for ambiguity. High scoring countries in this scale are more inclined to opt for stiff codes of behavior, guidelines, laws. Whereas more acceptance of differing thoughts and/or ideas are accepting for those scoring low in this dimension.

Long-term orientation (LTO) Is associated with the connection of the past with the current and future actions and/or challenges. Lower scoring countries tend to believe that traditions are honored and kept, and value steadfastness. High scoring countries believe more that adaptation and circumstantial, pragmatic problem-solving are necessary.

Table 3: Spearman correlations between Hofstede's cultural dimensions and the analyzed diversity measures. Abbreviations for diversity measures as in Table 1. Results for absolute and relative diversity measures are obviously the same and therefore reported only once. Note: *p<.05, **p<.01

	Vol. abs./rel.	Vol. $> 1‰$ abs./rel.	Entropy	Vol. μ	Vol. σ
Power Distance	-.183	.036	.132	.183	.022
Individualism	.459**	-.167	-.117	-.459**	-.414**
Masculinity	-.073	-.115	-.133	.073	.088
Uncertainty Avoidance	.057	.301*	.218	-.057	-.174
Long-Term Orientation	.106	.443**	.442**	-.106	-.442**
Indulgence	.217	-.300**	-.558**	-.217	.225

Indulgence (IND) Defines in general the happiness of a country. Countries scoring high in this dimension are related to a society that allows relatively free gratification of basic and natural human desires related to enjoying life and having fun (e.g., be in control of their own life and emotions). Whereas low scoring countries show more controlled gratification of needs and regulate it by means of strict social norms

4 Experiments and Results

A correlation analysis was performed to indicate the relationship between Hofstede's cultural dimensions and the diversity measurements, cf. Table 3. Spearman correlation ($r \in$ [-1,1]) is reported as the correlation coefficient to indicate the strength of the relationship. Statistically significant results at a level of $p < 0.05$ and $p < 0.01$ are denoted by * and **, respectively.

Investigating the table, we find moderate, highly significant correlations between the cultural dimension of *individualism* and the volume-based diversity measures. This correlation is positive for absolute volume (*Vol. abs.*) and negative for the mean and standard deviation of the volume measures that take actual playcount values into account (respectively, *Vol.* μ and *Vol.* σ). It seems reasonable that listeners from cultures in which individualism is important show higher diversity in terms of numbers of distinct genres they listen to. The negative correlation to the playcount-based volume measures signifies that these listeners do not listen only to a few genres very intensely (which would result in a higher *Vol.* μ and *Vol.* σ value), but instead spread their music listening time slightly more evenly over various genres (overall, resulting in lower *Vol.* μ and *Vol.* σ scores). Interestingly, the volume measure that restricts results by the per mille threshold does not show significant correlations to individualism. This is presumably due to the lower number of genres and styles considered in this case that does not account for a high enough amount of individualism.

For *long-term orientation*, we identify moderate positive, highly significant correlations with both volume- and entropy-based diversity measures. This can be explained by the reasonable assumption that cultures scoring high on aspects like flexibility, adaptation, and pragmatic problem-solving (according to the definition of long-term orientation) are more likely to listen to more diverse music, both in terms of unique genres listened to and entropy in their music distribution over genres. These countries' listening events are also more evenly spread over a variety of genres (lower *Vol.* σ scores).

For the cultural dimension of *indulgence*, we observe quite interesting and maybe surprising results. In fact, this dimension is highly significantly, negatively correlated to volume- and entropy-based diversity measures, in particular to the latter. Therefore, citizens of countries scoring high on indulgence, which means they tend to enjoy life and have a lot of fun, exhibit a smaller need for music diversity. This could, to some extent, be explained by a focus on certain genres that are commonly regarded as positive and happy, e.g., Pop, while avoiding music from dark genres, such as Death Metal.

The aspect of *uncertainty avoidance* is only slightly correlated to the relative volume diversity. *Power distance* and *masculinity* do not show significant correlation to any of the diversity measures.

5 Conclusions and Future Work

In the presented study, we found distinct correlations between volume- and entropy-based music diversity measures and Hofstede's cultural dimensions, which showed to be in line with, and extended, prior results reported in [8]. We identified moderate, highly significant correlations for the aspect of individualism and volume-based diversity measures. Highly individualist societies thus listen to more diverse genres. For long-term orientation and indulgence, we also found moderate, highly significant correlations; in these cases not only for volume-based, but also for entropy-based diversity measures. For long-term orientation, this means that countries whose population can be characterized as flexible, pragmatic, and eager to adapt to changes show a higher level of diversity in their music consumption behavior. Populations characterized by high indulgence (happiness and enjoying life) in contrast show a significantly lower desire for music diversity.

Approaching diversity on a country level enables the creation of proxy measures for personalization in situations where data is limited. For example, the new user problem in recommender systems when there is not enough behavioral data yet to make personalization inferences with. To address this problem, users' personality, among other aspects, has attracted interest to make inferences for personalization [4, 9, 21]. One way to extract personality is facilitated by the increasing connectedness of applications and social media (e.g., single sign-on buttons). This allows exploitation of social media data for personality acquisition, for instance, from Facebook [1, 3, 5, 15], Twitter [10, 17], or Instagram [6, 7, 20]. However, a connection with the user's social media account is still needed. Our results could be used to make inferences about the user's diversity needs based solely on their country, which is often available through the user's account information.

Future work will investigate diversity formulations that also take into account similarities between genres. In particular when using fine-grained genre terms, an approach based on extending the one presented in [16] may yield additional interesting findings. In addition, taking into consideration similarities and affinities between countries, e.g., via Wikipedia articles [14], may allow for a more decent modeling of culture. In this study we only focused on Hofestede's cultural dimensions. However, although less comprehensive, there are other cultural dimensions (e.g., GLOBE [12] and Trompenaar's [22] cultural dimensions) available. It would be nice to investigate the consistency between the different cultural dimensions in the future.

6 Acknowledgments

This research is supported by the Austrian Science Fund (FWF): P25655.

References

[1] M. D. Back, J. M. Stopfer, S. Vazire, S. Gaddis, S. C. Schmukle, B. Egloff, and S. D. Gosling. Facebook profiles reflect actual personality, not self-idealization. *Psychological Science*, 21:372–374, 2010.

[2] D. Bollen, B. P. Knijnenburg, M. C. Willemsen, and M. Graus. Understanding choice overload in recommender systems. In *Proceedings of the 4th ACM Conference on Recommender Systems*, pages 63–70, Barcelona, Spain, September 2010.

[3] F. Celli, E. Bruni, and B. Lepri. Automatic personality and interaction style recognition from facebook profile pictures. In *Proceedings of the*

ACM International Conference on Multimedia, pages 1101–1104, Orlando, Florida, USA, November 2014.

[4] B. Ferwerda and M. Schedl. Enhancing Music Recommender Systems with Personality Information and Emotional States: A Proposal. In *Proceedings of the 2nd Workshop on Emotions and Personality in Personalized Services (EMPIRE)*, Aalborg, Denmark, July 2014.

[5] B. Ferwerda, M. Schedl, and M. Tkalcic. Personality traits and the relationship with (non-) disclosure behavior on facebook. In *Proceedings of the 25th International World Wide Web Conference Companion (WWW)*, Montreal, Canada, April 2016.

[6] B. Ferwerda, M. Schedl, and M. Tkalčič. Predicting Personality Traits with Instagram Pictures. In *Proceedings of the 3rd Workshop on Emotions and Personality in Personalized Services (EMPIRE)*, Vienna, Austria, September 2015.

[7] B. Ferwerda, M. Schedl, and M. Tkalčič. Using Instagram Picture Features to Predict Users' Personality. In *Proceedings of the 22nd International Conference on MultiMedia Modeling (MMM 2016)*, Miami, USA, January 2016.

[8] B. Ferwerda, A. Vall, M. Tkalčič, and M. Schedl. Exploring Music Diversity Needs Across Countries. In *Proceedings of the 24th International Conference on User Modeling, Adaptation and Personalization (UMAP)*, Halifax, Canada, July 2016.

[9] B. Ferwerda, E. Yang, M. Schedl, and M. Tkalčič. Personality Traits Predict Music Taxonomy Preferences. In *ACM CHI '15 Extended Abstracts on Human Factors in Computing Systems*, Seoul, Republic of Korea, April 2015.

[10] J. Golbeck, C. Robles, M. Edmondson, and K. Turner. Predicting Personality from Twitter. In *Proceedings of the 3rd IEEE International Conference on Social Computing (SocialCom)*, Boston, USA, October 2011.

[11] G. Hofstede, G. J. Hofstede, and M. Minkov. *Cultures and Organizations: Software of the Mind*. McGraw-Hill, USA, 3rd edition, 2010.

[12] R. J. House, P. J. Hanges, M. Javidan, P. W. Dorfman, and V. Gupta. *Culture, leadership, and organizations: The GLOBE study of 62 societies*. Sage publications, 2004.

[13] S. Kitayama and H. Park. Cultural Shaping of Self, Emotion, and Well-Being: How Does It Work? *Social and Personality Psychology Compass*, 1(1):202–222, 2007.

[14] P. Laufer, C. Wagner, F. Flöck, and M. Strohmaier. Mining cross-cultural relations from Wikipedia — A study of 31 European food cultures. In *Proceedings of the ACM Web Science Conference (WebSci)*, Oxford, UK, June–July 2015.

[15] G. Park, H. A. Schwartz, J. C. Eichstaedt, M. L. Kern, M. Kosinski, D. J. Stillwell, L. H. Ungar, and M. E. Seligman. Automatic Personality Assessment Through Social Media Language. *Journal of Personality and Social Psychology*, 108, November 2014.

[16] M. Park, I. Weber, and M. N. amd Sarah Vieweg. Understanding Musical Diversity via Online Social Media. In *Proceedings of the 9th AAAI Conference on Web and Social Media (ICWSM)*, Oxford, England, UK, May 2015.

[17] D. Quercia, M. Kosinski, D. Stillwell, and J. Crowcroft. Our twitter profiles, our selves: Predicting personality with twitter. In *Proceedings of the 3rd IEEE International Conference on Social Computing (SocialCom)*, Boston, USA, October 2011.

[18] M. Schedl. The LFM-1b Dataset for Music Retrieval and Recommendation. In *Proceedings of the ACM International Conference on Multimedia Retrieval (ICMR)*, New York, USA, June 2016.

[19] B. Scheibehenne, R. Greifeneder, and P. M. Todd. Can there ever be too many options? A meta-analytic review of choice overload. *Journal of Consumer Research*, 37(3):409–425, 2010.

[20] M. Skowron, B. Ferwerda, M. Tkalčič, and M. Schedl. Fusing Social Media Cues: Personality Prediction from Twitter and Instagram. In *Proceedings of the 25th International World Wide Web Conference Companion (WWW)*, Montreal, Canada, April 2016.

[21] M. Tkalčič, B. Ferwerda, D. Hauger, and M. Schedl. Personality Correlates for Digital Concert Program Notes. In *Proceedings of the 22nd International Conference on User Modeling, Adaptation and Personalization (UMAP)*, Dublin, Ireland, June–July 2015.

[22] F. Trompenaars and C. Hampden-Turner. *Riding the Waves of Culture — Understanding Diversity in Global Business*. McGraw-Hill Education, 3rd edition, 2012.

[23] M. C. Willemsen, B. P. Knijnenburg, M. P. Graus, L. C. Velter-Bremmers, and K. Fu. Using latent features diversification to reduce choice difficulty in recommendation lists. In *Proceedings of the RecSys 2011 Workshop on Human Decision Making in Recommender Systems*, Chicago, USA, October 2011.

6 Implicit Acquisition of Users' Personal Characteristics From Social Media

Ferwerda, B., Schedl, M., & Tkalcic, M. (2014) To Post or Not to Post: The Effects of Persuasive Cues and Group Targeting Mechanisms on Posting Behavior. *In Proceedings of the 6th International Conference on Social Computing* (Stanford, CA, US).

Ferwerda, B., Schedl, M., & Tkalcic, M. (2015) Predicting Personality Traits with Instagram Pictures. *In Extended Proceedings of the 9th ACM Conference on Recommender Systems: 3rd Workshop on Emotions and Personality in Personalized Services* (Vienna, AT).

Ferwerda, B., Schedl, M., & Tkalcic, M. (2016) Using Instagram Picture Features to Predict Users' Personality. *In Proceedings of the 22nd International Conference on MultiMedia Modeling* (Miami, FL, US).

Ferwerda, B., Schedl, M., & Tkalcic, M. (2016) Personality Traits and the Relationship with (Non-)Disclosure Behavior on Facebook. *In Companion Proceedings of the 25th International Conference on World Wide Web: 7th International Workshop on Modeling Social Media* (Montreal, QB, CA).

Skowron, M., **Ferwerda, B.**, Tkalcic, M., & Schedl, M. (2016) Fusing Social Media Cues: Personality Prediction from Twitter and Instagram. *In Companion Proceedings of the 25th International Conference on World Wide Web* (Montreal, QB, CA).

To Post or Not to Post: The Effects of Persuasive Cues and Group Targeting Mechanisms on Posting Behavior

Published in: Proceedings of the 6th International Conference on Social Computing (SocialCom 2014; Stanford, CA, US)

BRUCE FERWERDA, Johannes Kepler University, Linz (AT)
MARKUS SCHEDL, Johannes Kepler University, Linz (AT)
MARKO TKALCIC, Johannes Kepler University, Linz (AT)

Abstract

When social network sites (SNSs) users intend to share content, they need to estimate the appropriateness of the content for their audience. Wrongly made estimations can result in regret about the posted content. A common strategy for users to minimize regret is to self-censor content. However, this also means that content that would have been safe to share may be left unshared. To solve sharing problems, SNSs have been focusing on improving group targeting mechanisms to give users more control over their content. As users still need to estimate the content appropriateness themselves, we asked whether improving these mechanisms is really the solution. We hypothesized that users' posting decisions consist of uncertainty and therefore providing guidance on whether it is safe to post would be more beneficial. To answer this we conducted two studies. In Study A we identified what kind of content users are self-censoring and what the reasons are. Study B was used to test and compare different solutions to limit the self-censored content found in Study A. We created a persuasive cue that predicted how the user's audience would possibly respond to the content and compared this with the effects of a group targeting mechanism. Among 215 participants we found that posting decisions consist of uncertainty and that persuasive cues are a more effective means to limit self-censorship, but can also warn users of content that is not safe to post. Making use of such cue can improve SNSs' sociability and reduce regret of wrong posting decisions.

Keywords: Persuasive Cues, Posting Behavior, Censoring, Social Networks

1 Introduction

Social network sites (SNSs) are rapidly growing with the adoption spreading out across a wide audience [8]. This continuous adoption have been changing online social practices and experiences [12]. The encouragement of sociability resulted in a growing social diversity in the user's social network. SNSs like Facebook have become an "all-friends-in-one-place" solution, meaning a larger number of social connections with a mix of strong and weak ties [9].

The trend of becoming an all-friends-in-one-place environment can create a "privacy dilemma," i.e., a conflict between the privacy needs of individuals and the need for sociability and content sharing. When privacy is protected, sociability and content sharing will be compromised and vice versa. In both cases the outcome is undesirable [9]. Previous research argued that these problems are exacerbated because users are experiencing difficulties to share content with specific groups in their social network [9, 17, 22]. As a result, users self-censor to anticipate regret of wrongly posted content [24]; they adjust their posting or eventually decide not posting at all. Sleeper et al. [22] found that reasons for self-censoring behavior are mainly concentrated around the anticipated feelings or opinions of the user's audience (e.g., not wanting to start an argument, worried to offend or hurt someone, feeling that the content would be boring or redundant, or feeling that the content went against the way users wanted to present themselves). This suggests that the reason for users to start censoring their content may not be the difficulties to target specific groups, but the perception users create on how their audience will respond to the posted content. Therefore, users may not seek a better way to target groups in their network to share content with when considering to post content, but a way to assure how their audience will respond to the content they want to post instead. Being able to target specific groups in an ever growing social network may rather be a strategy of users in order to better predict how their audience will respond to the content by shrinking the audience to a manageable size.

We created an online experiment to test this hypothesis in which we set up a persuasive cue that provide users with possible responses on the content against a group targeting mechanism. By comparing these two methods against each other we investigated the behavioral change these methods can induce in posting decisions and gain deeper insight in the underlying mechanisms of posting behavior. These insights provide knowledge on how to help users to make better posting decisions. This is important as SNSs are intended for people to create and share content about themselves as a result of voluntary disclosure among multiple users [16]. Wrongly made decisions in the posting process affects the functionality of SNSs: not posted content that should have been posted jeopardizes SNSs functionality as it limits sociability, while posted content that should not be posted contributes to regret of the poster.

2 Related work

In this section we discuss users' behavior when considering sharing content on SNSs. We continue the related work with how users are judging the appropriateness of the content followed by their strategy to reduce regret of posted content. Finally, we discuss recent work that tries to overcome sharing difficulties.

Behavior on SNSs
Users have difficulties in defining their social connections. The "friend" category in SNSs is very broad and ambiguous. Most users tend to list anyone they have met during the course of their lives and do not actively dislike [6]. This kind of behavior results in that the user's social network include a mix of strong and weak ties [13]. SNSs have created extensive sets of privacy controls that allow users to interact at different levels of sociability. Despite these privacy controls, users are not utilizing the privacy settings provided [5]. Although users are not utilizing the privacy settings, they still actively share their content [11]. The content considered for sharing undergoes the evaluation of possible regret. Shared content becomes regret because of unforeseen or ignored consequences, such as they want to be perceived in favorable ways, they do not think about the consequences of their post, or misjudge the culture and norms within their network [24].

The imagined audience
When considering content to share, users have a sense of audience. Strater and Lipford [23] found that the user's perceived audience shrinks over time; as users interact more with a certain group they start to perceive them as their primary audience and pay less attention to others. This behavior suggests that connections other than the user's perceived audience become less significant over time, thus also less important to take into consideration to share content with.

Users create an "imagined" audience to estimate the appropriateness of the content presentation [19]. They use cues from their social media environment to construct and augment knowledge about their audience [7]. However, as the interpretation of these cues are subjective, the created image may not be accurate and can be deviated from the actual readers. When the environment involves higher interaction with its readers, the awareness of the imagined audience becomes more prominent and thereby the specifics becomes more important [19]. A deviated image of the audience can therefore become problematic. Especially in SNSs where sharing is part of being social, the possible misconception users create can significantly influence the way users share their content.

Self-censorship

One of the strategies to minimize the risk of regret is to self-censor content [19, 22, 24]. Although self-censorship is an effective strategy to prevent regret, it also increases the chances that content that would have been safe is left unshared. In a qualitative study by Sleeper et al. [22] in the U.S., most commonly found content subject to self-censoring is external content (e.g., entertainment, politics) followed by personal opinions and updates. The most important reason connected to these contents is to control the self-presentation. To a slightly lesser extent, other reasons of censoring content were found: users did not want to start an argument or discussion, or were afraid to be boring or repetitive. These self-censoring reasons apply to almost half of the content considered to be shared [22].

Privacy settings

To help users in their sharing decisions, research has focused on providing different methods to increase control and transparency of disclosure behavior. Extensively investigated are methods to improve group targeting settings, such as machine-learning solutions [3] as well as ways to give users more control about what to share and with whom [21]. A recent system providing an advanced sharing method are "circles" on Google+ [17]. Although such settings provide users with more control, they do not help prevent users posting content that they should not post. Additionally, research about privacy settings indicate that users often do not use the custom privacy settings available, but adhere to standard settings instead [22, 23, 24]. These standard settings provide an "all or nothing" situation. Users are able to set their information open to everybody or protecting their whole profile by utilizing the setting to restrict it to "friends only." Although users choose for a specific setting, they still deliberately choose to self-censor information and not share content [23].

Persuasive cues

Some studies have been focusing on accommodating persuasive cues by providing a justified reason for the user to disclose information, such as giving a reason why it would be better to disclose [10], or appealing to the social norm by displaying what others have done [1, 4, 20]. The presentation of these kind of persuasive cues have been studied in different ways. For example Patil, Page, and Kobsa [20] used a descriptive way to present the aggregated privacy choices of one's social circle, while Besmer, Watson, and Lipford [4] chose for a visual approach to display the social norm.

3 The actual sharing problem

Based on the just discussed literature, we can conclude that improving group targeting mechanisms may not be the right solution as users do not make sufficiently use of them. Even if users choose a setting, they still rely on what they *think* is acceptable. Ajzen [2] defines this influence on the decision making as the subjective norm in which behavior is influenced by the perceptual judgments and beliefs of relevant others about the intended behavior. Results of previous research on SNSs indeed suggest that the underlying reasons for users to start censoring their content are the perceptions they create about responses of their audience on the content [19, 22, 23, 24]. What is shared is judged on its perceived safeness and appropriateness, as well as what is considered to be socially acceptable and normal within their social network [23]. As these judgments are subjectively created by the user, they are uncertain to some extent.

The work in this study is based on the assumption that posting decisions consist of uncertainty. We believe that by presenting a cue that provides possible responses of the user's audience on the content, we can minimize that uncertainty. Providing such cue will help users in their decision making process and prevent them from wrongly censoring or posting content. To the best of our knowledge, the usage of this kind of cues has not yet been explored in the context of content sharing. Our study was designed to answer the following main questions:

1. Will presenting possible responses of the audience change posting behavior in compliance?
2. Will providing a group targeting mechanism in a self-censoring state change posting behavior?
3. When in a self-censoring state, will presenting responses have a stronger effect on behavioral change than a group targeting mechanism?

Users create a subjective judgment about what is acceptable within the audience to which the content is shown. Therefore, we expect that when users know how the audience will possibly respond on the content, they will adjust their posting behavior towards the polarity of the responses. Presenting responses of the user's audience on the content will reduce self-censoring behavior, but also warns users of posting content that they should not. In both cases, this cue reduces regret of a wrongly made decision.

We expect group targeting mechanisms to alleviate self-censoring problems to some extent. By using group targeting mechanisms, users are gaining more control over their content. That is, to whom they share the content with. By targeting specific groups in the network, users are able to shrink their ever growing social network to a size in which they can make more easily judgments about the appropriateness of the content. We expect that it only alleviates the self-censoring problem partly, hence the appropriateness estimation of the content is still subjectively created by the user. Thus, giving users more control will have a smaller effect compared to presenting possible responses of the audience.

Given that we expect both methods to have a positive effect on posting decisions, there may be an accumulation of the effect size when both methods are combined. We expect that combining the methods will provide a positive interaction effect.

4 Methodology

The experiment that would provide insights in posting behavior was conducted in two steps, which we will refer to as *Study A* and *Study B*. The goal of Study A was to identify 1) content for posting, and 2) reasons for censorship. We used these outcomes in the design of Study B, where we observed the actual posting and censorship behavior. In Study B, we put the user in a posting position for a selected content. Then the participant was subjected to a self-censorship inducing scenario and put in one of the experimental conditions. Afterwards, the participant was asked again if they wanted to post the content. The difference between the first posting position and the second were the observed behavior changes.

4.1 Study A

In order to acquire the desired data (content for posting and reasons for censorship) we followed the experiment design laid out by Sleeper et al. [22]. The workflow of the study is depicted in Fig. 1.

We recruited 21 university students (8 males, aged between 20 26 years, median age 22 years) for a one week diary study and a concluding in-

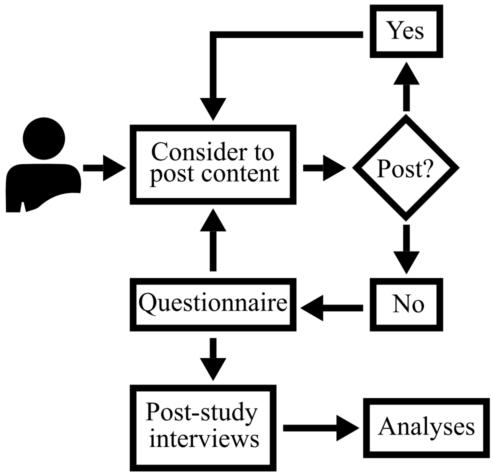

Figure 1: Study A: Workflow

terview. Participants were asked to fill-in a questionnaire every time they self-censored a content posting on Facebook. The questionnaire consisted of three questions: 1) context of the content, 2) content type (e.g., photo, video etc.), and 3) reason for censoring. After a week of reporting posting censorship, participants were invited for a post-study semi-structured interview where they discussed in more details the censored content and the reasons for censorship.

4.1.1 Data Coding and Analysis

The goal of data coding was to map the raw data from the diary study and interviews into a set of self-censored contents and reasons. The results of coding are reported in Fig. 2 and Fig. 3.

In total, participants reported 88 self-censored contents in the diary study. Following the coding scheme of Sleeper et al. [22], coding took place in four steps: 1) two researchers coded half of the content each, 2) based on these codes, the researchers created a set of higher level codes, 3) these were used to code the remaining halves (for each researcher), and 4) disagreements and inconsistencies were discussed jointly.

4.1.2 Results

Participants self-censored various kind of content (see Fig. 2). Most censored contents in our sample were about personal opinions and updates. Personal opinions were mostly about how participants felt about specific

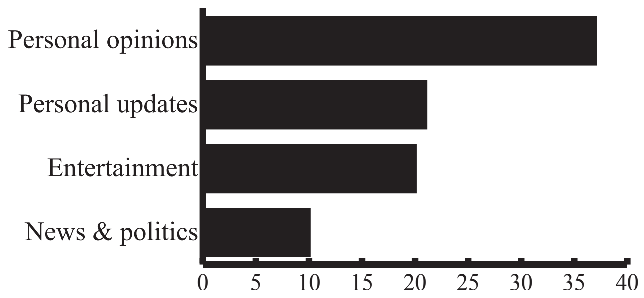

Figure 2: Number of self-censored content by type.

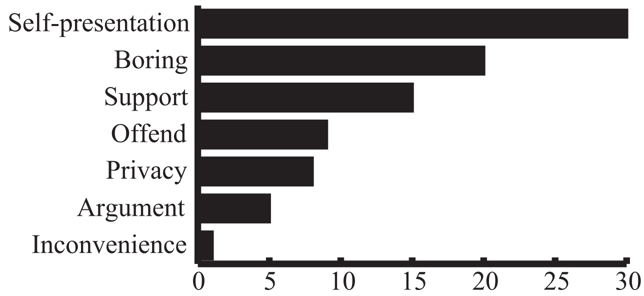

Figure 3: Number of self-censored reasons by type.

things happening in their life whereas personal updates were in general about happenings of participants throughout the day. These were mainly expressed as status updates on Facebook. Next to personal content, entertainment content (i.e., music, humoristic, sports, and fashion items) are most commonly censored. Finally, to a much lesser degree participants self-censored political and news items.

Responses of participants about the reason of self-censoring could be placed in one of the following categories (see Fig. 3):

- **Self-presentation:** worried that the content would hurt their own presentation.
- **Boring/repetitive:** concerned that the content would be perceived as boring or repetitive.
- **Support/sympathy:** worried that friends would not respond on the content (comment or like).
- **Argument:** did not want to get involved into an argument about the topic with anybody.
- **Offend:** felt that the content may hurt somebody.
- **Privacy:** felt that the content would violate privacy of others.

- **Inconvenience:** too much effort to post the content.

The main reason for participants to censor their content is because of self-presentation concerns. This is for a big part related to the personal opinions and updates categories. Participants were mainly worried that posting something about themselves would possibly create a negative image toward others. Other significant reasons is the believe that the content would be perceived as boring, or that they would not get the appropriate support and sympathy from others. To a lesser degree participants were worried about privacy, creating an argument, or offending someone.

In the interviews we additionally asked participants if they would post the self-censored content under different circumstances, such as being able to target specific groups in their network or if they would know how their social network would possibly respond. Striking was that all the participants said that for personal opinions and updates that there was nothing to change their decision. For all other content, participants seemed to be more open minded. Furthermore, we asked participants about disclosure concerns with certain groups in their social network. Depending on the content type, most participants expressed concerns about sharing content with some of the groups in their network except for "close friends." Participants did not show concerns about disclosing any content with this group.

4.1.3 Discussion

Our goal in this study was to better understand the contents and reasons for self-censoring behavior. Additionally, as our study took place with South Korean participants, previous findings of Sleeper et al. [22] gave us the opportunity to explore possible cultural differences. Most of our findings show agreements with the results of previous research. Although we found less varied content, the content types we found are in line with the findings of Sleeper et al. [22] among U.S. participants. Furthermore, a resemblance can be seen in the order of the major categories. The less variation in our study may be explained by the age range and occupation variation of our sample. We focused on university students between 20 and 30 only, while the sample of Sleeper et al. [22] varied in their occupation and age (20-51). Still, similar self-censoring reasons were found.

A compelling additional reason we found among South Koreans is the concern for support and sympathy. Our participants expressed a substantial concern of getting no responses on their postings. This longing for support and sympathy may be connected to cultural dimensions. As most Asian cultures, Koreans tend to have a collectivistic nature that may explain the importance of this reason [15]. Despite this difference, our findings show in general similar trends between Americans and Koreans indicating that cultural differences may not be so prominently present as thought.

4.2 Study B

Study B represents the core experiment with which we wanted to get insights into how a group targeting mechanism and a response prediction on the content can influence the posting behavior. We developed an online experiment where the users 1) first chose a content, then 2) decided whether to post it or not, then 3) were subjects to a manipulation (we had manipulations dealing with user groups and with predicted response; see §4.2.1 for details), and finally 4) decided again whether to post the content or not. This experiment design, outlined in Fig. 4, allowed us to gather the data needed to make the conclusions to our initial research questions.

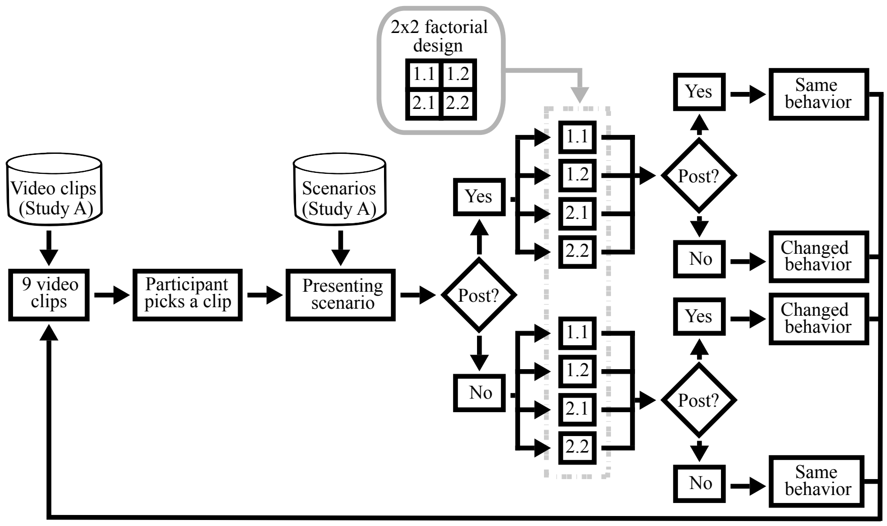

Figure 4: Study B: Workflow

We recruited 215 participants (104 males, age ranging from 20 to 30 years, median age 24 years) among university students in South Korea. Participants were asked to post content through a Facebook-like web application. The application was a visual clone of Facebook running on our server.

In a pre-step we chose five groups that were to be used in some of the manipulations (see §4.2.1). Participants were asked to define two groups of friends out of their social network. In addition to these two groups we added three more: *friends*, *public* (default Facebook groups), and the group *close friends* (participants in Study A suggested that they would have had a different posting behavior if there was the possibility to target such a group).

In the next step we put the user in a self-censoring situation. Self-censoring is one of the possible results of a decision made about content that users are considering to post. In other words; when a user has a positive attitude to post the content but feels hesitant to continue the posting. Based on the outcomes of Study A (see §4.1.2), we selected the set of contents that were offered for posting to the participants. This set was composed of 9 different video clips of roughly 1 minute of length: movie (2x), music (2x), humor (2x), TV-series (1x), politics (1x), and news (1x). Participants chose one of the content items that they would consider to share. This ensured the necessary positive attitude towards posting the content. To ensure the possibility to hesitate on posting the content, participants were then presented with a reason for not posting (chosen from the set of reasons collected in Study A; see §4.1.2) and were asked whether they wanted to post the content. If the participant decided not to post, self-censorship has occurred as the participant had a positive attitude to post by choosing their own content.

In the next step the participants were exposed to one of the four conditions that could have an effect on their posting behavior. After being exposed to the conditions the participants were asked again if they wanted to post the content. For each participant, the experiment flow was designed in such a way that they were exposed to all four conditions in a random order.

4.2.1 Manipulations

We wanted to observe the influence of two factors on the posting behavior: a) group targeting mechanism, and b) predicted audience's response. Hence we used a 2x2 within-subject factorial design consisting of the following manipulation conditions (see Tab. 1): (1.1) with group targeting mechanism and with predicted audience's response, (1.2) with group targeting mechanism only, (2.1) with predicted audience's response only, and (2.2) without manipulation (control condition). Using a within-subject design the participants went four times through the procedure. For each round the content choices, as well as the reasons for self-censoring, differed. To cancel out order effects, participants were randomly assigned to one of the conditions until all four were met.

Audience's response prediction
This condition was designed to present possible responses of the user's social network members. In order to effectively capture the effect of the manipulation, we designed it to oppose the participant's choice of posting.

Condition	Group targeting mechanism	Predicted audience response
(1.1)	Yes	Yes
(1.2)	Yes	No
(2.1)	No	Yes
(2.2)	No	No

Table 1: Manipulations

Since participants could choose whether or not to follow the self-censoring scenario, we created two different textual messages based on work of Patil et al. [20]: 1) positive/comforting message, and 2) negative/warning message. When participants decided to censor the content, a comforting message was presented trying to change their decision (e.g., "We analyzed your social network and based on their [participant's audience] responses on similar content they would like this posting."). A warning message was presented when participants did not censor their content. That is, when they decided to post the content regardless the self-censoring scenario provided (e.g., "We analyzed your social network and based on their [participant's audience] responses on similar content they will not like this posting."). Next to the text messages, we also included a visual method (based on work of Besmer et al. [4]) where we used a bar indicating the percentage of the user's social network that would like the posting. For the comforting message we set the bar on 95%, and for the warning message we set the bar on 5%. This condition alone was present in the manipulation ((2.1); see Tab. 1).

Group targeting mechanism

In this condition we displayed the five groups of users (two participant generated groups, *public*, *friends*, and *close friends*) and required the subject to choose one of the groups when posting the content. We did not define a default group in order to avoid situations in which participants would just click through or get influenced by the default option. This condition alone was present in the manipulation ((1.2); see Tab. 1).

Combined audience's response prediction and group targeting mechanism

We combined the audience's response prediction messages and the group targeting mechanism condition in order to investigate the existence of an accumulative effect. Both conditions were presented in the manipulation ((1.1); see Tab. 1).

Control
To ensure that the observed effects are not due to the fact that participants could think their previous decision over, we included a control condition. This condition gave participants the opportunity to change their previous posting decision without the addition of a guiding message or added functionality. To this end, we provided just a message asking if they wanted to change their previously made decision.

4.2.2 Measurements

In order to investigate the behavioral change the conditions induce in posting behavior, we captured three variables: 1) the posting behavior of participants before the condition presented (post or not post), 2) the experimental condition presented to the participants (i.e., (1.1), (1.2), (2.1) or (2.2)), and 3) the posting decision after the condition presented (post or not post). The influence of the condition was measured by whether participants changed their posting decision after the condition was presented. Additionally, to gain a better understanding about the final posting decision made by the participants, we asked them to write down the reasons for their behavior.

4.2.3 Results

As we gave participants the choice to follow the self-censoring scenario we created, or to ignore it, we obtained different starting posting behaviors. We divided our analyses in order to investigate the behavioral change of the manipulations on not posting (self-censoring) and posting behavior. In this section we first discuss the results of participants that followed the self-censoring scenario. That is, whether the manipulations could induce a behavioral change from self-censoring (not posting) to posting. We continue with the results of our analysis where we discuss the results of participants that ignored the self-censoring scenario. In other words, whether the manipulations could induce a behavioral change from posting to not posting (self-censoring).

In general, our interest was the effect of the manipulations on posting behavior and the effect size of each. We performed a repeated measures logistic regression to test this, by using a generalized linear model (GENLIN) with a binomial distribution and a logit link function.

Not posting (self-censoring)
In this section we present our results with the self-censoring cases ($n = 90$).

The GENLIN's goodness-of-fit is reflected in the quasi likelihood criterion ($QIC = 188.434$) and the corrected quasi likelihood under independence model criterion ($QICC = 118.631$). A low difference between QIC and QICC indicates that the model has a good correlation structure and the predictors obtained fit the model well.

To investigate the effects of the logistic regression, odds ratios (OR) are reported. We first assessed the baseline odds (control condition) in order to see the effects of the manipulation conditions. When participants did not receive the audience's responses on the content nor able to target specific user groups in their social network, we found the odds that they changed their initial posting decision (censoring) to be 0.207 ($CI = 0.076$ to 0.564, Wald Chi-Square = 9.479, $p = 0.002$). In other words, when participants were just given the opportunity to readjust their decision, 0.207 participants decided to change their posting decision for every participant that did not. The lack of a higher odds ratio in a certain direction supports our assumption that posting decisions in a self-censoring state is uncertain as participants' posting decision fluctuates.

Exploring the effect of presenting participants with possible responses of their audience, we found an increase of the baseline odds. When participants knew how their audience would respond on the to be posted content, the baseline odds increased by a factor 9.302 ($CI = 2.548$ to 33.955, Wald Chi-Square = 11.397, $p = 0.001$). This means that the odds of participants changing their initial posting decision in this condition is 1.93. We also found an increase of the baseline odds when participants were only given a group targeting mechanism. The baseline odds increased by a factor 5.375 ($CI = 1.748$ to 16.529, Wald Chi-Square = 8.609, $p = 0.003$). In other words, the odds of participants posting their content when they could target specific groups is 1.11. Given these results, we can conclude that both conditions induce a posting decision change. However, the odds are higher when participants knew how their audience would respond.

The interaction effect of combining the two manipulation conditions show a minimal change in the odds ratio ($OR = 0.015$, $CI = 0.042$ to 0.563, Wald Chi-Square = 7.968, $p = 0.005$). This suggest that there is an effect between knowing how the audience would respond and being able to target user groups. However, the odds increase is negligible small.

The choice that participants could make between different content types could have played a role in the effect of the conditions. To investigate this, we dummy coded the content type in order to add them as a covariate. The only (marginally) significant results were found for political ($OR = -0.131$, $CI = 0.016$ to 1.061, Wald Chi-Square = 3.626, $p = 0.057$) and news items ($OR = -0.241$, $CI = 0.056$ to 1.033, Wald Chi-Square = 3.671, $p = 0.055$). Surprising is the negative odds ratio of both items, indicating a decrease in the odds of changing posting behavior. Analyzing the quantitative data obtained, revealed that participants took this content more serious. They found this content *heavy loaded*; the topics discussed were serious and having an *wrong* opinion about this could significantly influence the way others would look at them.

Posting

When asking participants ($n = 125$) why they did not follow the self-censoring scenario, they responded that they found the reasons given not severe enough for them to start censoring the content they really liked. As participants indicated to have a high positive attitude toward this content, it gave us the change to explore to which extent uncertainty consist in posting behavior. Prior to the analysis, we developed additional expectations. As we believe that posting decisions consist of uncertainty, we expected that the conditions involving the audience's response would be able to change posting behavior. Furthermore, as the group targeting mechanism condition did not have any guiding message, we expected this to not have any effect on posting decisions. That is, we expected this condition to have the same effect as the control condition.

The two quasi-likelihood criteria to assess the goodness-of-fit indicate a fairly good model fit ($QIC = 125.996$ and $QICC = 126.317$). Results show a baseline odds of 5.716 ($CI = 2.035$ to 16.054, Wald Chi-Square = 10.9464, $p = 0.001$). Meaning that 5.716 participants stayed with their initial decision of posting the content for every participant that did not. This suggest that participants were more confident about their posting as the odds of staying with their prior decision is much higher than changing it. The main effect of the audience's response condition show a decrease of the baseline odds ($OR = -0.325$, $CI = 0.092$ to 1.140, Wald Chi-Square = 3.081, $p = 0.079$). Meaning that 1.85 participants changed their posting to not post for every participant that kept on posting. As expected, a non significant main effect was found for the group targeting condition ($OR = 5.030$, $CI = 0.604$ to 41.893, Wald Chi-Square = 2.231, *ns*). Additionally, a non significant interaction effect was found (OR=.171, CI=.017 to 1.684, Wald Chi-Square=2.291, *ns*). Meaning that there is no difference between the audience's response condition with or without the option to target groups.

As with the previous section, the political content revealed a marginally significant effect ($OR = -0.172$, $CI = 0.024$ to 1.249, Wald Chi-Square = 3.626, $p = 0.057$). The negative relation indicates an increase of the odds ratio of changing posting behavior. That is, participants were more likely to change their decision from posting to not posting for this kind of content.

4.2.4 Discussion

In Study B we investigated three different methods to influence users' posting decisions on SNSs by presenting: possible responses of the user's audience, group targeting mechanism, and a combination of the two methods. Our results show that all three methods have an effect on users' posting decisions. By comparing these different methods, we were able to get a deeper understanding about the needs of users when posting content. Looking at the effect of each method, knowing whether their audience would like the posting seem to be more beneficial for users than able to target specific groups in their social network. It gives support to our notion that users are uncertain in their posting decisions as they were easily influenced. This uncertainty is highlighted in the results of the control condition in the self-censoring condition. Results showed that participants frequently reconsidered their self-censoring behavior when they were given the opportunity. When participants already decided to post the content, they were less prone to reconsider and were more likely to stick with their posting. They seem to be confident and sure about the content. However, this confidence is still affected when presented with how their audience would respond to the content. When participants knew that the content would not perceived well, they started to come back at their posting decision.

Despite the fact that all three methods provided significant results on behavioral change, we found no support for our expectation of an accumulative effect for the combined condition. This may be explained by the fact that we guided participants to share content with the close friends group. The close friends group was chosen because the results of Study A indicated that users did not perceive disclosure problems with this group. Strater and Lipfort [23] noted that users are starting to see the group with which they interact most with as their primary audience. As users do not feel any disclosure problems with the close friends group, they may interact most with this group already and therefore seeing this group as their primary audience. Presenting general responses of the user's audience (audience's response condition) or more specific as in the combined condition, would not make a difference for the user as they depict the same primary audience.

We also found differences among content type. These differences can be explained by the "load." We used different content genres that can be divided in two groups: 1) heavy loaded content, and 2) light loaded content. The heavy loaded content are more about serious items and comprised political and news content types. Movie, music, humoristic, and TV-series fall in the light loaded content category as they consist of frivolous and entertaining items. Participants tend to respond differently to possible responses of their audience depending on this content load. Our results indicate that a comforting message has less effect on heavy loaded content than a warning message. This suggest that for this kind of content, users are more cautious and prefer to err on the safe side when having the sightliest doubt. The quantitative data we obtained about the reasons for participant's final posting decision indicate that the content load indeed plays a prominent role. Especially for the political content, participants were very careful by expressing their opinion as it indicates their political side. For light loaded content, these concerns were less present. Although this kind of content can also express a certain preference, participants felt that they could "laugh it away" when it would play against them. For this type of content participants seemed to be more sensitive for a comforting message that gives them that little push to post the content. However, when they are feeling confident about the content, the effect of the warning message decreases. This behavior is seen when participants did not follow the self-censoring scenario we presented to them. Additionally, it can also explain why a part of the participants did not follow the scenario. As they felt they could easily repair damage made to their self-presentation, they found the reasons given not severe enough to start censoring the content. This changeable behavior that we could induce by presenting possible responses accentuate the uncertainty that exists when users are trying to post their content on SNSs.

5 Conclusion

Results suggest that posting behavior consist of uncertainty. Especially when users are self-censoring their content, they tend to change their decision quite frequently when given the chance. Although users seem to be more sure and confident about posted content, results indicate that some degree of uncertainty still exist. Group targeting mechanisms seem to contribute to influence posting decisions to some extent. Improving group targeting mechanisms can only partly take away uncertainty by allowing users to shrink their social network to a size that they can easily estimate the content appropriateness. However, users can still make wrong estimations. Results show that by providing predictions about how the audience would respond to the content significantly help users to make better posting decisions; when to self-censor the content, or when it is safe to post. With regards to our research questions (see §3) this means that: 1) presenting users with possible responses of their audience do result in an adjusted posting towards the polarity of the responses, 2) group targeting mechanisms influence posting behavior in a self-censoring state, and 3) audience's response predictions do have a stronger effect on posting behavior than group targeting mechanisms.

Additionally, the posting guidance provided by the audience's response predictions depends on the content type. Users seem to be more cautious and suspicious of posting content that is serious and heavy loaded. Posting this content can damage their self-presentation which may be difficult to repair. Therefore, users are less sensitive for positive guidances for posting, and more sensitive for negative guidances (when there is any doubt of hurting their self-presentation, they will err on the safe side of their posting behavior; not posting the content). For light loaded content, users can use a little push to persuade them to continue posting. Even if this kind of content will not be received well by their audience, the consequences for their self-presentation are less severe and will be more easy to repair as the content is light and not serious. Therefore, participants seem to care less about what kind of responses the content will produce.

The effects found of giving guidance in user's posting decisions give raise to the question whether research should continue focusing on improving group targeting mechanisms. As SNSs are made for sociability, greater advantage should be taken of the information that can be derived from a user's social network instead of trying to limit sociability by providing restrictions.

6 Limitations and future work

Study A relied on qualitative, self-reported data, and a relatively small sample was used. Although results seem to be consistent with findings of Sleeper et al. [22], we still lack the ability to generalize our findings. Furthermore, both studies (A and B) used a specific subset of actual SNS users; university students between 20s and 30s. Brandtzaeg et al. [9] argue that variation can occur among different age ranges and occupations. For example, older and younger users have other privacy perceptions and therefore could respond differently to the manipulations we provided. To alleviate these limitations, a bigger and wider sample need to be obtained in future studies in order to properly map the social diversity on SNSs.

In Study B, we made use of a combined condition where we merged the audience's response condition with the group targeting condition. To measure the influence of the message effectively, we tried to guide sharing behavior toward one specific group; close friends. An interesting direction would be to investigate the effects when given posting guidance for each user group separately to see if users have separate posting thresholds for different user groups. For example, users may desire a higher percentage of approval of the content when considering sharing content with close friends than with everybody.

Our findings could have been influenced by cultural dimensions. As we already stressed out in the discussion of Study A, the additional reason found for self-censoring may be explained by the collectivistic nature of South Koreans. The results of Study B need to be interpreted with some precaution as they could have been influenced by cultural dimensions too. One explanation of our findings is that self-esteem in collectivistic cultures is not derived through idiosyncrasy [14], but rather through harmony with the group [25]. This makes people in these cultures tend to fit in rather than to stand out [18]. Therefore, our results could have been compromised by participants trying to fit in by following the cue we provided (audience's response and combined conditions). Future studies should try to answer this question by using a more individualistic society.

Lastly, as we focused specifically on content sharing, our findings might be well applicable to other areas involving information disclosure.

7 Acknowledgments

This research is supported by the Austrian Science Fund (FWF): P25655; and by the EU FP7/2007-2013 through the PHENICX project under grant agreement no.601166. Furthermore, we want to thank Sunwha Lee and Yeunju Kim for their assistance.

References

[1] Acquisti, A., John, L.K., Loewenstein, G.: The impact of relative standards on the propensity to disclose. Journal of Marketing Research 49(2), 160–174 (2012)

[2] Ajzen, I.: The theory of planned behavior. Organizational behavior and human decision processes 50(2), 179–211 (1991)

[3] Amershi, S., Fogarty, J., Weld, D.: Regroup: Interactive machine learning for on-demand group creation in social networks. In: Proc. CHI. pp. 21–30. ACM (2012)

[4] Besmer, A., Watson, J., Lipford, H.R.: The impact of social navigation on privacy policy configuration. In: Proc. SOUPS. p. 7. ACM (2010)

[5] Bonneau, J., Preibusch, S.: The privacy jungle: On the market for data protection in social networks. In: Economics of information security and privacy, pp. 121–167. Springer (2010)

[6] Boyd, D.: Friends, friendsters, and myspace top 8: Writing community into being on social network sites. First Monday 11(12) (2006)

[7] Boyd, D.: Why youth (heart) social network sites: The role of networked publics in teenage social life. MIT Press, Cambridge, MA (2007)

[8] Brandtzæg, P.B., Heim, J.: Why people use social networking sites. In: Online communities and social computing, pp. 143–152. Springer (2009)

[9] Brandtzæg, P.B., Lüders, M., Skjetne, J.H.: Too many facebook "friends"? content sharing and sociability versus the need for privacy in social network sites. Intl. Journal of Human–Computer Interaction 26(11-12), 1006–1030 (2010)

[10] Consolvo, S., Smith, I.E., Matthews, T., LaMarca, A., Tabert, J., Powledge, P.: Location disclosure to social relations: why, when, & what people want to share. In: Proc. CHI. pp. 81–90. ACM (2005)

[11] Debatin, B., Lovejoy, J.P., Horn, A.K., Hughes, B.N.: Facebook and online privacy: Attitudes, behaviors, and unintended consequences. Journal of Computer-Mediated Communication 15(1), 83–108 (2009)

[12] Ellison, N.B., Lampe, C., Steinfield, C.: Feature social network sites and society: current trends and future possibilities. interactions 16(1), 6–9 (2009)

[13] Ellison, N.B., Steinfield, C., Lampe, C.: The benefits of facebook âĂIJfriends:âĂİ social capital and college studentsâĂŹ use of online social network sites. Journal of Computer-Mediated Communication 12(4), 1143–1168 (2007)

[14] Goncalo, J.A., Staw, B.M.: Individualism–collectivism and group creativity. Organizational Behavior and Human Decision Processes 100(1), 96–109 (2006)

[15] Gudykunst, W.B., Matsumoto, Y.: Cross-cultural variability of communication in personal relationships. Communication in personal relationships across cultures pp. 19–56 (1996)

[16] Joinson, A.N.: Looking at, looking up or keeping up with people?: motives and use of facebook. In: Proc. CHI. pp. 1027–1036. ACM (2008)

[17] Kairam, S., Brzozowski, M., Huffaker, D., Chi, E.: Talking in circles: selective sharing in google+. In: Proc. CHI. pp. 1065–1074. ACM (2012)

[18] Kitayama, S., Markus, H.R., Lieberman, C.: The collective construction of self esteem: Implications for culture, self, and emotion. Kluwer Academic/Plenum Publishers (1995)

[19] Marwick, A.E.: I tweet honestly, i tweet passionately: Twitter users, context collapse, and the imagined audience. New Media & Society 13(1), 114–133 (2011)

[20] Patil, S., Page, X., Kobsa, A.: With a little help from my friends: can social navigation inform interpersonal privacy preferences? In: Proc. CSCW. pp. 391–394. ACM (2011)

[21] Sadeh, N., Hong, J., Cranor, L., Fette, I., Kelley, P., Prabaker, M., Rao, J.: Understanding and capturing people's privacy policies in a mobile social networking application. Personal and Ubiquitous Computing 13(6), 401–412 (2009)

[22] Sleeper, M., Balebako, R., Das, S., McConahy, A.L., Wiese, J., Cranor, L.F.: The post that wasn't: exploring self-censorship on facebook. In: Proc. CSCW. pp. 793–802. ACM (2013)

[23] Strater, K., Lipford, H.R.: Strategies and struggles with privacy in an online social networking community. In: Proc. BCS HCI. pp. 111–119. British Computer Society (2008)

[24] Wang, Y., Norcie, G., Komanduri, S., Acquisti, A., Leon, P.G., Cranor, L.F.: I regretted the minute i pressed share: A qualitative study of regrets on facebook. In: Proc. SOUPS. p. 10. ACM (2011)

[25] Wink, P.: Beyond ethnic differences: Contextualizing the influence of ethnicity on individualism and collectivism. Journal of Social Issues 53(2), 329–349 (1997)

Personality Traits and the Relationship with (Non-)Disclosure Behavior on Facebook

Published in: Companion Proceedings of the 25th World Wide Web Conference (WWW 2016): 7th International Workshop on Modeling Social Media (MSM 2016; Montreal, QC, CA)

BRUCE FERWERDA, Johannes Kepler University, Linz (AT)
MARKUS SCHEDL, Johannes Kepler University, Linz (AT)
MARKO TKALCIC, Free University of Bolzano, Bolzano (IT)

Abstract

Applications increasingly use personality traits to provide a personalized service to the user. To acquire personality, social media trails showed to be a reliable source. However, until now, analysis of social media trails have been focusing on *what* has been disclosed: content of disclosed items. These methods fail to acquire personality when there is a lack of content (non-disclosure). In this study we do not look at the disclosed content, but whether disclosure occurred or not. We extracted 40 items of different Facebook profile sections that users can disclose or not disclose. We asked participants to indicate to which extent they disclose the items in an online survey, and additionally asked them to fill in a personality questionnaire. Among 100 participants we found that users' personality can be predicted by solely looking at whether they disclose particular sections of their profiles. This allows for personality acquisition when content is missing.

Keywords: Facebook, Disclosure, Non-disclosure, Personality, Personality Prediction

1 Introduction

Social networking sites (SNSs) are becoming increasingly connected with applications, such as recommender systems. The interconnectedness with SNSs lets users automatically import their information to the application by making use of a single sign-on (SSO) mechanism to authenticate. This allows users to save a considerable amount of registration time, and makes them able to use the application right away.

Before SNSs release users' profile information to an application, users need to accept a consent form that states which parts of the profile is going to be accessed by the application. Besides accessing users' basic profile information, applications often ask for additional permissions for accessing other parts of users' profile [2]. By granting access to other parts of the profile, applications are able to unobtrusively infer users' preferences and thereby able to provide the new user a more personalized experience.

User preferences can be inferred explicitly or implicitly. For example, Facebook user profiles consist of sections where users can explicitly disclose entertainment content (e.g., music, movies, books) they like, which makes inferring user's preferences straight forward. When explicit information is unavailable, an implicit approach can be adopted. Research has shown that it is possible to infer personality traits from content of social media trails (e.g., Facebook; [1, 9, 13, 16], and Twitter; [8, 14], Instagram [5, 6]). It has been shown that personality consist of reliable cues to create proxy measures about users' behavior, preference, and taste (e.g., [4, 15, 17]). However, both methods heavily rely on disclosed content. When sections are not disclosed, and thereby, content is missing, both methods fail to infer user preferences.

In this study we do not rely on the *content* of disclosed sections, but solely whether sections are disclosed, and especially *not* disclosed. To investigate the relationship between personality traits and (non-)disclosure behavior, we focus on Facebook. Facebook is one of the most popular and interconnected SNS, which in addition allows users to create an extensive user profile, with the ability to control for disclosure by assigning separate privacy settings to each section. This makes Facebook a suitable platform to study the relationship between personality traits and (non-)disclosure behavior of different user profile sections.

Our work makes several contributions. We provide insights into the relationship between (non-)disclosure behavior of profile sections and personality traits. Our findings could be used by applications to infer personality when content data is missing, hence allowing to exploit the benefits of personality to address, for example, cold start problems [18], adaptive user interfaces [7], or music recommendations [3].

We conducted an online survey where we extracted all the user's profile sections of Facebook, and asked participants to indicate for each section the items they disclose or not. Additionally, we asked them to fill in the Big Five Inventory (BFI) questionnaire in order to assess their personality.

Among 100 participants we found distinct relationships between disclosed and not disclosed user's profile sections and personality traits. In the remainder of this paper we continue with the related work, materials, results, discussion, limitations and future work, and conclusion.

2 Related work

In this work, we focus specifically on personality traits. Personality traits have shown to be an enduring factor with relationships to one's taste, preference, and interest (e.g., [4, 15, 17]). For example, a finding of Rawlings and Ciancarelli show that extraverts have a preference for pop music [15].

Several models have been developed to categorize personality, of which the five-factor model (FFM) is the most well known and widely used. The FFM categorizes personality into five general dimensions that describes personality in terms of: openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism [12].

As personality is such an enduring factor, knowing one's personality provides information about a person's taste, preference, and interest without the need of directly related data. Hence, personality is a useful measurement for personalized systems, such as recommender systems to provide an improved user experience (e.g., [3, 7, 18]). For example, Tkalcic et al. propose a method to overcome the cold start problem of new users by incorporating personality data to enhance the nearest-neighbor measurement [18]. Similarly, Ferwerda et al. use personality traits to adjust the user interface in order to match different music browsing strategies [7]. Hu and Pu showed that personality-based recommender system create an advantage (e.g., higher system loyalty of users) over systems that do not incorporate personality [10].

In order to incorporate personality information into applications, research has given attention to the implicit acquisition of personality from social media trails (e.g., Facebook [9, 13, 16], Twitter [8, 14]). For example, personality has been linked to Facebook use, such as the number of friends [16]. Others have shown personality correlations with natural language features on Twitter [8, 14]. Although, prior research has been able to infer personality traits from social media, they relied on content analyses. When content data is missing (e.g., no information is disclosed), these methods fail to infer personality traits. However, as personality is related to human behavior, which sections users disclose or not may provide indicators about their personality. More specifically, we believe that sections that users decide not to disclose is related to certain personality traits. This provides opportunities to infer personality when content data is missing.

3 Materials

To investigate the relationship between (non-)disclosure of Facebook's user profile sections and personality traits, we extracted all the items available in a user's profile. We closely observed an average Facebook profile, and extracted in total 40 items of three different sections of a Facebook profile (i.e., about, interest, and like sections; see Table 1).

About section:	
1	Work
2	Education
3	Professional skills
4	Current city
5	Hometown
6	Places lived
7	Mobile phone
8	Website
9	Email
10	Address
11	Birth date
12	Gender
13	Interested in
14	Religious views
15	Language
16	Political views
17	Relationship
18	Family members
19	About you (e.g., short description about yourself)
20	Other names (e.g., nickname)
Interest section:	
21	Music (i.e., listen later)
22	Movies (i.e., watched and want to watch)
23	TV-shows (i.e., watched and want to watch)
24	Books (i.e., read and want to read)
Like section:	
25	Movies
26	Television
27	Music
28	Books
29	Sports teams
30	Athletes

31	Inspirational People
32	Restaurants
33	Games
34	Activities
35	Interests
36	Sports
37	Foods
38	Clothing
39	Websites
40	Other

Table 1: Facebook's disclosure items with the corresponding section of occurrence.

In the survey, participants were asked to indicate to which extent they disclosed the information of the respective item (*To everybody, To friends only, Custom setting, Don't know the setting,* or *Don't disclose*), by answering the following question of the corresponding section: "In the '{section},' I disclose my {item}..." After all the disclosure questions were answered, participants were asked to fill in the 44-item BFI personality questionnaire (5-point Likert scale; Disagree strongly - Agree strongly [11]) to identify the FFM factors.

We recruited 126 participants through Amazon Mechanical Turk. Participation was restricted to those located in the United States with a very good reputation ($\geq$95% HIT approval rate and $\geq$1000 HITs approved). Additional comprehension testing questions were used to filter out fake and careless entries. The Mahalanobis distance was calculated to check for outliers. This left us with 100 completed and valid responses. Age (18-64, median 30) and gender (49 male, 51 female) information indicated an adequate distribution.

4 Results

To find the relationship between personality traits and disclosure behavior, we dichotomized the responses of the disclosure scale (To everybody, To friends only, Custom setting, Don't know the setting, or Don't disclose). Although we asked participants for their disclosure setting, providing a third-party application access to one's profile disregards that. An application will have access to the sections that a user granted access to, regardless of the disclosure setting in the profile. Hence, we recoded the responses "To everybody," "To friends only," "Custom setting," "Don't know the setting," to 1, as this means that participants had something filled in. The "Don't disclose" responses were recoded as 0.

A correlation analysis was performed to indicate the relationship between personality traits and disclosure behavior (Table 2). Point-biserial correlation (r ϵ [-1,1]) is reported as the correlation coefficient. [1] Below the results related to each personality trait. A positive correlation indicates that participants scoring high in the personality trait show a higher tendency to engage in disclosure behavior of the respective item in their user profile, while a negative correlation indicates the opposite effect.

Openness to experience. The openness to experience factor correlates with several items in the "About" section. We found negative correlations with the "Current city" (r=-.24, p=.02), "Hometown" (r=-.25, p=.01), "Mobile phone" (r=-.22, p=.03), "Website" (r=-.22, p=.03), and "Address" (r=-.24, p=.02). Additionally, we found a relationship of openness to experience with "Birth date" (r=-.018, p=.08). Negative correlations indicate a decreased tendency to engage in disclosing these items.

Conscientiousness. For the conscientiousness personality trait we found some relationships with items in the "About" section. We found correlations with "Current city" (r=-.20, p=.05), "Hometown" (r=-.18, p=.07), and "Birth date" (r=-.018, p=.07). Additionally, we found a correlation with the "Other" item in the "Like" section (r=-.19, p=.06). Results show a negative correlation meaning that conscientiousness participants indicated to be less likely to disclose these items.

Extraversion. Significant correlations were found in the "About" section and extraversion. We found correlations with "Email" (r=.23, p=.02), and "Birth date" (r=-.22, p=.03). Additionally, we found several positive correlations with items in the "Like" section and extraversion: "Restaurant" (r=.22, p=.03), "Games" (r=.18, p=.08), "Activities" (r=.21, p=.04), "Interests" (r=.17, p=.09), "Food" (r=.24, p=.02), and "Clothing" (r=.19, p=.06). Except for email and birth date, the items show a positive relationship with extraversion; indicating a higher tendency to disclose.

Agreeableness. The only correlation we found with the agreeableness personality factor is with "Places lived" in the "About" section (r=-.20, p=.04). The negative correlation indicates that agreeable participantes are less likely to engage in disclosing this item.

[1] The magnitude of the reported correlations are commonly seen in personality related research [5, 8, 9, 13, 14, 16].

Neuroticism. A correlation was found between "Birth date" and neuroticism (r=.17, p=.09). The positive coefficient indicate a positive relationship with disclosing birth date and the neuroticism trait.

		O	C	E	A	N
4	Current city	**-.24***	**-.20ˆ**	-.08	-.08	.01
5	Hometown	**-.25***	**-.18ˆ**	-.08	-.13	-.05
6	Places lived	-.12	-.12	-.08	**-.20***	-.01
7	Mobile phone	**-.22***	-.12	-.01	-.05	.10
8	Website	**-.22***	.01	.16	.02	-.16
9	Email	-.16	.09	**-.23***	.13	-.13
10	Address	**-.24***	-.02	.14	-.04	-.15
11	Birth date	**-.18ˆ**	**-.18ˆ**	**-.22***	-.12	**.17ˆ**
32	Restaurant	.03	-.06	**.22***	-.06	.09
33	Games	.10	.01	**.18ˆ**	.02	-.13
34	Activities	.05	.03	**.21***	.06	-.08
35	Interests	.09	-.04	**.17ˆ**	-.06	-.05
37	Foods	.01	-.18	**.24***	.01	-.11
38	Clothing	-.05	-.06	**.19ˆ**	.01	-.09
40	Other	-.05	**-.19ˆ**	.08	-.09	.02

Note. ˆp<0.1, *p<0.05

Table 2: Correlation Matrix of the profile items disclosure against the personality traits: (O)penness, (C)onscientiousness, (E)xtraversion, (A)greeableness, (N)euroticism. Only items that show significant levels of p<0.1 are reported.

5 Personality prediction

As we found significant correlations between personality traits and disclosure behavior, we explored personality prediction based on disclosure behavior. We trained a 10-fold cross-validation regression model with 10 iterations by using the *Radial Basis Function*. To indicate the differences between the predicted and observed values, we report the *root-mean-square error* (RMSE; see Table 3). The RMSE of each personality trait relates to a [1,5] scale.

Personality	RMSE	1	2	3
Openness to experience	0.73	0.73	0.68	0.69
Conscientiousness	0.73	0.69	0.66	0.76
Extraversion	0.99	0.95	0.90	0.88
Agreeableness	0.73	0.74	0.69	0.79
Neuroticism	0.83	0.95	0.95	0.85

Table 3: Personality prediction with the root-mean-square error (RMSE). Left RMSE column shows the results of the current study. Columns numbered 1, 2, and 3 show RSME scores of Ferwerda et al. [5, 6] and Quercia et al. [14] respectively.

To see how well our prediction performs, we compared our results with prior work of Ferwerda et al. [5, 6], and Quercia et al. [14], as they used a similar approach for their analyses. Ferwerda et al. [5, 6] extracted personality using characteristics of Instagram images, and Quercia et al. [14] uses Twitter users' characteristics (e.g., popularity, highly read; see Table 3). By disregarding content and only looking at whether sections are disclosed or not, we show that we can approach similar RSME scores as prior research analyzing social media content. Similarly, we found the most difficult traits to predict are extraversion and neuroticism.

6 Discussion

We found that personality traits are correlated with disclosing or not disclosing different parts of the user profile on Facebook. Most significant correlations are found for openness to experience, extraversion, and agreeableness. Our results indicate a relation between openness to experience and *non-disclosure* behavior in the about section of a profile, while for extraversion it is mainly *disclosure* behavior in the like section of a profile. A non-disclosure relationship was found between the places lived and agreeableness.

Additionally, we were able to identify some correlations with conscientiousness and neuroticism. The conscientiousness trait shows overlapping disclosure behavior with the openness to experience trait, whereas the neuroticism trait shows a more distinct pattern; a positive relationship was found of neuroticism on disclosing the birth date.

Furthermore, we show that the extracted Facebook items can be used to predict personality traits. Comparing with prior work (i.e., [5, 6, 14]), we found similar patterns in personality prediction; prediction is most successful for openness to experience, conscientiousness, and agreeableness, but more difficult traits are conscientiousness and neuroticism.

7 Limitations and future work

Although our results indicate correlations with disclosing behavior and personality traits, there are also several limitations to our study. Due to constraints of the Facebook API, we decided to use self-report measurements to capture disclosure behavior. There is a possibility that this self-report measure did not accurately capture all the disclosure behavior of participants. Additionally, our sample size is relatively small (n=100). We, therefore, adopted a more lenient significance level to reveal correlations with all personality traits. Reported findings would benefit from a larger sample size.

By using Amazon Mechanical Turk we focused only on participants based in the United States. However, what people disclose may be influenced by culture [19]. Future work should take cultural differences into account. Finally, we focused specifically on Facebook user profile disclosures. Interesting would be to see how and whether (non-)disclosure behavior on other platforms (e.g., Twitter, Instagram, Pinterest) are able to indicate personality traits as well.

8 Conclusion

Our results suggest that personality traits can be inferred by analyzing whether users disclose or not disclose sections in their profile. Being able to infer personality traits without content information, enables the creation of measurements to estimate personality traits even when there is no content data available. This makes it possible to facilitate personality based applications (e.g., [7, 18]) with personality approximations to create a personalized experience.

9 Acknowledgments

This research is supported by the Austrian Science Fund (FWF): P25655; and by the EU FP7/2013-2016 through the PHENICX project under grant agreement no.601166.

References

[1] Back, M.D., Stopfer, J.M., Vazire, S., Gaddis, S., Schmukle, S.C., Egloff, B., Gosling, S.D.: Facebook profiles reflect actual personality, not self-idealization. Psychological science (2010)

[2] Chia, P.H., Yamamoto, Y., Asokan, N.: Is this app safe?: a large scale study on application permissions and risk signals. In: WWW. ACM (2012)

[3] Ferwerda, B., Schedl, M.: Enhancing music recommender systems with personality information and emotional states: A proposal. EMPIRE (2014)

[4] Ferwerda, B., Schedl, M., Tkalcic, M.: Personality & emotional states: Understanding users' music listening needs. UMAP 2015 Extended Proceedings

[5] Ferwerda, B., Schedl, M., Tkalcic, M.: Predicting personality traits with instagram pictures. In: In EMPIRE. ACM, New York, NY, USA (2015)

[6] Ferwerda, B., Schedl, M., Tkalcic, M.: Using instagram picture features to predict users' personality. In: MultiMedia Modeling. Springer (2016)

[7] Ferwerda, B., Yang, E., Schedl, M., Tkalcic, M.: Personality traits predict music taxonomy preferences. In: CHI'15 Extended Abstracts. ACM (2015)

[8] Golbeck, J., Robles, C., Edmondson, M., Turner, K.: Predicting personality from twitter. In: SocialCom. pp. 149–156. IEEE (2011)

[9] Gosling, S.D., Gaddis, S., Vazire, S., et al.: Personality impressions based on facebook profiles. ICWSM (2007)

[10] Hu, R., Pu, P.: Acceptance issues of personality based recommender systems. In: RecSys. ACM (2009)

[11] John, O.P., Donahue, E.M., Kentle, R.L.: The big five inventory: Versions 4a and 54, institute of personality and social research. UC Berkeley (1991)

[12] McCrae, R.R., John, O.P.: An introduction to the five-factor model and its applications. J. of pers. (1992)

[13] Park, G., Schwartz, H.A., Eichstaedt, J.C., Kern, M.L., Kosinski, M., Stillwell, D.J., Ungar, L.H., Seligman, M.E.: Automatic personality assessment through social media language. Journal of personality and social psychology 108(6), 934 (2015)

[14] Quercia, D., Kosinski, M., Stillwell, D., Crowcroft, J.: Our twitter profiles, our selves: Predicting personality with twitter. In: SocialCom. pp. 180–185. IEEE (2011)

[15] Rawlings, D., Ciancarelli, V.: Music preference and the five-factor model of the neo personality inventory. Psychology of Music 25(2), 120–132 (1997)

[16] Ross, C., Orr, E.S., Sisic, M., Arseneault, J.M., Simmering, M.G., Orr, R.R.: Personality and motivations associated with facebook use. Computers in Human Behavior 25(2), 578–586 (2009)

[17] Tkalcic, M., Ferwerda, B., Hauger, D., Schedl, M.: Personality correlates for digital concert program notes. UMAP 2015, Springer LNCS 9146 (2015)

[18] Tkalcic, M., Kunaver, M., Košir, A., Tasic, J.: Addressing the new user problem with a personality based user similarity measure. In: DEMRA (2011)

[19] Wang, Y., Norice, G., Cranor, L.F.: Who is concerned about what? a study of american, chinese and indian users' privacy concerns on social network sites. In: Trust and trustworthy comp. Springer (2011)

Predicting Personality Traits with Instagram Pictures

Published in: Extended Proceedings of the 9th ACM Conference on Recommender Systems (RecSys 2015): 3rd Workshop on Emotions and Personality in Personalized Systems (EMPIRE 2015; Vienna, AT)

BRUCE FERWERDA, Johannes Kepler University, Linz (AT)
MARKUS SCHEDL, Johannes Kepler University, Linz (AT)
MARKO TKALCIC, Johannes Kepler University, Linz (AT)

Abstract

Instagram is a popular social networking application, which allows photo-sharing and applying different photo filters to adjust the appearance of a picture. By applying photo filters, users are able to create a style that they want to express to their audience. In this study we tried to infer personality traits from the way users take pictures and apply filters to them. To investigate this relationship, we conducted an online survey where we asked participants to fill in a personality questionnaire, and grant us access to their Instagram account through the Instagram API. Among 113 participants and 22,398 extracted Instagram pictures, we found distinct picture features (e.g., hue, brightness, saturation) that are related to personality traits. Our findings suggest a relationship between personality traits and the way users want to make their pictures look. This allow for new ways to extract personality traits from social media trails, and new ways to facilitate personalized systems.

Keywords: Instagram, Personality, Photo Filters, Picture Features

1 Introduction

Instagram is a popular mobile photo-sharing, and social networking application with currently over 300 million active users. [1] Instagram lets users easily connect with other social networking platforms (e.g., Facebook, Twitter, Tumblr, and Flickr) to share the taken pictures on, and enables users to apply filters to their pictures. At this moment Instagram offers 25 predefined photo filters that soften and color shift picture properties, for users to customize and modify their pictures to create the desired visual style.

[1] `https://instagram.com/press/` (accessed: 02/23/2015)

The ease with which a photo filter can be applied allow users to express a personal style and create a seeming distinctiveness with the customized pictures. Through the shared content and the way of applying filters, users are able to reveal a lot about themselves to their social network. With that, the question arises: What do Instagram pictures tell about the user? Or more specifically: What do Instagram pictures say about the personality of the user?

Personality traits have shown to consist of cues to infer users' behavior, preference, and taste (e.g., [3, 11, 13]). Hence, knowing one's personality can provide important cues for systems to cater to a personalized user experience. It can provide systems with estimations about user preferences without the use of extensive questionnaires or observations.

There has been an increased interest in how to use personality in systems (e.g., [2, 4, 14]), and how to automatically extract personality from online behavior trails (e.g., Facebook [1, 6, 9, 12], Twitter [5, 10]). In this work we join the personality extraction research by specifically focusing on the relationship between the picture features of an Instagram collection and the personality traits of the user.

Our work makes several contributions. We contribute to personality research by showing relationships between personality traits and the visual style of users' Instagram pictures. Additionally, we contribute to new ways to extract personality from social media (i.e., Instagram). To the best of our knowledge, we are the first to analyze (Instagram) pictures in relation with personality traits.

We conducted an online survey where we asked participants to fill in the widely used, Big Five Inventory (BFI) personality questionnaire, and grant us access to the content of their Instagram account. We extracted 22,398 Instagram pictures of 113 users, and analyzed them on several features (e.g., hue, brightness, saturation). Distinct correlations were found between personality traits and picture features.

In the remainder of the paper we will continue with related work, materials, features, results, discussion, and conclusion.

2 Related work

There is an increase of psychological research that investigate the relationship between personality and real-world behavior. Personality is known as an enduring factor and has shown to be related to a person's taste, preference, and interest (e.g., [3, 11, 13]). For example, Rawlings and Ciancarelli found relationships between personality traits and music genre preferences [11], while Tkalcic et al. found relationships between personality and classical music [13].

To categorize personality, several models have been developed. The five-factor model (FFM) is the most well known and widely used one. It categorizes personality into five general dimensions that describe personality in terms of: openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism [8].

Based on psychological findings of personality in relation to real-world behavior, there is an emergent interest in how to use and implement these findings into applications (e.g., [2, 4, 14]). It can provide useful proxy measures for applications to cater to a more personalized service. For example, Tkalcic et al. proposes to use personality traits to enhance the nearest-neighborhood measurement for overcoming the cold-start problem (i.e., recommending items to new users) in recommender systems [14]. Ferwerda et al. provide a way to use personality traits for adjusting the user interface of music applications to fit user's music browsing styles [4].

Personality traits have shown to consist of valuable information for personalized system, but hard to acquire (e.g., extensive, time-consuming questionnaires). Therefore, as applications are increasingly interconnected (e.g., social media), research started to focus on how to extract personality information from (online) behavior trails. Current research on Facebook (e.g., [1, 6, 9, 12]) and Twitter (e.g., [5, 10]) have shown to consist of reliable cues to infer personality traits from. With our work we add to the personality extraction research by showing the relationship between personality traits and picture features on Instagram.

3 Materials

To investigate the relationship between personality traits and picture features, we asked participants to fill in the 44-item BFI personality questionnaire (5-point Likert scale; Disagree strongly - Agree strongly [7]). The questionnaire include questions that aggregate into the five basic personality traits of the FFM. Additionally, we asked participants to grant us access to their Instagram account through the Instagram API in order to crawl their pictures. From hereon, we define the picture-collection term as all the Instagram pictures of a single user.

We recruited 126 participants through Amazon Mechanical Turk. Participation was restricted to those located in the United States, and also to those with a very good reputation to avoid careless contributions. Several comprehension-testing questions were used to filter out fake and careless entries. The Mahalanobis distance was calculated to check for outliers. This left us with 113 completed and valid responses. Age (18-64, median 30) and gender (54 male, 59 female) information indicated an adequate distribution.

4 Features

For each picture in a picture-collection that was crawled, we extracted several features. The extracted features are discussed below. Most of the features are color-based, some are content-based. For color-based features we use the color space that is most closely related to the human visual system, i.e., the Hue-Saturation-Value (HSV) color space [15].

Brightness. For each picture, we calculated the average brightness and variance across all the pixels in the picture. Pictures that have a high average brightness tend to be bright, obviously. These features represent how light/dark a picture is and how much contrast there is in the picture, respectively. Pictures that have a high variance tend to have both dark and light areas, whereas pictures with a low variance tend to be equally bright across the picture. Furthermore, we divided the brightness axis into three equal intervals and counted the share of pixels that fall into each of these intervals (low/mid/high brightness). Pictures that have a high value in the *low brightness* feature tend to be darker, those that have a high value in the *mid brightness* feature tend to have mostly neither dark nor bright areas, while those pictures that have a high value in the *high brightness* feature tend to have lots of bright areas.

Saturation. We calculated the average saturation and the variance for each picture. Pictures with low average saturation tend to be bleak, colorless, while pictures with high saturation have more vivid colors. Pictures with a high saturation variance tend to have both bleak and vivid colors. Here we also divided the saturation axis into three equal intervals and calculated the share of pixels that fall into each interval (low/mid/high saturation). pictures that have a high value in the *low saturation* tend to have more bleak colors, those with a high value in the *mid saturation* feature tend to have neither bleak nor vivid colors while those pictures that have a high value in the *high saturation* feature tend to have vivid colors across most of the picture area.

Pleasure-Arousal-Dominance (PAD). As the filters on Instagram intend to create a certain expression, we adopted the PAD model of Valdez and

Merhabian [16]. They created general rules of the expression of pleasure, arousal, and dominance in a picture as a combination of brightness and saturation levels:

1. Pleasure = .69 Brightness + .22 Saturation
2. Arousal = -.31 Brightness + .60 Saturation
3. Dominance = -.76 Brightness + .32 Saturation

Hue-related features We extracted features that represented the prevalent hues in pictures. We chose many features that represent various aspects of the hues. For each of the basic colors (red, green, blue, yellow, orange, and violet) we counted the share of pixels that fall into each color. As the discrete color clustering of the hue dimension is nonlinear and subjective, we also divided the hue into 10 equal intervals and calculated the share of pixels for each interval. These intervals are hard to describe with subjective color descriptions. Furthermore, we calculated the share of pixels that fall into cold (violet, blue, green) and warm (yellow, red, orange) colors.

Content-based features. Beside the color-centric features we also performed picture content analysis. We counted the number of faces and the number of people in each picture. We used the standard Viola-Jones algorithm [17]. A manual inspection of the Viola-Jones face detector results revealed some false positives (e.g., a portrait within the picture) and false negatives (e.g., some rotated and tilted faces). However, in general the users who tended to take pictures of people (e.g., selfies) had a higher number of average number of faces/people per picture than those users who tended to take mostly still photographs.

5 Results

We crawled 22,398 pictures, and extracted all the features per picture for each picture-collection. As the features in the picture-collections show a symmetrical distribution, we calculated mean values for each feature to create a measurement of central tendency. The mean values of the features were used to calculate the correlation matrix (see Table 1). Pearson's correlation ($r \in$ [-1,1]) is reported to indicate the linear relationship between personality and picture features.

The correlation matrix shows several features related with personality traits. We will discuss the results related to each personality trait below. Besides significant correlations of p<.05, we decided to report marginally significant results as well (i.e., significant levels of .1 >p>.05).

Openness to experience. The openness factor correlates positively with the feature *green*, meaning that open users tend to take pictures with a lot of green or applied a filter to express more greenness. We observed a negative correlation with the feature *brightness mean*, which means that open users tend to upload pictures that are low on brightness. This was further confirmed by the positive correlation on *brightness low*, and the negative correlation on *brightness high*. These correlations show that the pictures of open users show more dark areas, and less bright areas.

Openness is also correlated with the *saturation mean*. This indicate that the pictures of open users consist of more saturated, vivid colors. We also observed a positive correlation with the feature *saturation variance*, which means that open users upload pictures that have both vivid and bleak colors.

A marginally significant correlation was observed for the warm/cold features. Pictures of open users contained less warm colors (i.e., red, orange), but more cold colors (i.e., blue, green). Also, the pictures of open users tend to express less pleasure, but more arousal and dominance. Additionally, their pictures consist of less faces and people.

Conscientiousness. A marginally significant (positive) correlation was found between the *saturation variance* feature and conscientiousness. This indicate that conscientious users upload pictures consisting of bleak and vivid colors.

Extraversion. We mostly found marginally significant correlations with the picture features and the extraversion. Extraverts tend to upload pictures with less *red* and *orange*, but with more *green* and *blue* tones. Additionally, their pictures tend to be darker (*brightness low*), but tend to consist of both vivid and bleak colors (*saturation variance*). Additionally, the emotion that the pictures of extraverts consist, are low on *pleasure*, but high on *dominance*.

Agreeableness. A marginally significant correlation was found between agreeableness and the *brightness mid* feature. This means that the pictures of agreeable users do not show emphasized bright or dark areas, but are more in between.

Neuroticism. A marginally significant correlation was found on *brightness mean*, *brightness low*, and *brightness high*. The positive correlation with *brightness mean* indicate that extravert users tend to upload pictures that are high on brightness. This is also reflected in the *brightness low* (negative correlation) and *brightness high* (positive correlation) features. Additionally,

	O	C	E	A	N
Red	-0.06	0.02	**-0.17**$^{\wedge}$	-0.05	0.03
Green	**0.17**$^{\wedge}$	0.14	**0.23**$^{\wedge\wedge}$	0.03	-0.12
Blue	-0.01	0	**0.17**$^{\wedge}$	0.02	-0.01
Yellow	0.01	0.04	0.01	0.14	-0.07
Orange	-0.03	-0.07	**-0.16**$^{\wedge}$	-0.02	0.06
Violet	0	-0.06	-0.09	-0.07	0.06
Bright.mean	**-0.25***	-0.1	**-0.19**$^{\wedge}$	-0.07	**0.22**$^{\wedge}$
Bright.var.	0.06	0	0	-0.07	0.05
Bright.low	**0.28****	0.09	**0.16**$^{\wedge}$	-0.05	**-0.16**$^{\wedge}$
Bright.mid	-0.09	0.06	0.04	**0.15**$^{\wedge}$	-0.06
Bright.high	**-0.2**$^{\wedge}$	-0.12	**-0.18**$^{\wedge}$	-0.08	**0.21**$^{\wedge}$
Sat.mean	**0.16**$^{\wedge}$	0.06	0.03	-0.04	0
Sat.var.	**0.2**$^{\wedge\wedge}$	**0.16**$^{\wedge}$	**0.19**$^{\wedge\wedge}$	0.1	-0.05
Sat.low	-0.08	-0.02	0.02	0.07	0.01
Sat.mid	0.08	-0.09	0.02	0.07	0.01
Sat.high	0.13	0.1	0.04	-0.01	0.01
Warm	**-0.05**$^{\wedge\wedge}$	-0.04	-0.2	0	0.03
Cold	**0.05**$^{\wedge\wedge}$	0.04	0.2	0	-0.03
Pleasure	**-0.19**$^{\wedge\wedge}$	-0.08	**-0.18**$^{\wedge}$	-0.09	**0.22**$^{\wedge\wedge}$
Arousal	**0.23***	0.09	0.1	0	-0.08
Dominance	**0.28****	0.11	**0.17**$^{\wedge}$	0.05	**-0.18**$^{\wedge\wedge}$
# of faces	**-0.16**$^{\wedge}$	0.03	0.11	-0.11	-0.03
# of people	**-0.22**$^{\wedge\wedge}$	-0.05	-0.07	-0.01	0.07

Note. $^{\wedge}p$<0.1, $^{\wedge\wedge}p$<.05, $^{*}p$<.01, $^{**}p$<.001.

Table 1: Correlation Matrix of the picture features against the personality traits: (O)penness, (C)onscientiousness, (E)xtraversion, (A)greeableness, (N)euroticism.

correlations were found in the emotion expression of the pictures of extraverts. They show to adjust their pictures to express more pleasure but less dominance.

6 Personality prediction

Given that we found significant correlations between picture features and personality traits, we explored personality prediction based on these features. We trained our predictive model with the *radial basis function network* classifier in Weka, with a 10-fold cross-validation. We report the *root-mean-square error* (RMSE) in Table 2. The RMSE of each personality trait relates to the [1,5] score scale.

Personality	**RMSE**
Openness to experience	0.73
Conscientiousness	0.69
Extraversion	0.95
Agreeableness	0.74
Neuroticism	0.95

Table 2: Personality prediction with the root-mean-square error (RMSE).

Personality	**Picture properties**
Openness to experience	Green, low brightness, high saturation, cold colors, few faces
Conscientiousness	Saturated and unsaturated colors
Extraversion	Green and blue tones, low brightness, saturated and unsaturated colors
Agreeableness	Few dark and bright areas
Neuroticism	High brightness

Table 3: Picture properties in relation to personality traits. The properties apply for the pictures of users who score high in the respective personality trait.

The RMSE values that we found are low and comparable with previous work on personality extraction from social media trails. For example, Quercia et al. [10] looked at the relationship between personality traits and Twitter usage and reported RSME scores of 0.69, 0.76, 0.88, 0.79, and 0.85, respectively for, openness to experience, conscientiousness, extraversion, agreeableness, and neuroticsm. Our results, as well as the results of prior work show that the most difficult traits to predict are the extraversion and the neuroticism personality traits.

7 Discussion

We found Instagram picture features to be correlated with personality. Results show that the most strongly significant correlations are found in the openness to experience personality trait. Although we found weaker significant levels for the other personality traits, we were still able to find distinct correlations. See Table 3 for a summary of our findings.

Based on the found correlations, we also explored the prediction of the personality traits based on the picture features. Compared with the findings of prior work (i.e., [10]), we were able to find similar results and patterns. The most successful personality traits to predict are openness to experience, conscientiousness, and agreeableness, whereas the more difficult traits are extraversion and neuroticism.

8 Limitations and future work

Our study contains limitations that need to be considered. Although we were able to obtain a fair amount of Instagram pictures (n=22,398), our personality measurement was limited to 113 participants. Given that we only had personality information of 113 participants to find relationships with picture features, we decided to give attention to the marginally significant results as well (i.e., significant levels of .1 $>p>$.05). A bigger sample size to assess personality traits should provide more conclusive results about the marginally significant effects that we found in this study.

In this study we solely focused on participants based in the United States. However, color interpretation and meaning could be influenced by cultural factors. Therefore, cultures could engage in different behavior of picture taking and applying filters. Future work should address this.

9 Conclusion

Our results suggest that there is a relation between personality traits and the way of taking and applying filters to pictures. By analyzing pictures of Instagram, personality traits can be inferred. Implicitly extracting personality from social media trails makes it possible to facilitate systems in order to provide a personalized experience. For example, it can help recommender systems to overcome the cold-start problem [14], or provide a personalized user interface [4].

10 Acknowledgments

This research is supported by the Austrian Science Fund (FWF): P25655; and by the EU FP7/2013-2016 through the PHENICX project under grant agreement no.601166.

References

[1] Back, M.D., Stopfer, J.M., Vazire, S., Gaddis, S., Schmukle, S.C., Egloff, B., Gosling, S.D.: Facebook profiles reflect actual personality, not self-idealization. Psychological science (2010)

[2] Ferwerda, B., Schedl, M.: Enhancing music recommender systems with personality information and emotional states: A proposal. In: Proceedings of the 2nd EMPIRE Workshop. pp. 1613–0073 (2014)

[3] Ferwerda, B., Schedl, M., Tkalcic, M.: Personality & emotional states: Understanding users' music listening needs. UMAP 2015 Extended Proceedings (2015)

[4] Ferwerda, B., Yang, E., Schedl, M., Tkalcic, M.: Personality traits predict music taxonomy preferences. In: CHI'15 Extended Abstracts. ACM (2015)

[5] Golbeck, J., Robles, C., Edmondson, M., Turner, K.: Predicting personality from twitter. In: IEEE Third Conference on Social Computing. pp. 149–156 (2011)

[6] Gosling, S.D., Gaddis, S., Vazire, S., et al.: Personality impressions based on facebook profiles. ICWSM

[7] John, O.P., Donahue, E.M., Kentle, R.L.: The big five inventory: institute of personality and social research. UC Berkeley, CA (1991)

[8] McCrae, R.R., John, O.P.: An introduction to the five-factor model and its applications. Journal of personality 60(2), 175–215 (1992)

[9] Park, G., Schwartz, H.A., Eichstaedt, J.C., Kern, M.L., Kosinski, M., Stillwell, D.J., Ungar, L.H., Seligman, M.E.: Automatic personality assessment through social media language. (2014)

[10] Quercia, D., Kosinski, M., Stillwell, D., Crowcroft, J.: Our twitter profiles, our selves: Predicting personality with twitter. In: IEEE Third Conference on Social Computing. pp. 180–185 (2011)

[11] Rawlings, D., Ciancarelli, V.: Music preference and the five-factor model of the neo personality inventory. Psychology of Music 25(2), 120–132 (1997)

[12] Ross, C., Orr, E.S., Sisic, M., Arseneault, J.M., Simmering, M.G., Orr, R.R.: Personality and motivations associated with facebook use. Computers in Human Behavior 25(2), 578–586 (2009)

[13] Tkalcic, M., Ferwerda, B., Hauger, D., Schedl, M.: Personality Correlates for Digital Concert Program Notes. UMAP pp. 1–6 (2015)

[14] Tkalcic, M., Kunaver, M., Košir, A., Tasic, J.: Addressing the new user problem with a personality based user similarity measure. In: DEMRA (2011)

[15] Tkalcic, M., Tasic, J.: Colour spaces: perceptual, historical and applicational background. IEEE EUROCON. Computer as a Tool pp. 304–308 (2003)

[16] Valdez, P., Mehrabian, A.: Effects of color on emotions. Journal of Experimental Psychology: General 123(4), 394 (1994)

[17] Viola, P., Jones, M.J.: Robust Real-Time Face Detection. International Journal of Computer Vision 57(2), 137–154 (May 2004)

Using Instagram Picture Features to Predict Users' Personality

Published in: Proceedings of the 22nd International Conference on MultiMedia Modeling (MMM 2016; Miami, FL, US)

BRUCE FERWERDA, Johannes Kepler University, Linz (AT)
MARKUS SCHEDL, Johannes Kepler University, Linz (AT)
MARKO TKALCIC, Free University of Bolzano, Bolzano (IT)

Abstract

Instagram is a popular social networking application, which allows photo-sharing and applying different photo filters to adjust the appearance of a picture. By applying these filters, users are able to create a style that they want to express to their audience. In this study we tried to infer personality traits from the way users manipulate the appearance of their pictures by applying filters to them. To investigate this relationship, we studied the relationship between picture features and personality traits. To collect data, we conducted an online survey where we asked participants to fill in a personality questionnaire, and grant us access to their Instagram account through the Instagram API. Among 113 participants and 22,398 extracted Instagram pictures, we found distinct picture features (e.g., relevant to hue, brightness, saturation) that are related to personality traits. Our findings suggest a relationship between personality traits and these picture features. Based on our findings, we also show that personality traits can be accurately predicted. This allow for new ways to extract personality traits from social media trails, and new ways to facilitate personalized systems.

Keywords: Instagram, Personality, Photo Filters, Picture Features

1 Introduction

Instagram is a popular mobile photo-sharing, and social networking application, with currently over 300 million active users a month, over 70 billion pictures shared, with an average of 70 million new pictures a day. [1] Instagram is interconnected with an abundance of social networking sites (e.g., Facebook, Twitter, Tumblr, and Flickr) to let its users share their pictures on.

[1] `https://instagram.com/press/` (accessed: 08/07/2015)

In addition, it encourages users to apply filters to modify the color appearance of their pictures. At this moment Instagram offers 25 predefined photo filters that allow users to customize and modify their pictures to create the desired visual style.

The ease with which a photo filter can be applied allow users to express a personal style and create a seeming distinctiveness with the customized pictures. Through the shared content and the way of applying filters, users are able to reveal a lot about themselves to their social network. With that, the question arises: What do Instagram pictures tell about the user? Or more specifically: What do Instagram pictures say about the personality of the user?

Personality traits have shown to consist of cues to infer users' behavior, preference, and taste (e.g., [9, 26, 28]). Hence, knowing one's personality can provide important information for systems to create a personalized user experience. It can provide systems with estimations about user preferences, and avoid the use of extensive questionnaires or observations.

There is an increasing interest in implement personality in systems (e.g., [8, 11, 29]), and the implicit acquisition of personality from online behavior trails (e.g., Facebook [2, 5, 13, 24, 27], Twitter [12, 25], Flickr [7], video blogs [3, 4]). In this work we join the personality extraction research. We specifically focus on the relationship between the personality of Instagram users and the way they manipulate their pictures by using photo filters, in order to create a visual style.

Our work makes several contributions. We contribute to personality research, by showing relationships between personality traits and the visual style of users' Instagram pictures. Additionally, we contribute to new ways to extract personality from social media (i.e., Instagram). To the best of our knowledge, we are the first to investigate the relationship between personality traits and how users try to create a visual style by applying photo filters.

An online survey was conducted where we: 1) asked participants to fill in the widely used Big Five Inventory (BFI) personality questionnaire, and 2) grant us access to the content of their Instagram account. We extracted 22,398 Instagram pictures of 113 users, and analyzed them on several color-centric picture features (e.g., related to hue, saturation, value). We found distinct correlations between personality traits and picture features, and show that personality can be accurately predicted from the picture features. [2]

[2]Our preliminary results of this work can be found in [10].

In the remainder of the paper we will continue with related work, materials, features, results, discussion, and conclusion.

2 Related work

Personality has shown to be an enduring factor that can be related to a person's taste, preference, and interest (e.g., [9, 26, 28]). For example, Rawlings and Ciancarelli found relationships between personality traits and music genre preferences [26], while Tkalcic et al. found relationships between personality and classical music [28]. These relationships indicate that personality information can be used to create useful proxy measures for applications to cater to a more personalized service (e.g., [8, 11, 29]). For example, Tkalcic et al. propose to use personality to enhance the nearest-neighborhood measurement for overcoming the cold-start problem (i.e., recommending items to new users) in recommender systems [29]. Ferwerda et al. provide a way to use personality for adjusting the user interface of music applications to fit a user's music browsing style [11].

In order to measure personality, several models have been developed. The five-factor model (FFM) is the most well known and widely used one in the computing community [32], and categorizes personality into five general dimensions (traits), that describe personality in terms of: openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism [21]. However, unless using extensive, and time-consuming questionnaires, acquiring personality traits is still a challenging task.

There is an emergent interest in how to implicitly acquire personality traits based on behavioral data (for an overview see [32]). Research has shown it is feasible to compute personality from behavioral data such as mobile phone usage (e.g.,[6, 23]), or with acoustic and visual cues through cameras and microphones (e.g., [1, 19, 22]). With the increasing connectedness of people, recent research has started to focus on personality acquisition from online behavior trails, such as, video blogs (vlogs) [3, 4], Facebook behavior (e.g., [2, 13, 24, 27]) and profile pictures [5], Twitter behavior (e.g., [12, 25]), and Flickr pictures [7].

Little work has been done on personality extraction from pictures. Celli et al. focused on the content of Facebook profile pictures (e.g., facial close-ups, facial expressions, alone or with others) to extract personality [5]. Other work of Cristani et al. showed that personality can be extracted from the visual features of Flickr pictures [7]. Flickr attracts a lot of advanced photographers as an image hosting platform, and thereby consist of more serious and higher quality pictures. Instagram, on the other hand, targets snapshot pictures taken with the mobile phone, and puts emphasis on applying predefined filters. The different usage and interactions on Flickr and Instagram attract different audiences, and therefore personality prediction may be based on different cues. As Instagram is known for its photo filters to create certain effects, we decided to focus more on color-properties; how users manipulate their pictures with help of the filters to achieve a certain expression, rather then picture content.

3 Materials

To investigate the relationship between personality traits and picture features, we asked participants to fill in the 44-item BFI personality questionnaire (5-point Likert scale; Disagree strongly - Agree strongly [16]). The questionnaire include questions that aggregate into the five basic personality traits of the FFM. The distribution of each personality trait can be found in Figure 1. Additionally, we asked participants to grant us access to their Instagram account through the Instagram API, in order to crawl their pictures. From hereon, we define the picture-collection term as *all* the Instagram pictures of a single user.

We recruited 126 participants through Amazon Mechanical Turk, a popular recruitment tool for user-experiments [18]. Participation was restricted to those located in the United States, and also to those with a very good reputation ($\geq$95% HIT approval rate and $\geq$1000 HITs approved) [3] to avoid careless contributions. Several comprehension-testing questions were used to filter out fake and careless entries. The Mahalanobis distance was calculated to check for outliers. This left us with 113 completed and valid responses. Age (18-64, median 30) and gender (54 male, 59 female) information indicated an adequate distribution. Pictures of each participant were crawled after the study. This resulted in a total of 22,398 pictures.

4 Features

As the goal in this study was to see how Instagram users manipulate their pictures with photo filters, we extracted mainly color related picture features. Based on the assumption that Instagram users' personality is manifested through the way photo filters are applied, a set of features that relate to color were selected based on the work of Machajdik and Hanbury [20]. For

[3] HITs (Human Intelligence Tasks) represent the assignments a user has participated in on Amazon Mechanical Turk prior to this study.

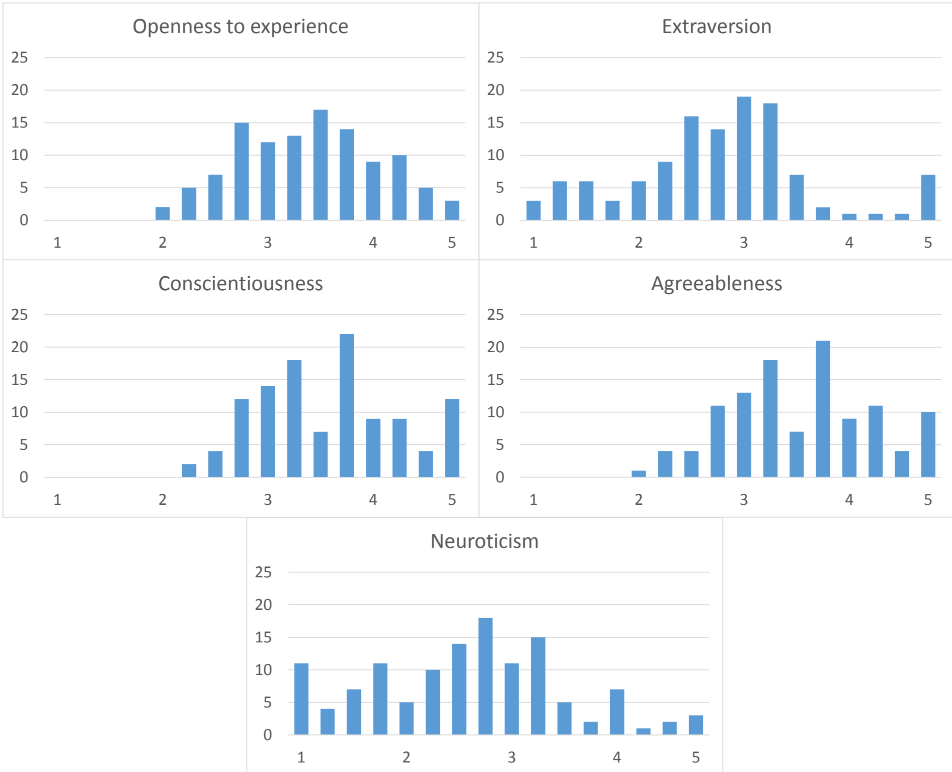

Figure 1: Collected personality distribution for each personality trait.

the color-centric features, the color space that is most closely related to the human visual system was selected. That is, the Hue-Saturation-Value (HSV) color space [30]. The H parameter describes the hue (i.e., the color quality of each pixel) from orange through yellow, green , blue, violet to red, on a scale from 0 to 1. The S parameter describes how saturated the color is. That is, the share of white in a picture (high share of white means low saturation). The parameter V (value) represents the brightness of the color. Beside the color-centric features, we additionally performed basic picture content analysis, by counting the number of faces and the number of people in each picture.

Hue-related features. We divided the range of the H parameter into intervals that correspond to the hues: orange, yellow, green, blue, violet, and red. For each of these intervals we counted the number of pixels in an image that fall into this interval and divided it with the number of all pixels in the image. This yielded the share of the image surface that a hue covers.

Furthermore, we merged the cold colors (i.e., green, blue violet) and the warm colors (i.e., orange, red, yellow) into the respective shares across an image. On a user level, the features: orange, yellow, green, blue, violet, red, warm and cold, are the average values of the color shares among all the pictures of a user.

Saturation-related features. For each picture we calculated the average saturation and the variance. Images with low average saturation tend to be bleak and colorless, while pictures with high saturation have more vivid colors. Pictures with a high saturation variance tend to have both bleak and vivid colors. Here, we also divided the saturation axis into three equally spaced intervals and calculated the share of pixels that fall into each interval (low, mid, and high saturation). Pictures that have a high value in the *low saturation* tend to have more bleak colors, those with a high value in the *mid saturation* feature tend to have neither bleak nor vivid colors, and those that have a high value in the *high saturation* feature tend to have vivid colors across most of the image area.

Value-related features. For each image we calculated the average value (*value mean*) and variance (*value variance*) across all the pixels in the image. These features represent how light or dark a picture is and how much contrast it reveals, respectively. Pictures that have a high variance tend to have both dark and light areas, whereas pictures with a low variance tend to be equally bright across the image. Furthermore, we divided the value axis into three equally spaced intervals and counted the share of pixels that fall into each of these intervals (*value low, mid, and high*).

Pleasure-Arousal-Dominance (PAD). As the filters Instagram users can apply to their pictures, are intended to create certain expressions, we adopted the PAD model of Valdez and Merhabian [31], which contains general rules of the expression of pleasure, arousal, and dominance in a picture, and models these as a combination of brightness (value) and saturation levels:

1. Pleasure = .69 Value + .22 Saturation
2. Arousal = -.31 Value + .60 Saturation
3. Dominance = -.76 Value + .32 Saturation

Content-based features. In addition to the color-centric features, we computed two extra features for each user: 1) the average number of human faces across all the images of a user, and 2) the average number of full people bodies across all the images of a user.

To extract the number of faces in a picture, we used the Viola-Jones algorithm [33]. We trained the algorithm to recognize the number of faces in an image, by using the Haar-like features and the AdaBoost classifier. For extracting the number of full bodies we used the Histogram of Oriented Gradient (HOG) features with a Support Vector Machine (SVM) classifier. To achieve this, we employed Matlab's Computer Vision System Toolbox. [4]

5 Results

We divided the results section into two parts: the first part discusses the correlations we found, and the second part we discuss our personality regressor to predict personality traits based on the picture features.

5.1 Correlations

Although the main goal of this study was to investigate the relationship between personality and picture features, we decided to explore whether we could find correlations with personality and the usage of certain filters. We shortly discuss our results on the latter, and then continue with the correlations of the picture features.

5.1.1 Instagram filters

Besides crawling the pictures, we also crawled descriptives about the filters each participant applied to their pictures, in order to explore whether certain personality traits are related to a more frequent use of some filters. Of the collected 22,398 pictures from 113 participants, 1487 unique filters were applied in total (this also include the "Normal" filter which means that no filter is applied). This brings that on average participants use 13 (13.15) different photo-filters to the pictures they upload to their picture-collection. Given that Instagram offers 25 different filters to apply, participants seem to use a whole range of filters. This is not so surprising, as the result of applying a filter depends on how the original picture looks like. So it is possible that the visual characteristics of a picture is eventually the same, but achieved by using different filters.

With using Pearson's correlation ($r \in$ [-1,1]) to indicate the linear relationship between personality and photo filters, we found the following correlations: a positive correlation between the conscientiousness personality traits and the "Kelvin" filter (r=.203, p=.044), a negative correlation between

[4] http://www.mathworks.com/help/vision/index.html

the agreeableness personality trait, the “Crema" (r=-.205, p=.042) and the “Gotham" (r=-.204, p=.042) filter. Additionally, we found a positive correlation between the neuroticism personality traits and the “Hudson" (r=.224, p=.026) filter. These results imply that in general conscientiousness participants make more use of the Kelvin filter, agreeable participants used less the Crema and Gotham filter, and neurotic participants applied more the Hudson filter. Besides these four significant correlations, there were no statistical significant correlations found with the remaining 21 photo filters. Given the low number of significant correlations that we found, and that the applied filter does not tell anything about how the end result looks like, we decided to not further pursue this direction, and continue our analyses focusing on the picture features.

5.1.2 Picture features

For each picture from a crawled picture-collection (i.e., all the pictures of a participant's Instagram account), we extracted all the features that are described in Section 4 (Features). As the features in participants' picture-collection show a normal distribution, we calculated mean values for each feature to create a measurement of central tendency, to represent the whole picture-collection of each participant. The mean values of the features were used to calculate the correlation matrix (see Table 1). Pearson's correlation ($r \epsilon$ [-1,1]) is reported to indicate the linear relationship between personality and picture features. The correlation matrix shows several features related with personality traits. [5] We discuss the results related to each personality trait below.

Openness to experience: Openness to experience was found to correlate with the *saturation mean*. This indicates that the pictures of open participants consist of more saturated, vivid colors. We also observed a positive correlation with the feature *saturation variance*, which means that open participants share pictures that have both vivid and bleak colors. Furthermore, a negative correlation with the feature *value mean* was found, indicating that open participants tend to share pictures that are low on brightness. This was further confirmed by the positive correlation on *brightness low*, and the negative correlation on *value high*. These correlations show that the pictures of open participants show more dark areas, and less bright areas. Also, a correlation was observed for the warm and cold features. Pictures of more open participants contained less warm colors (i.e., red, orange), but more cold colors (i.e., blue, green). Also, their pictures tend to

[5]The magnitude of the correlations are commonly found in relationships between personality traits and social media trails (e.g., [2, 12, 13, 24, 25, 27])

	O	C	E	A	N
Red	-0.06	0.02	**-0.17**^	-0.05	0.03
Green	**0.17**^	0.14	**0.23**^^	0.03	-0.12
Blue	-0.01	0	**0.17**^	0.02	-0.01
Yellow	0.01	0.04	0.01	0.14	-0.07
Orange	-0.03	-0.07	**-0.16**^	-0.02	0.06
Violet	0	-0.06	-0.09	-0.07	0.06
Saturation mean	**0.16**^	0.06	0.03	-0.04	0
Saturation variance	**0.20**^^	**0.16**^	**0.19**^^	0.10	-0.05
Saturation low	-0.08	-0.02	0.02	0.07	0.01
Saturation mid	0.08	-0.09	0.02	0.07	0.01
Saturation high	0.13	0.10	0.04	-0.01	0.01
Value mean	**-0.25***	-0.10	**-0.19**^	-0.07	**0.22**^
Value variance	0.06	0	0	-0.07	0.05
Value low	**0.28****	0.09	**0.16**^	-0.05	**-0.16**^
Value mid	-0.09	0.06	0.04	**0.15**^	-0.06
Value high	**-0.20**^	-0.12	**-0.18**^	-0.08	**0.21**^
Warm	**-0.05**^^	-0.04	-0.20	0	0.03
Cold	**0.05**^^	0.04	0.20	0	-0.03
Pleasure	**-0.19**^^	-0.08	**-0.18**^	-0.09	**0.22**^^
Arousal	**0.23***	0.09	0.10	0	-0.08
Dominance	**0.28****	0.11	**0.17**^	0.05	**-0.18**^^
# of faces	**-0.16**^	0.03	0.11	-0.11	-0.03
# of people	**-0.22**^^	-0.05	-0.07	-0.01	0.07

Note. ^p<0.1, ^^p<.05, *p<.01, **p<.001.

Table 1: Correlation Matrix of the picture features against the personality traits: (O)penness, (C)onscientiousness, (E)xtraversion, (A)greeableness, (N)euroticism.

express less *pleasure*, but more *arousal* and *dominance*. Additionally, their pictures consist in general of fewer faces and people.

Conscientiousness: A positive correlation was found between the *saturation variance* feature and conscientiousness. This indicates that conscientious participants more frequently shared pictures consisting of bleak and vivid colors.

Extraversion: We found correlations between the picture features and extraversion. Extraverts tend to create pictures with less *red* and *orange*, but with more *green* and *blue* tones. Additionally, their pictures tend to be darker (*brightness low*), but consist of both vivid and bleak colors (*saturation variance*). Also the emotion that the pictures of extraverts consist, are low on *pleasure*, but high on *dominance*.

Agreeableness: A positive correlation was found between agreeableness and the *brightness mid* feature. This means that the pictures of agreeable participants do not show emphasized bright or dark areas, but are more in between.

Neuroticism: A correlation was found on *value mean*, *value low*, and *value high*. The positive correlation with *value mean* indicate that participants scoring higher on the neurotic trait tend to share pictures that are high on brightness. This is also reflected in the *value low* (negative correlation) and *value high* (positive correlation) features. Additionally, correlations were found in the emotion expression of the pictures of extraverts. Result show that they adjust their pictures to express more pleasure but less dominance.

5.2 Personality Regressor

Given the correlations that were found, we developed a personality regressor based on the features reported in Table 1. As prior work on prediction personality traits from Twitter behavior uses the same method of analyses [25], we compare our performance against their results. We trained our predictive model with several classifiers in Weka, with a 10-fold cross-validation with 10 iterations. For each classifier we used, we report the *root-mean-square error* (RMSE) in Table 2, to indicate the root mean square difference between predicted and observed values. The RMSE of each personality trait relates to the [1,5] score scale.

	RMSE			
Personality traits	**Radial Basis Function Network**	**Random Forests**	**M5 Rules**	**M5 Rules Quercia et al. [25]**
Openness to experience	**0.68**	0.71	0.77	0.69
Conscientiousness	**0.66**	**0.67**	**0.73**	0.76
Extraversion	0.90	0.95	0.96	0.88
Agreeableness	**0.69**	**0.71**	**0.78**	0.79
Neuroticism	0.95	1.01	0.97	0.85

Table 2: Comparison of different classifiers to predict personality prediction compared to prior work of Quercia et al. [25]. Numbers in bold represent the results that outperform prior work. Root-mean-square error (RMSE) is reported ($r \in$ [1,5]).

In line with prior work of Quercia et al. [25], we started to train our predictive model with the *M5' rules* [34]. Although our results do not outperform

Personality	Picture properties
Openness to experience	More green tones, lower in brightness, higher in saturation, more cold colors, fewer faces and people
Conscientiousness	Mix of saturated and unsaturated colors
Extraversion	More green and blue tones, lower in brightness, mix of saturated and unsaturated colors
Agreeableness	Fewer dark and bright areas
Neuroticism	Higher in brightness

Table 3: Interpretation and summary of the correlations found between personality traits and picture properties. The properties apply for the pictures of participants who score high in the respective personality trait.

prior work on most facets, we do find similar trends: extraversion and neuroticism are the hardest personality traits to predict. As applying M5' rules did not result in any improvement, we applied the *random forests* classifier. Random forests are known to have a reasonable performance when the features consist of high amounts of noise [15]. The results show slight improvement over M5' rules in general, but for the neuroticism personality trait, the prediction got worse. Finally, we tried using the radial basis function (RBF) network, which is a neural network that has shown to work well on smaller datasets [17]. Results show that the RBF network outperforms the M5' rules, as well as the random forest classifier. Compared to prior work, the RBF network outperforms prediction of openness to experience, conscientiousness, and agreeableness. Also with the RBF network, predictions is most difficult for extraversion and neuroticism. Although we do not outperform prior work on extraversion and conscientiousness, our results do not differ much. In general, we show that with picture features we can achieve better personality prediction than prior work on Twitter data.

6 Discussion

We found Instagram picture features to be correlated with personality. A summary and interpretation of the picture features can be found in Table 3. We found that most correlations appear in the openness to experience personality trait. Even though, less and weaker correlations were found for the other personality traits, we were still able to observe distinct correlations.

Based on the identified correlations between personality and picture features, we created a personality regressor. The results that we achieved with our prediction model show that personality can be accurately predicted from picture features on Instagram. The results of the personality regressor show similar patterns as prior work on personality extraction from social media (i.e., Twitter) [25], and were able to outperform in predicting most of the personality traits. We found that the easiest and most successful prediction are for the openness to experience, conscientiousness, and agreeableness personality traits, whereas the more difficult traits are extraversion and neuroticism.

7 Future work & Limitations

Our study contains limitations that need to be considered. Although we were able to obtain a fair amount of Instagram pictures (n=22,398), our personality measurement was limited to 113 participants. Given that we only had personality information of 113 participants to find relationships with picture features, results would benefit from a bigger sample size.

In this study we solely focused on participants based in the United States. However, color interpretation and meaning could be influenced by cultural factors. Therefore, cultures could engage in different behavior of picture taking [14] and applying filters. Future work should address this. Furthermore, we decided to mainly focus on the color-centric features of the pictures we crawled, as Instagram's main focus is on applying photo filters. However, content analyses of the pictures would be an interesting next step to conduct.

8 Conclusion

With this study we show that personality of Instagram users can be accurately predicted from color-centric features of the pictures they post. Being able to implicitly extract personality from social media trails, gives possibilities to facilitate systems in order to provide a personalized experience. For example, it can help recommender systems to overcome the cold-start problem.

9 Acknowledgments

This research is supported by the Austrian Science Fund (FWF): P25655; and by the EU FP7/'13-'16 through the PHENICX project, grant agreement no.601166.

References

[1] Aran, O., Gatica-Perez, D.: Cross-domain personality prediction: from video blogs to small group meetings. In: Proceedings of the 15th ACM on International conference on multimodal interaction. pp. 127–130. ACM (2013)

[2] Back, M.D., Stopfer, J.M., Vazire, S., Gaddis, S., Schmukle, S.C., Egloff, B., Gosling, S.D.: Facebook profiles reflect actual personality, not self-idealization. Psychological science (2010)

[3] Biel, J.I., Aran, O., Gatica-Perez, D.: You are known by how you vlog: Personality impressions and nonverbal behavior in youtube. In: ICWSM. Citeseer (2011)

[4] Biel, J.I., Gatica-Perez, D.: The youtube lens: Crowdsourced personality impressions and audiovisual analysis of vlogs. Multimedia, IEEE Transactions (2013)

[5] Celli, F., Bruni, E., Lepri, B.: Automatic personality and interaction style recognition from facebook profile pictures. In: Proceedings of the ACM International Conference on Multimedia. pp. 1101–1104. ACM (2014)

[6] Chittaranjan, G., Blom, J., Gatica-Perez, D.: Mining large-scale smartphone data for personality studies. Personal and Ubiquitous Computing 17(3), 433–450 (2013)

[7] Cristani, M., Vinciarelli, A., Segalin, C., Perina, A.: Unveiling the multimedia unconscious: Implicit cognitive processes and multimedia content analysis. In: Proceedings of the 21st ACM international conference on Multimedia (2013)

[8] Ferwerda, B., Schedl, M.: Enhancing music recommender systems with personality information and emotional states: A proposal. In: Proc. EMPIRE workshop (2014)

[9] Ferwerda, B., Schedl, M., Tkalcic, M.: Personality & emotional states: Understanding users' music listening needs. UMAP 2015 Extended Proceedings (2015)

[10] Ferwerda, B., Schedl, M., Tkalcic, M.: Predicting personality traits with instagram pictures. In: Proceedings of the 3rd Workshop on Emotions and Personality in Personalized Systems 2015. pp. 7–10. ACM (2015)

[11] Ferwerda, B., Yang, E., Schedl, M., Tkalcic, M.: Personality traits predict music taxonomy preferences. In: CHI'15 Extended Abstracts. ACM (2015)

[12] Golbeck, J., Robles, C., Edmondson, M., Turner, K.: Predicting personality from twitter. In: IEEE Third Conference on Social Computing. pp. 149–156 (2011)

[13] Gosling, S.D., Gaddis, S., Vazire, S., et al.: Personality impressions based on facebook profiles. ICWSM (2007)

[14] Huang, C.M., Park, D.: Cultural influences on facebook photographs. International Journal of Psychology 48(3), 334–343 (2013)

[15] Humston, E.M., Knowles, J.D., McShea, A., Synovec, R.E.: Quantitative assessment of moisture damage for cacao bean quality using two-dimensional gas chromatography combined with time-of-flight mass spectrometry and chemometrics. Journal of Chromatography A 1217(12), 1963–1970 (2010)

[16] John, O.P., Donahue, E.M., Kentle, R.L.: The big five inventory: institute of personality and social research. UC Berkeley, CA (1991)

[17] Khot, L.R., Panigrahi, S., Doetkott, C., Chang, Y., Glower, J., Amamcharla, J., Logue, C., Sherwood, J.: Evaluation of technique to overcome small dataset problems during neural-network based contamination classification of packaged beef using integrated olfactory sensor system. LWT-Food Science and Technology (2012)

[18] Kittur, A., Chi, E.H., Suh, B.: Crowdsourcing user studies with mechanical turk. In: Proceedings of the SIGCHI conference on human factors in computing systems. pp. 453–456. ACM (2008)

[19] Lepri, B., Subramanian, R., Kalimeri, K., Staiano, J., Pianesi, F., Sebe, N.: Connecting meeting behavior with extraversion&# x2014; a systematic study. Affective Computing, IEEE Transactions on 3(4), 443–455 (2012)

[20] Machajdik, J., Hanbury, A.: Affective image classification using features inspired by psychology and art theory. In: Proceedings of the international conference on Multimedia. pp. 83–92. ACM (2010)

[21] McCrae, R.R., John, O.P.: An introduction to the five-factor model and its applications. Journal of personality 60(2), 175–215 (1992)

[22] Mohammadi, G., Vinciarelli, A.: Automatic personality perception: Prediction of trait attribution based on prosodic features. Affective Computing, IEEE Transactions on 3(3), 273–284 (2012)

[23] de Montjoye, Y.A., Quoidbach, J., Robic, F., Pentland, A.S.: Predicting personality using novel mobile phone-based metrics. In: Social computing, behavioral-cultural modeling and prediction, pp. 48–55. Springer (2013)

[24] Park, G., Schwartz, H.A., Eichstaedt, J.C., Kern, M.L., Kosinski, M., Stillwell, D.J., Ungar, L.H., Seligman, M.E.: Automatic personality assessment through social media language. (2014)

[25] Quercia, D., Kosinski, M., Stillwell, D., Crowcroft, J.: Our twitter profiles, our selves: Predicting personality with twitter. In: IEEE Third Conference on Social Computing. pp. 180–185 (2011)

[26] Rawlings, D., Ciancarelli, V.: Music preference and the five-factor model of the neo personality inventory. Psychology of Music 25(2), 120–132 (1997)

[27] Ross, C., Orr, E.S., Sisic, M., Arseneault, J.M., Simmering, M.G., Orr, R.R.: Personality and motivations associated with facebook use. Comput Hum Behav (2009)

[28] Tkalcic, M., Ferwerda, B., Hauger, D., Schedl, M.: Personality Correlates for Digital Concert Program Notes. UMAP pp. 1–6 (2015)

[29] Tkalcic, M., Kunaver, M., Košir, A., Tasic, J.: Addressing the new user problem with a personality based user similarity measure. In: DEMRA (2011)

[30] Tkalcic, M., Tasic, J.: Colour spaces: perceptual, historical and applicational background. IEEE EUROCON. Computer as a Tool pp. 304–308 (2003)

[31] Valdez, P., Mehrabian, A.: Effects of color on emotions. Journal of Experimental Psychology: General 123(4), 394 (1994)

[32] Vinciarelli, A., Mohammadi, G.: A survey of personality computing. Affective Computing, IEEE Transactions on 5(3), 273–291 (2014)

[33] Viola, P., Jones, M.J.: Robust Real-Time Face Detection. INT J COMPUT VISION 57(2), 137–154 (May 2004)

[34] Witten, I.H., Frank, E.: Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann (2005)

Fusing Social Media Cues: Personality Prediction from Twitter and Instagram

Published in: Companion Proceedings of the 25th World Wide Web Conference (WWW 2016; Montreal, QC, CA)

MARCIN SKOWRON, Johannes Kepler University, Linz (AT) &
Austrian Research Institute for Artificial Intelligence, Vienna (AT)
BRUCE FERWERDA, Johannes Kepler University, Linz (AT)
MARKO TKALCIC, Free University of Bolzano, Bolzano (IT)
MARKUS SCHEDL, Johannes Kepler University, Linz (AT)

Abstract

Incorporating users' personality traits has shown to be instrumental in many personalized retrieval and recommender systems. Analysis of users' digital traces has become an important resource for inferring personality traits. To date, the analysis of users' explicit and latent characteristics is typically restricted to a single social networking site (SNS). In this work, we propose a novel method that integrates text, image, and users' meta features from two different SNSs: Twitter and Instagram. Our preliminary results indicate that the joint analysis of users' simultaneous activities in two popular SNSs seems to lead to a consistent decrease of the prediction errors for each personality trait.

Keywords: Social Media Mining, User Modeling, Personality Computing

1 Introduction and Related Work

In recent years, social networking sites (SNSs) have become a popular means for information exchange and social interactions. Users' presence and their online activities spread across different platforms, each with their own interactive and content-oriented characteristics. The user-generated content has been shown to yield important insights into users' interests, preferences, and sentiments toward various topics. However, to date, the analysis of users' explicit and latent characteristics has been typically restricted to a single SNS.

Personality is a psychological construct accounting for individual differences in people [9]. The most widely used *Five-Factor Model* comprises five traits:

Openness, Conscientiousness, Extraversion, Agreeableness and *Neuroticism* [6]. As users leave rich traces of their digital activities through SNSs, the collected data has become an important resource for inferring users' personality traits. The research domain has been gathered under the umbrella of *personality computing* [9]. One of the earliest works was published by Golbeck et al. [2], in which a variety of features related to a user's social network membership and language use in microblogs were investigated. Quercia et al. [7] applied features related to the user's activity and reputation in Twitter. Kosinski et al. [4] have collected a large dataset of Facebook users and demonstrated the predictive value of users' *Likes* to infer personality traits. The frequently applied approaches focused on data from a single SNS. Nowadays, as users often provide clues that enable the linkage between their respective profiles in different SNSs, there is an increase of amount of multi-network data. In this work, we present a range of multimodal personality regressors incorporating user information from two SNSs and evaluate them against regressors trained on data acquired from a single SNS.

2 Method

Foundational concepts of social network research, including Lewin's field theory, motivate joint analysis of different networks of an individual, including the relations between personality traits and their manifestations in different SNSs. Previous studies have demonstrated that the visual [1], linguistic [2] and meta [7] features extracted from users' generated online content can be instrumental for inferring personality traits. As the present work analyses users' activities in two popular SNSs, Instagram and Twitter, primarily used to share images and text, we developed a pipeline that extracts image, linguistic, and users' meta features related to their reputation and influence. For Instagram images, the annotations include: Pleasure-Arousal-Dominance (PAD), brightness, saturation, hue-related and content-based features such as person's face or full body, drawing inspiration from features used in the emotion detection technique proposed in [5]. For extraction of linguistic features from tweets and Instagram image captions, we adapted our annotation system [8], that integrates natural language processing resources (e.g., LIWC, ANEW) and classifiers (e.g., Dialog Acts, Sentiment). We also included users' reputation and influence meta features based on users' publicly available counts: number of followers and followees, 'Klout' and adaptation of 'TIME' influence scores [7].

To obtain compact feature representations of the accessible information, we compute for each extracted feature: mean, standard deviation, minimum,

	O	C	E	A	N	AVG
B	0.75	0.75	1.04	0.67	0.98	1.14
T_m	0.74	0.78	1.16	0.71	1.07	0.89
I_i	0.80	0.70	0.98	0.74	0.90	0.83
T_{lm}	0.75	0.73	0.92	0.71	0.80	0.78
$T_m I_i$	0.62	0.66	0.92	0.69	0.92	0.77
I_l	0.62	0.66	0.92	0.69	0.92	0.76
T_l	0.73	0.66	0.96	0.63	0.75	0.75
$T_m I_l$	0.65	0.68	0.86	0.60	0.79	0.72
$T_l I_l$	0.61	0.68	0.86	0.63	0.76	0.71
$T_m I_{li}$	**0.51**	0.68	0.86	**0.55**	0.88	0.70
$T_l I_{li}$	0.64	**0.65**	0.87	**0.55**	**0.73**	0.69
$T_{lm} I_{li}$	0.53	0.67	**0.71**	0.56	0.83	**0.66**

Table 1: RMSE with features from (T)witter and (I)nstagram; and (B)aseline score: average value for a dimension. Subscripts indicate sets of features used: (l)inguistic, i(mage), (m)eta.

maximum and median. We address the curse of dimensionality and noise reduction through subsampling with the F-statistic. We use random forest regression to build a low variance and low bias model of personality trait characteristics by averaging over regression tree decisions. The variable importance ranking induced by random forests further reduces the number of features considered for each personality trait.

2.1 Data collection

We recruited native English SNS users of high reputation located in the United States of both, Instagram and Twitter, via Amazon Mechanical Turk. An administered questionnaire asking for their informed consent included the 44-item Big Five Inventory personality questionnaire [3] and quality assurance cross-checks and comprehension questions. The aggregated answers were used to infer participants' five basic personality traits (openness to experience, conscientiousness, extraversion, agreeableness, neuroticism); each trait scored from 1 to 5. We then crawled data from participants' SNSs accounts, filtering out those with fewer than 30 Instagram images or 30 tweets: our final set comprises of 62 users with sufficient amount of data in both SNSs.

	RMSE		MAE		PCC	
	T [7]	**$T_{Im}I_{li}$**	**T** [2]	**$T_{Im}I_{li}$**	**FB** [4]	**$T_{Im}I_{li}$**
O	0.69	**0.51**	0.12	**0.11**	0.43	**0.74**
C	0.76	**0.67**	0.14	**0.11**	0.29	**0.76**
E	0.88	**0.71**	**0.16**	0.17	0.40	**0.65**
A	0.79	**0.50**	0.12	0.12	0.30	**0.34**
N	0.85	**0.73**	0.19	**0.16**	0.30	**0.71**
AVG	0.79	**0.73**	0.15	**0.13**	0.30	**0.64**

Table 2: Personality traits regression accuracy of $T_{Im}I_{li}$ along the state-of-the-art systems inferred from different SNSs: (T)witter, (I)nstagram, FB - Facebook, in terms of RMSE, MAE - Mean Absolute Error, PCC - Pearson Correlation Coefficient.

3 Results and Conclusions

Table 1 presents personality regressor performances over different sets of features, using root-mean square error (RMSE) calculated over 5 independent, 10-fold cross-validation runs, one for each personality trait. Results indicate that both, the set of features selected and the choice of the SNS or their combination, yield differences in regressor performance. The best results, overall and for each personality trait, are obtained integrating features extracted from both SNSs. The overall best performing regressor, $T_{lm}I_{li}$, uses a complete set of features (linguistic, image, and meta). It is also most informative for Extraversion (RMSE: 0.71). Conscientiousness (RMSE: 0.65), Agreeableness (RMSE: 0.55) and Neuroticism (RMSE: 0.73) are best regressed with the combination of linguistic features of tweets and captions with image features, while Openness (RMSE: 0.51) using Twitter's linguistic and meta plus Instagram image features. As there is no common, publicly available personality data-set that includes data from multiple SNSs, in Table 2 we present the results of the $T_{lm}I_{li}$ regressor along the state-of-the-art personality regression systems over different data-sets. In the table we adopt the metrics used in the original papers [7, 4, 2]. Note that each study used different data sets acquired from different SNSs, and while the presented results provide insights on the performance of each regressor, different regressors are not directly comparable.

Our novel approach to Internet user personality recognition integrates user data from two SNSs. The presented method can easily be adapted to other SNSs and for inferring other types of explicit and latent user characteristics. Our first results indicate the potential of the proposed joint, multi-modal analysis of user-generated data from different SNSs, justifying the ongoing detailed investigation, which also addresses theoretical issues such as soundness conditions of such integration and includes a large scale evaluation of the inferred users' personality traits scores applied in recommendation systems.

4 Acknowledgments

This research is supported by the Austrian Science Fund (FWF): P25655, P27530-N15.

References

[1] Celli, F., Bruni, E., Lepri, B.: Automatic Personality and Interaction Style Recognition from Facebook Profile Pictures. In: Proc. ACM Multimedia -MM '14'. pp. 1101–04. ACM (2014), http://dl.acm.org/citation.cfm?doid=2647868.2654977

[2] Golbeck, J., Robles, C., Edmondson, M., Turner, K.: Predicting Personality from Twitter. In: Proc. PASSAT and SocialCom. pp. 149–156. IEEE (2011), http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6113107

[3] John, O.P., Donahue, E.M., Kentle, R.L.: The big five inventory—versions 4a and 54. UCB, Inst. of Personality and Social Research, Berkeley, CA (1991)

[4] Kosinski, M., Stillwell, D., Graepel, T.: Private traits and attributes are predictable from digital records of human behavior. PNAS 110(15), 5802–5 (2013), http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3625324\&tool=pmcentrez\&rendertype=abstract

[5] Machajdik, J., Hanbury, A.: Affective image classification using features inspired by psychology and art theory. In: Proc. ACM Multimedia - MM '10. pp. 83–92. ACM Press (2010), http://portal.acm.org/citation.cfm?doid=1873951.1873965

[6] McCrae, R.R., John, O.P.: An Introduction to the Five-Factor Model and its Applications. J. Pers. 60(2), 175 – 215 (1992)

[7] Quercia, D., Kosinski, M., Stillwell, D., Crowcroft, J.: Our twitter profiles, our selves: Predicting personality with twitter. In: Proc. PASSAT and SocialCom. pp. 180–185. IEEE (Oct 2011), http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6113111

[8] Skowron, M., Theunis, M., Rank, S., Borowiec, A.: Effect of affective profile on communication patterns and affective expressions in interactions with a dialog system. In: ACII 2011, pp. 347–356. Springer (2011)

[9] Vinciarelli, A., Mohammadi, G.: A Survey of Personality Computing. IEEE Trans. Aff. Comp. 5(3), 273–291 (2014), http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6834774

7 User-Centric Evaluation of a Music Recommender System

Ferwerda, B., Graus, M., Vall, A., Tkalcic, M., & Schedl, M. (2017) How Item Discovery Enabled by Diversity Leads to Increased Recommendation List Attractiveness. *In Submission*.

Ferwerda, B., Graus, M., Vall, A., Tkalcic, M., & Schedl, M. (2016) The Influence of Users' Personality Traits on Satisfaction and Attractiveness of Diversified Recommendation Lists. *In Extended Proceedings of the 10th ACM Conference on Recommender Systems: 4th Workshop on Emotions and Personality in Personalized Systems* (Boston, MA, US).

How Item Discovery Enabled by Diversity Leads to Increased Recommendation List Attractiveness

Published in: In Submission (2017)

BRUCE FERWERDA, Johannes Kepler University, Linz (AT)
MARK GRAUS, Eindhoven University of Technology, Eindhoven (NL)
ANDREU VALL, Johannes Kepler University, Linz (AT)
MARKO TKALCIC, Free University of Bolzano, Bolzano (IT)
MARKUS SCHEDL, Johannes Kepler University, Linz (AT)

Abstract

Applying diversity to a recommendation list has been shown to positively influence the user experience. A higher perceived diversity is argued to have a positive effect on the attractiveness of the recommendation list and on the difficulty to make a choice. In a user study we presented 100 participants with several personalized lists of recommended music artists varying in levels of diversity. Participants were asked to assess these lists on perceived diversity and attractiveness, the experienced choice difficulty and discovery (i.e., the extent the list enriches their taste). We found that recommendation list attractiveness is influenced by two effects: 1) by diversity mediated through discovery; diverse recommendation lists are perceived more attractive if they enrich the user's taste or 2) by the list familiarity; a higher list familiarity contributes to a higher list attractiveness. We additionally revealed how individual differences (i.e., personality and familiarity) moderate the effects found.

Keywords: Diversity, Recommender Systems, User-Centric Evaluation

1 Introduction

Recommender systems are usually designed in such a way that they provide the most relevant items to the user. However, this often results in a set of recommendations that are too similar to each other and thereby not covering the full spectrum of the user's interest [1]. Diversifying the recommendations can positively influence the user experience of the user [6, 11, 12, 13]. The diversification of the recommendations has been shown to contribute to the attractiveness of the recommendation list, which can reduce the choice difficulty and increase the user's choice satisfaction [6, 11].

Although recommendation diversification has been shown to have a positive influence on the perceived attractiveness of the recommendation list. The reason behind this phenomenon has not yet been fully investigated though. In this paper we take a deeper look at what leads to a higher attractiveness of the recommendation list, which subsequently leads to a lower choice difficulty. By testing differently diversified music recommendation lists, we show that the attractiveness of the lists can be increased by list diversity *if* the items contribute to enriching the taste of the user. However, although the effect is weaker, list attractiveness can also be increased by presenting a list that users are familiar with. Next to the known moderators, such as expertise and preference strength [9, 12], we show that users' personality plays a significant moderating role on aspects of the user experience. Our findings provide new insights on how to effectively create recommendation lists in order to increase list attractiveness and thereby decrease user's choice difficulties.

2 Related Work

Diversification of recommendations has been shown to influence the user experience. Too similar recommendations or lack of diversity can have negative consequences [1]. Ziegler et al. [13] argued that recommendation list diversity can have positive effects on the user perception of the recommendation lists, and proposed a diversity algorithm to vary the recommendation list diversity without affecting the prediction accuracy. Willemsen et al. [11, 12] investigated the influence of diversity on movie recommendations and found that diversity has a positive effect on the attractiveness of the recommendation set, the difficulty to make a choice, and eventually on the choice satisfaction.

Personal characteristics (e.g., preference strength and expertise [9, 12]) have been identified as moderators that influence the relationship between diversification and the user experience. Another personal characteristic that may play a role is the user's personality. Personality traits have been shown to be an enduring factor with relationships to one's taste, preference, and interest (e.g., [2, 10]), and therefore lends itself to create user models with. The five-factor model (FFM) is the most known and widely used model, which categorizes personality into five general dimensions that describe personality in terms of: openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism [5].

3 Data Preparation & Procedures

We created differently diversified music recommendation lists for participants in order to investigate the effects underlying the increase of recommendation list attractiveness. Since we created the recommendation lists off-line, we separated the study in two parts. In the first part participants were recruited and their *complete* Last.fm listening history was crawled in order to create the recommendation lists. After the lists were created, participants from the first part were invited for the second part where they were asked to assess the diversified recommendation lists.

We recruited 254 participants through Amazon Mechanical Turk for the first part of the study. Participation was restricted to those located in the United States with a very good reputation ($\geq$95% HIT approval rate and $\geq$1000 HITs approved) and a Last.fm account with at least 25 listening events. Furthermore, they were asked to fill in the 44-item Big Five Inventory personality questionnaire [5] to measure the FFM. Control questions were asked to filter out fake and careless contributions. A compensation of $1 was provided. We crawled the complete listening history of each participant and aggregated the listening events to represent artist and playcount (i.e., number of times listened to an artist).

In order to prepare the music recommendation lists for each participant, we complemented our data with the LFM-1b dataset [8]. [1] This dataset consists of the complete listening histories of 120,322 Last.fm users from different countries. Since our participants were all located in the United States, we only used United States users of the LFM-1b dataset to complement our dataset with. This resulted in 10,255 additional users, which we also aggregated into artist and playcount for each user. The final dataset consists of user, artist, and artist playcount triplets with a total of 387,037 unique artists for the creation of the recommendation lists.

We used the weighted matrix factorization algorithm by [4] on our final dataset to calculate the recommended items. This algorithm is specifically designed to deal with datasets consisting of implicit feedback (e.g., artist playcounts). We optimized the factorization hyper-parameters by conducting grid-search and picking the setting that yielded the best 5-fold cross-validated mean percentile rank. Specifically, using 20 factors, confidence scaling factor α=40, regularization weight λ=1000 and 10 iterations of alternating least squares, we achieved the best 5-fold cross-validated mean percentile rank of 1.78%. [2] Afterwards we factorized the whole user-artist

[1] Available at `http://www.cp.jku.at/datasets/LFM-1b/`

[2] See [4] for details on the hyper-parameters and the definition of the mean percentile rank metric.

triplets using this set of hyper-parameters.

The recommended items were diversified as was done in [11, 12] by using the method of [13]. By using the latent features as the basis of diversification instead of additional metadata like genre information (as is done in content-based recommender systems) guarantees that diversity is manipulated in line with user preferences. Previous research demonstrated that this way of diversifying recommendations is perceived accordingly by users [11].

A greedy selection to optimize the intra-list similarity [1] was run on the top 200 recommended artists (i.e., the 200 artists with highest predicted relevance) to maximize the distances between item vectors in the matrix factorization space. This algorithm starts with a recommendation set consisting of the artist with highest predicted relevance. In an iterative fashion items are added to the recommendation set until it contains 10 items.

In each step of the iteration, for each candidate item i the sum of all distances from its item vector to each item vector in the recommendation set is calculated: $c_i = \sum_{j=1}^{z} d(i,j)$, where z is the number of items in the recommendation set and $d(i,j)$ is the Euclidean distance between two item vectors i and j. All candidate items are ranked based on decreasing value of c_i (P_{c_i}) and on predicted relevance (P_{r_i}). A weighting factor β is introduced to balance the trade-off between predicted relevance and diversity. For each candidate item the combined rank is calculated following $w_i^* = \beta * P_{c_i} + (1 - \beta) * P_{r_i}$. The item with the highest combined rank is added to the recommendation set and the next step is taken until 10 items are selected.

β was manipulated to achieve different levels of diversification. In the described implementation β=1 corresponds to maximum diversity, β=0 corresponds to maximum predicted relevance. We compared recommendation lists for different values of β in terms of the sum of distances between the latent features scores of items in the recommendation set and their average range. The list for β=0.4 showed to fall halfway between maximum relevance and maximum diversity. Thus, the final β levels for diversification were set at β=0 (low), β=0.4 (medium), and β=1 (high).

After the recommendation lists were created, emails were sent out to all participants to invite them for the second part of the study. We created a login screen so that we could retrieve the personalized recommendation lists for each participant. After the log in, the participant was sequentially presented with a recommendation list for three times, with each time a different level of diversity (i.e., low, medium, or high. The order of presentation was randomized). Each recommended artist was enriched with metadata from Last.fm (i.e., picture, genre, top 10 songs with the number of listeners and playcounts), which was shown when hovered over the name in the list. Additionally, example songs were provided by clicking on the artist name (new browser screen linked to the artist's YouTube page). Participants were asked to answer questions about perceived diversity, choice simplicity, recommendation attractiveness, music discovery, and list familiarity [3] before moving on to the next list. These questions needed to be answered for each of the three lists.

After the participant assessed all three recommendation lists, we performed a manipulation check by placing the three lists next to each other (randomly ordered) and asked the participant to rank order the lists by diversity. The study ended with a short questionnaire about their music expertise (Goldsmiths Musical Sophistication Index [7]) and their general preference strength (adapted from [12]).

There were 103 participants who returned for the second part of the study. We included several control questions to filter out careless contributions, which left us with 100 participants for the analyses. Age: 18-65 (median 28), gender: 54 male, 46 female, and were compensated with $2.

4 Results

4.1 Manipulation Check

A Wilcoxon signed-rank test was used to test the perceived diversity levels of the recommendation lists. Results show an increase of perceived diversity by comparing the low diversity (M=1.28) against the medium (M=2.05, r=.60, Z=10.370, p<.001) and high condition (M=2.65, r=.80, Z=13.784, p<.001). A significant diversity increase was also found between medium and high (r=.45, Z=7.711, p<.001).

4.2 Measures

Items in the questionnaire were assessed using a confirmatory factor analysis (CFA) to determine whether they convey the predicted constructs. After deleting questions with high cross-loadings and low commonalities, the model consisting of six constructs showed a good fit: χ^2(390)=695.4, p<.001,

[3] Questions measuring perceived diversity, choice simplicity and recommendation attractiveness were adapted from [12]. We created the questions measuring music discovery and list familiarity in order to understand the relationship between perceived diversity and recommendation attractiveness.

CFI=.96, TLI=.95, $RMSEA$=.05. [4] The constructs with their items are shown below (5-point Likert scale; Disagree strongly-Agree strongly. Cronbach's alpha (α) and the average variance extracted (AVE) of each construct showed good values (i.e., α>.8, AVE>.5), indicating convergent validity. The square root of the AVE for each construct is higher than any of the factor loading of the respective construct; indicating good discriminant validity. For the standardized personality and music expertise questionnaires see [5] and [7]):

Choice Simplicity (AVE=.820, α=.932):

- I would find it easy to choose an artist to listen to because it stands out from the rest.
- I would find it difficult to choose an artist to listen to because all recommendations were equally bad.
- Many artists had comparable good aspects.

Recommendation Attractiveness (AVE=.881, α=.980):

- I am satisfied with the list of recommended artists.
- In most ways the recommended artists were close to ideal.
- The list of artist recommendations meet my exact needs.
- I would give the recommended artists a high rating.
- The list of artists showed too many bad items.
- The list of artists was attractive.
- The list of recommendations matched my preferences.

Discovery (AVE=.914, α=.955):

- The recommendations broadened my taste.
- The recommendations deepened my taste.

List Familiarity (AVE=.871, α=.953):

- I am familiar with the recommended artists.
- I did not know the artists from the list.
- I already listen to the artists that were recommended.

Perceived Diversity (AVE=.816, α=.957):

- The list of artists was varied.

[4] Cutoff values for the fit indices are proposed to be: CFI>.96, TLI>.95, and $RSMEA$<.05 [3].

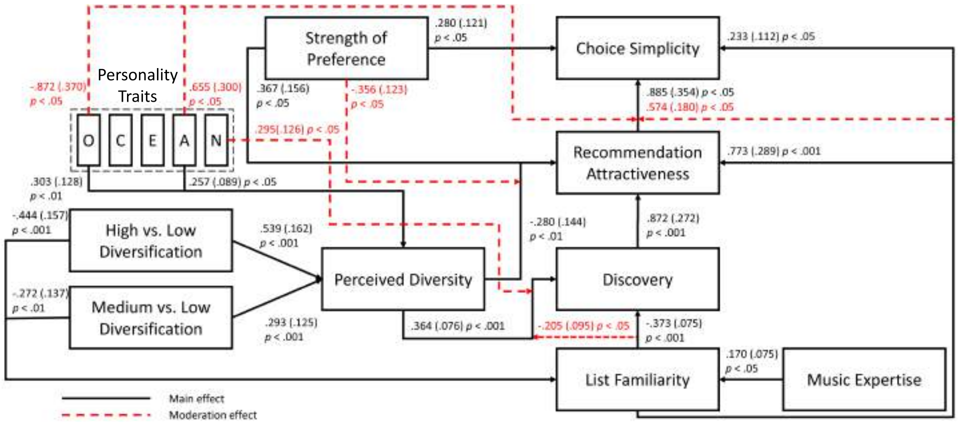

Figure 1: Path model. The numbers near the arrows denote the estimated means and the standard deviations. Personality: (O)penness to experience, (C)onscientiousnes, (E)xtraverion, (A)greeableness, (N)euroticsm.

- All the artists were similar to each other.
- Most artists were from the same genre.
- Many of the artists in the lists differed from other artists in the list.
- The artists differed a lot from each other on different aspects.

Strength of Preference (single item):

- I know what kind of music I like.

4.3 SEM Model

We created a path model with the subjective constructs of the CFA together with the personality and music expertise constructs using structural equation modeling (SEM) following the framework of Knijnenburg et al. [6] as a guideline (Figure 1). No order effects were observed in the order of presentation of the lists. The fit statistics show that the model has a good fit: χ^2(431)=701.4, p<.001, CFI=.96, TLI=.96, $RMSEA$=.04. Only effects of p<.05 are included in the model. The diversity conditions medium and high are compared against the low diversity condition.

The medium and high diversification both show an increase in perceived diversity. Additionally, some of the personality traits show a moderation effect on how the diversity of the recommendation list is perceived: those who scored higher on openness to experience and agreeableness perceive a higher degree of diversity of the lists.

A higher perceived diversity has a positive influence on the discovery of music (i.e., enriching ones taste). This effect is higher for those scoring high on the neuroticism scale, but lower for those who are familiar with the list. Discovery is furthermore influenced by list familiarity; more familiar with the list has a negative effect on the discovery of new music. The familiarity with the recommendation list is dependent on the level of diversity as well as on the music expertise. A higher list diversification results in lower list familiarity, while a higher music expertise results in higher familiarity.

The attractiveness of the recommendations is influenced by several constructs: perceived diversity, discovery, strength of preference, and list familiarity. Perceived diversity has a negative effect on the recommendation attractiveness, this effect gets stronger for those having a predefined preference. However, the effect becomes positive when it goes through the discovery of music; a higher attractiveness is achieved when the diversified list is able to enrich the music taste of people than when only recommending diverse items. Furthermore, strength of preference and list familiarity have both a positive effect on the perceived attractiveness.

The simplicity of making a choice is influenced by the recommendation attractiveness as well as by the strength of preference and list familiarity. The more attractive the recommendations are, the easier it becomes to make a choice. This effect is moderated by two personality traits and by list familiarity. The strength of recommendation attractiveness on choice simplicity is lower for those scoring high on openness to experience, but higher for those scoring high on agreeableness and those who are familiar with the list. Also the preference strength and list familiarity plays a positive role on the ease of making a choice.

5 Conclusion & Discussion

We showed to be able to effectively increase the diversity in the list of recommendations as it was reflected in the participants' increased perceived diversity. By increasing the list diversity and thereby increasing its perceived diversity, we show how the perceived recommendation list attractiveness is influenced. Our results indicate that recommendation list attractiveness is influenced by two effects: 1) by diversity mediated through discovery; diverse recommendation lists are perceived more attractive if they enrich the user's taste, or 2) by the list familiarity; a higher familiarity contributes to a higher attractiveness. Although these two ways were identified to increase list attractiveness, list diversification yields a stronger effect to increase the attractiveness and to subsequently increase choice simplicity.

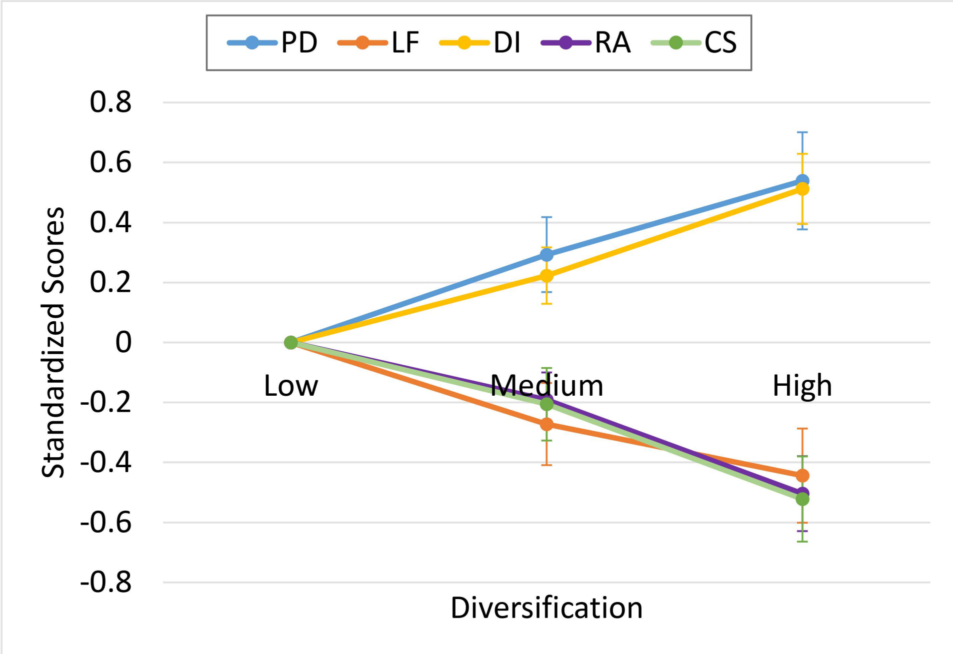

Figure 2: Marginal means with error bars of one std. error of the mean: perceived diversity (PD), list familiarity (LF), discovery (DI), recommendation attractiveness (RA), choice simplicity (CS).

Despite the new insights we provided on how diversification affects recommendation list attractiveness, our net effect of diversification on choice simplicity and recommendation attractiveness is negative (Figure 2). This in contrast to prior work [11, 12] showing opposite effects. A possible explanation for this opposing effect may be a domain dependency (movies in cited work versus music in this study). Specifically the range in which we diversified (i.e., top 200) may be too broad for the music domain. Although we took the same range as prior work, music consist of more variable components (e.g., variations in genre, artist, context). By diversifying within the top 200, we may have created recommendation lists with items that are too far outside the spectrum of the user's taste; resulting in increasingly less attractive lists. We will further investigate the reason behind our negative net effect of diversification in future work.

In addition, we identified that personality traits play a significant moderating role in the relationship between recommendation attractiveness and choice simplicity: open people seem to have more difficulties to make a choice out of attractive items than those scoring high on agreeableness. This may be because open people have a preference for variety [5] and therefore a diverse list, but with equally attractive items, makes it more difficult to make a choice. Agreeable people, on the other hand, tend to be tactful [5]; a diverse list would allow them to make more tactical choices.

Although prior work has shown that recommendation list diversity leads to a decrease in choice difficulties by increasing the list attractiveness [12], we provide new insights by showing that this effect occurs by enriching one's taste through the ability to discover new items. Furthermore, personal characteristics (e.g., expertise and strength preference) have been identified to be important moderators (e.g., [9, 12]). With personality we found an additional personal characteristic that should be taken into account.

We were unable to effectively measure choice satisfaction. Due to the within-subjects design of our experiment we did not ask participants to make a choice, but rather assess the recommendation lists. However, prior work (e.g., [12, 13]) has shown that choice simplicity leads to a higher choice satisfaction in recommender systems. We will address this question in future work.

6 Acknowledgments

Supported by the Austrian Science Fund (FWF): P25655.

References

[1] Castells, P., Hurley, N.J., Vargas, S.: Novelty and diversity in recommender systems. In: Recommender Systems Handbook, pp. 881–918. Springer (2015)

[2] Ferwerda, B., Schedl, M., Tkalcic, M.: Personality & emotional states: Understanding users' music listening needs. UMAP 2015 Extended Proceedings

[3] Hu, L.t., Bentler, P.M.: Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural equation modeling: a multidisciplinary journal 6(1), 1–55 (1999)

[4] Hu, Y., Koren, Y., Volinsky, C.: Collaborative filtering for implicit feedback datasets. In: ICDM (2008)

[5] John, O.P., Donahue, E.M., Kentle, R.L.: The big five inventory: Versions 4a and 54, institute of personality and social research. UC Berkeley (1991)

[6] Knijnenburg, B.P., Willemsen, M.C., Gantner, Z., Soncu, H., Newell, C.: Explaining the user experience of recommender systems. UMUAI (2012)

[7] Müllensiefen, D., Gingras, B., Stewart, L., Musil, J.: Goldsmiths musical sophistication index (2012)

[8] Schedl, M.: The LFM-1b Dataset for Music Retrieval and Recommendation. In: Proceedings of the ACM International Conference on Multimedia Retrieval (ICMR). New York, USA (June 2016)

[9] Scheibehenne, B., Greifeneder, R., Todd, P.M.: What moderates the too-much-choice effect? Psychology & Marketing 26(3), 229–253 (2009)

[10] Tkalcic, M., Ferwerda, B., Hauger, D., Schedl, M.: Personality correlates for digital concert program notes. UMAP 2015, Springer LNCS 9146 (2015)

[11] Willemsen, M.C., Graus, M.P., Knijnenburg, B.P.: Understanding the role of latent feature diversification on choice difficulty and satisfaction. UMUAI (under revision)

[12] Willemsen, M.C., Knijnenburg, B.P., Graus, M.P., Velter-Bremmers, L.C., Fu, K.: Using latent features diversification to reduce choice difficulty in recommendation lists. RecSys 11, 14–20 (2011)

[13] Ziegler, C.N., McNee, S.M., Konstan, J.a., Lausen, G.: Improving recommendation lists through topic diversification. WWW p. 22 (2005)

The Influence of Users' Personality Traits on Satisfaction and Attractiveness of Diversified Recommendation Lists

Published in: Extended Proceedings of the 10th ACM Conference on Recommender Systems (RecSys 2016): 4th Workshop on Emotions and Personality in Personalized Systems (EMPIRE 2016; Boston, MA, US)

BRUCE FERWERDA, Johannes Kepler University, Linz (AT)
MARK GRAUS, Eindhoven University of Technology, Eindhoven (NL)
ANDREU VALL, Johannes Kepler University, Linz (AT)
MARKO TKALCIC, Free University of Bolzano, Bolzano (IT)
MARKUS SCHEDL, Johannes Kepler University, Linz (AT)

Abstract

Diversifying recommendations has shown to be a good means to counteract on choice difficulties and overload, and is able to positively influence subjective evaluations, such as satisfaction and attractiveness. Personal characteristics (e.g., domain expertise, prior preference strength) have shown to influence the desired level of diversity in a recommendation list. However, only personal characteristics that are directly related to the domain have been investigated so far. In this work we take personality traits as a general user model and show that specific traits are related to a preference for different levels of diversity (in terms of recommendation satisfaction and attractiveness). Among 103 participants we show that conscientiousness is related to a preference for a higher degree of diversification, while agreeableness is related to a mid-level diversification of the recommendations. Our results have implications on how to personalize recommendation lists (i.e., the amount of diversity that should be provided) depending on users' personality.

Keywords: Diversity, Recommender Systems, User-Centric Evaluation, Personality

1 Introduction

Providing users with a diversified list of recommendations has shown to have positive effects on the user experience. With an abundance of choices available nowadays, providing diversity in the recommendations can counteract on the negative psychological effects that users may experience, such as choice overload and choice difficulties [26]. These negative effects are caused by recommender systems, which are originally designed to output recommendations that are closest to the user's interest. The closer to the user's interest, the higher the accuracy of the recommender system algorithm, but also results in recommendations that are often too similar to each other (e.g., same level of attractiveness to the user). This does not only increase the chance of choice overload and choice difficulties to the user, but also increases the possibility of not covering the full spectrum of the user's interest [3].

Although prior research has shown that recommendation diversity has positive effects on the user experience, differences between diversity needs of users have not been given a lot of attention. Domain expertise and prior choice preferences have shown to play a role in the amount of diversity desired by the user [2, 6, 26]. Others have shown that diversity needs can also be related to cultural dimensions [7, 14]. In this work we consider personality traits as an indicator of satisfaction and attractiveness on differently diversified music recommendation lists.

The use of personality as a general model for users has gained increased interest. Several works revealed personality-based relationships with users' behavior, preferences, and needs (e.g., [10, 15, 25]), how to implicitly acquire personality traits of users from social media trails (e.g., Facebook [1, 4, 12, 20], Twitter [16, 21], and Instagram [11, 13, 24]), and how personality traits can be implemented into a personalized system [8, 9]. With our work we contribute to the personality research by providing more insights into personality-related diversity needs. We found among 103 participants that the conscientiousness and agreeableness personality traits play a role in the desired amount of diversity in a recommendation list. While conscientious participants showed a higher degree of satisfaction and attractiveness with the more diversified recommendations, agreeable participants were more satisfied and found the list more attractive with medium amount of diversity in the recommendations.

2 Related work

The positive effects of recommendation list diversity has been shown by several researchers. Bollen et al. [2] and Willemsen et al. [26] investigated the influence of diversity on movie recommendations and found that diversity has a positive effect on the attractiveness of the recommendation set, the difficulty to make a choice, and eventually on the choice satisfaction. Besides the positive effects of diversification, also personal characteristics play a role on the attractiveness of the diversified recommendation list (e.g., strength of prior preference or domain expertise [2, 23]). Bollen et al. [2] found that expertise in the domain showed a positive effect on the item attractiveness.

The personal characteristics that have been identified so far are domain specific to the kind of recommendations. However, a more general personal characteristic may be present that influences the subjective evaluations with the diversified recommendations. Personality has shown to be an enduring factor, which can relate to one's taste, preference, and interest (e.g., [5, 10, 25]). Chen et al. [5] and Wu et al. [27] showed relationships with personality and preference for diversification based on different movie characteristics (e.g., genre, artist, director). Ferwerda et al. [10] showed that music preferences can be related to the personality of the listener, whereas Tkalcic et al. [25] found relationships between personality traits and the preference of being exposed to certain amounts of multimedia meta-information.

In this work we investigate whether personality traits can be considered a personal characteristic that influences the subjective evaluations of diversified recommendation lists. To this end, we rely on the widely used five-factor model (FFM), which categorizes personality into five general dimensions: openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism [19].

3 Data Preparation & Procedures

We created differently diversified music recommendation lists in order to investigate the influence of personality traits on the subjective evaluation of the recommendation lists. Since we created the recommendation lists offline, we separated the study in two parts. In the first part participants were recruited and their *complete* Last.fm listening history was crawled in order to create the recommendation lists. After the lists were created, participants from the first part were invited for the second part where they were asked to assess the diversified recommendation lists.

We recruited 254 participants through Amazon Mechanical Turk for the first part of the study. Participation was restricted to those located in the United States with a very good reputation ($\geq$95% HIT approval rate and $\geq$1000 HITs approved) and a Last.fm account with at least 25 listening events. Furthermore, they were asked to fill in the 44-item Big Five Inventory personality questionnaire [19] to measure the FFM. Control questions were asked to filter out fake and careless contributions. A compensation of $1 was provided. We crawled the complete listening history of each participant and aggregated the listening events to represent artist and playcount (i.e., number of times listened to an artist).

In order to prepare the music recommendation lists for each participant, we complemented our data with the LFM-1b dataset [22]. [1] This dataset consists of the complete listening histories of 120,322 Last.fm users from different countries. Since our participants were all located in the United States, we only used the United State users of the LFM-1b dataset to complement our dataset. This resulted in 10,255 additional users, which we also aggregated into artist and playcount for each user. The final dataset consists of user, artist, and artist playcount triplets with a total of 387,037 unique artists for the creation of the recommendation lists.

We used the weighted matrix factorization algorithm of [18] on our final dataset to calculate the recommended items. This algorithm is specifically designed to deal with datasets consisting of implicit feedback (e.g., artist playcounts). We optimized the factorization hyper-parameters by conducting grid-search and picking the setting that yielded the best 5-fold cross-validated mean percentile rank. Specifically, using 20 factors, confidence scaling factor α=40, regularization weight λ=1000 and 10 iterations of alternating least squares, we achieved the best 5-fold cross-validated mean percentile rank of 1.78%. [2] Afterwards we factorized the whole user-artist triplets using this set of hyper-parameters.

The recommended items were diversified as was done in [26] by using the method of [28]. By using the latent features as the basis of diversification instead of additional metadata like genre information (as is done in content-based recommender systems) guarantees that diversity is manipulated in line with user preferences. Previous research demonstrated that this way of diversifying recommendations is perceived accordingly by users [26].

A greedy selection to optimize the intra-list similarity [3] was run on the top 200 recommended artists (i.e., the 200 artists with highest predicted

[1] Available at `http://www.cp.jku.at/datasets/LFM-1b/`

[2] See [18] for details on the hyper-parameters and the definition of the mean percentile rank metric.

relevance) to maximize the distances between item vectors in the matrix factorization space. This algorithm starts with a recommendation set consisting of the artist with highest predicted relevance. In an iterative fashion items are added to the recommendation set until it contains 10 items.

In each step of the iteration, for each candidate item i the sum of all distances from its item vector to each item vector in the recommendation set is calculated: $c_i = \sum_{j=1}^{z} d(i,j)$, where z is the number of items in the recommendation set and $d(i,j)$ is the Euclidean distance between two item vectors i and j). All candidate items are ranked based on decreasing value of c_i (P_{c_i}) and on predicted relevance (P_{r_i}). A weighting factor β is introduced to balance the trade-off between predicted relevance and diversity. For each candidate item the combined rank is calculated following $w_i^* = \beta * P_{c_i} + (1-\beta) * P_{r_i}$. The item with the highest combined rank is added to the recommendation set and the next step is taken until 10 items are selected.

β was manipulated to achieve different levels of diversification. In the described implementation β=1 corresponds to maximum diversity, β=0 corresponds to maximum predicted relevance. We compared recommendation lists for different values of β in terms of the sum of distances between the latent features scores of items in the recommendation set and their average range. The list for β=0.4 showed to fall halfway between maximum relevance and maximum diversity. Thus, the final β levels for diversification were set at β=0 (low), β=0.4 (medium), and β=1 (high).

After the recommendation lists were created, emails were sent out to all participants to invite them for the second part of the study. We created a login screen so that we could retrieve the personalized recommendation lists for each participant. After the log in, the participant was sequentially presented with a recommendation list for three times, with each time a different level of diversity (i.e., low, medium, or high). The order of presentation was randomized. Each recommended artist was enriched with metadata from Last.fm (i.e., picture, genre, Top-10 songs with the number of listeners and playcounts), which was shown when hovered over the name in the list. Additionally, example songs were provided by clicking on the artist name (new browser screen linked to the artist's YouTube page). Participants were asked to answer questions about perceived diversity, recommendation satisfaction, and recommendation attractiveness [3] before moving on to the next list. These questions needed to be answered for each of the three lists.

[3]Questions measuring perceived diversity and recommendation attractiveness were adapted from [26].

After the participant assessed all three recommendation lists, we performed a manipulation check by placing the three lists next to each other (randomly ordered) and asked the participant to rank order the lists by diversity.

There were 103 participants who returned for the second part of the study. We included several control questions to filter out careless contributions, which left us with 100 participants for the analyses. Age: 18-65 (median 28), gender: 54 male, 46 female, and were compensated with $2.

4 Results

4.1 Manipulation Check

A Wilcoxon signed-rank test was used to test the perceived diversity levels of the recommendation lists. Results show an increase of perceived diversity by comparing the low diversity (M=1.28) against the medium (M=2.05, r=.60, Z=10.370, p<.001) and high condition (M=2.65, r=.80, Z=13.784, p<.001). A significant diversity increase was also found between medium and high (r=.45, Z=7.711, p<.001).

4.2 Measures

Items in the questionnaire were assessed using a confirmatory factor analysis (CFA) with repeated ordinal dependent variables and a weighted least squares estimator to determine whether the questions convey the predicted constructs. After deleting questions with high cross-loadings and low commonalities, the model consisting of three constructs showed a good fit: χ^2(32)=108.6, p<.001, CFI=.99, TLI=.98, $RMSEA$=.06. [4] The constructs with their items are shown below (5-point Likert scale; Disagree strongly-Agree strongly). The Cronbach's alpha (α) and the average variance extracted (AVE) of each construct showed good values (i.e., α>.8, AVE>.5), indicating convergent validity. Also, the square root of the AVE for each construct is higher than any of the factor loadings (FL) of the respective construct, which indicates good discriminant validity.

Perceived Diversity (AVE=.723, α=.887):

- The list of artists was varied. (FL=.858)
- Many of the artists in the lists differed from other artists in the list. (FL=.837)

[4] Cutoff values for a good model fit are proposed to be: CFI>.96, TLI>.95, and $RSMEA$<.05 [17].

- The artists differed a lot from each other on different aspects. (FL=.855)

Recommendation Satisfaction (AVE=.821, α=.932):

- I am satisfied with the list of recommended artists. (FL=.927)
- In most ways the recommended artists were close to ideal. (FL=.905)
- The list of artist recommendations meet my exact needs. (FL=.885)

Recommendation Attractiveness (AVE=.771, α=.931):

- I would give the recommended artists a high rating. (FL=.874)
- The list of artists showed too many bad items. (FL=-.830)
- The list of artists was attractive. (FL=.914)
- The list of recommendations matched my preferences. (FL=.893)

4.3 Analysis

We used a repeated measures ANOVA in order to investigate the influence of personality traits on the subjective evaluations of the diversified music recommendation lists. Below the results of personality traits on the different subjective evaluations are provided. The effects between diversity levels are all compared against the low diversity condition.

4.3.1 Personality on Perceived Diversity

Results show that Mauchly's test is not violated (χ^2(2)= .115, p=.944), so sphericity can be assumed. The results show that there are no significant main effects of the different personality traits on perceived diversity. However, a general difference in perceived diversity can be assumed (F(2, 22)=51.029, p<.001). Exploring the differences between the levels of diversified recommendation lists show that there is an increase in perceived diversity when comparing the low diversified list against the medium (F(1, 11)=11.596, p<.001) and the high diversified lists (F(1, 11)=31.191, p< .001). This confirms once more that our diversification was effective and was perceived as such by the participants.

4.3.2 Personality on Recommendation Satisfaction

Mauchly's test shows that sphericity is not violated (χ^2(2)= 1.830, p=.401), and therefore no correction is needed. Assessing the effect of the different personality traits on the recommendation satisfaction, the following personality traits show a main effect: conscientiousness (F(4, 22)=2.454, p<.05) and agreeableness (F(4, 22)=3.886, p<.05). Additional analyses by looking at the levels between the diversity levels (i.e., low, medium, and high diversification) show that conscientious participants are increasingly satisfied when provided a higher degree of diversity: medium diversity (F(2, 11)=3.994, p<.05) and high diversity (F(2, 11)=4.036, p<.05). However, the satisfaction differences for agreeable participants show a higher satisfaction for the medium diversification (F(2, 11)=9.660, p<.05) than for the high diversification (F(2, 11)=4.036, p<.05).

4.3.3 Personality on Recommendation Attractiveness

Assessing Mauchly's test shows that there is no violating of sphericity (χ^2(2)= 1.860 p=.395). Also here, results show main effects for the conscientiousness (F(4, 22)=3.157, p<.05) and agreeableness (F(4, 22)=3.469, p<.05) personality traits. By looking at the differences between the levels of diversification, we found similar patterns as with satisfaction. Results show that conscientious participants were increasingly more attracted to more diversified recommendation lists: medium (F(2, 11)=2.955, p<.05), high (F(2, 11)=7.866, p<.05). Participants scoring high on the agreeableness personality traits show to be more attracted to the medium (F(2, 11)=5.933, p<.05) diversified list than to the high (F(2, 11)=5.314, p<.05) diversified list.

5 Conclusion & Discussion

Our results show that certain personality traits (i.e., conscientiousness and agreeableness) are related to the subjective evaluations of diversified recommendation lists. We found that conscientious people judged a higher degree of diversity more attractive and were more satisfied with it, whereas agreeable people showed to have more interest (i.e., list attractiveness and satisfaction) in a medium degree of diversity.

The relationships that we found can be used in personality-based systems as proposed in [8]. With the increased connectedness of applications, such as recommender systems, with social networking sites, users' personality can be acquired without the need of behavioral data in the application (e.g., via Facebook [1, 4, 12, 20], Twitter [16, 21], or Instagram [11, 13, 24]). By identifying relationships with users' personality traits, such as in this work, cross-domain inferences about users' preferences and needs can be made and implemented to provide a personalized experience to users.

6 Acknowledgments

Supported by the Austrian Science Fund (FWF): P25655.

References

[1] Back, M.D., Stopfer, J.M., Vazire, S., Gaddis, S., Schmukle, S.C., Egloff, B., Gosling, S.D.: Facebook profiles reflect actual personality, not self-idealization. Psychological Science 21, 372–374 (2010)

[2] Bollen, D., Knijnenburg, B.P., Willemsen, M.C., Graus, M.: Understanding choice overload in recommender systems. In: Proceedings of the fourth ACM conference on RecSys. pp. 63–70. ACM (2010)

[3] Castells, P., Hurley, N.J., Vargas, S.: Novelty and diversity in recommender systems. In: Recommender Systems Handbook, pp. 881–918. Springer (2015)

[4] Celli, F., Bruni, E., Lepri, B.: Automatic personality and interaction style recognition from facebook profile pictures. In: Proceedings of the ACM MM (2014)

[5] Chen, L., Wu, W., He, L.: How personality influences users' needs for recommendation diversity? In: Proceeding of CHI'13 EA. ACM (2013)

[6] Ekstrand, M.D., Harper, F.M., Willemsen, M.C., Konstan, J.A.: User perception of differences in recommender algorithms. In: Proceedings of the 8th ACM Conference on Recommender systems. pp. 161–168. ACM (2014)

[7] Ferwerda, B., Schedl, M.: Investigating the relationship between diversity in music consumption behavior and cultural dimensions: A cross-country analysis. In: Proc. of the 1st Workshop on SOAP (2016)

[8] Ferwerda, B., Schedl, M.: Enhancing Music Recommender Systems with Personality Information and Emotional States: A Proposal. In: Proceedings of the 2nd Workshop on EMPIRE (2014)

[9] Ferwerda, B., Schedl, M.: Personality-Based User Modeling for Music Recommender Systems. In: Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2016). Riva del Garda, Italy (2016)

[10] Ferwerda, B., Schedl, M., Tkalcic, M.: Personality & emotional states: Understanding users' music listening needs. UMAP 2015 Extended Proceedings

[11] Ferwerda, B., Schedl, M., Tkalčič, M.: Predicting Personality Traits with Instagram Pictures. In: Proceedings of the 3rd Workshop on EMPIRE (2015)

[12] Ferwerda, B., Schedl, M., Tkalčič, M.: Personality Traits and the Relationship with (Non-)Disclosure Behavior on Facebook. In: Companion of the 25th International WWW Conference (2016)

[13] Ferwerda, B., Schedl, M., Tkalčič, M.: Using Instagram Picture Features to Predict Users' Personality. In: Proceedings of the 22nd International Conference on MMM. Miami, USA (January 2016)

[14] Ferwerda, B., Vall, A., Tkalčič, M., Schedl, M.: Exploring Music Diversity Needs Across Countries. In: Proceedings of the 24th International Conference on UMAP. Halifax, Canada (July 2016)

[15] Ferwerda, B., Yang, E., Schedl, M., Tkalčič, M.: Personality Traits Predict Music Taxonomy Preferences. In: ACM CHI '15 EA (2015)

[16] Golbeck, J., Robles, C., Edmondson, M., Turner, K.: Predicting Personality from Twitter. In: Proceedings of the 3rd International Conference on SocialCom (2011)

[17] Hu, L.t., Bentler, P.M.: Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural equation modeling: a multidisciplinary journal 6(1), 1–55 (1999)

[18] Hu, Y., Koren, Y., Volinsky, C.: Collaborative filtering for implicit feedback datasets. In: ICDM (2008)

[19] John, O.P., Donahue, E.M., Kentle, R.L.: The big five inventory: Versions 4a and 54, institute of personality and social research. UC Berkeley (1991)

[20] Park, G., Schwartz, H.A., Eichstaedt, J.C., Kern, M.L., Kosinski, M., Stillwell, D.J., Ungar, L.H., Seligman, M.E.: Automatic Personality Assessment Through Social Media Language. Journal of Personality and Social Psychology 108 (November 2014)

[21] Quercia, D., Kosinski, M., Stillwell, D., Crowcroft, J.: Our twitter profiles, our selves: Predicting personality with twitter. In: Proceedings of the 3rd International Conference on SocialCom (2011)

[22] Schedl, M.: The LFM-1b Dataset for Music Retrieval and Recommendation. In: Proceedings on ICMR (2016)

[23] Scheibehenne, B., Greifeneder, R., Todd, P.M.: What moderates the too-much-choice effect? Psychology & Marketing 26(3), 229–253 (2009)

[24] Skowron, M., Ferwerda, B., Tkalčič, M., Schedl, M.: Fusing Social Media Cues: Personality Prediction from Twitter and Instagram. In: Companion Proceedings of the 25th International WWW Conference (2016)

[25] Tkalcic, M., Ferwerda, B., Hauger, D., Schedl, M.: Personality correlates for digital concert program notes. UMAP 2015, Springer LNCS 9146 (2015)

[26] Willemsen, M.C., Knijnenburg, B.P., Graus, M.P., Velter-Bremmers, L.C., Fu, K.: Using latent features diversification to reduce choice difficulty in recommendation lists. RecSys 11, 14–20 (2011)

[27] Wu, W., Chen, L., He, L.: Using personality to adjust diversity in recommender systems. In: Proceedings of the 24th ACM Conference on Hypertext and Social Media. pp. 225–229. ACM (2013)

[28] Ziegler, C.N., McNee, S.M., Konstan, J.a., Lausen, G.: Improving recommendation lists through topic diversification. WWW p. 22 (2005)

8 General Conclusion

This dissertation provides a novel comprehensive way for creating a personalized music recommender system with personality and cultural information. To come back to the two RQs that this dissertation addresses:

1. How do personality traits and cultural background relate to behavior, preferences, and needs?
2. How can personality traits be implicitly acquired from SNSs?

In order to answer RQ1, several studies were conducted (see §5 [13, 19, 23, 24, 25, 52] and §7 [16, 15]) covering a wide range of different facets for a music recommender system. In §5.1 [24] and §5.2 [25] studies were conducted that are useful for creating user interfaces. The results of these studies show that there is a relationship between music browsing strategies and personality traits. With the abundance of music, these results could be applied to adjust the interface to present music that is already in line with the user's preferred way of music browsing to present an interface that is littered with different possibilities.

The study in §5.3 [19] looked at personality related music preferences and how a preference changes depending on the mood so that a better music recommendation can be given to the user. In §5.4 [52] a study was conducted that showed relationships between personality and supportive multimedia material about music. The studies conducted in §5.5 [23], §5.6 [13], §7.1 [16], and §7.2 [15] looked at diversity in music listening behavior and needs to better understand the amount of diversity that should be provided in a playlist. The studies in §5.5 [23] and §5.6 [13] looked at diversity from a country perspective, and analyzed listening behavior. In §7.1 [16] and §7.2 [15] diversity was investigated on a personality level, recommendations were provided, and a user-centric evaluation was performed to assess different levels of diversity.

For RQ2, several studies were conducted involving SNSs, which are shown in the works under §6 [18, 20, 21, 22, 49]. The work in §6.1 [18] revealed the psychological mechanism underlying sharing and posting behavior on Facebook so that mechanisms can be created to promote sharing and posting of content. Promoting sharing and positing behavior is necessary as implicit personality acquisition is usually dependent on the content available. The studies conducted in §6.2 [20], §6.3 [22], §6.4 [21], and §6.5 [49] showed how to implicitly acquire personality traits from SNSs. The works of §6.2 [20] and §6.3 [22] showed the implicit personality acquisition from Instagram picture features (i.e., the characteristics of the modified pictures). In §6.4 [21], we showed how to implicitly acquire personality from Facebook when user data is limited, and §6.5 [49] showed the improvement of personality acquisition by fusion SNSs data from Instagram and Twitter.

Although various new insights were gained in this dissertation, it was not possible to create and test a full-fledged music recommender system based on all the works. During the course of the studies it became clear that it would be challenging to gather enough eligible participants for the works independently, let alone gather enough participants to test a whole system with all the results implemented. Therefore, unfortunately, the user-centric evaluation was only done on music recommendation diversity (see the works in §7 [15, 16]). Nevertheless, by studying diverse areas of a music recommender system, this dissertation provides a holistic and novel view of how personality and cultural information can be applied to provide a personalized experience to the user (see §4 [13, 14, 17]). The studies do not only have implications as a whole, but have separate scientific contributions as well. The results of the works in §5 [13, 19, 23, 24, 25, 52] and §7 [15, 16] contribute to the personality and cultural research by showing new relationships with behavior, preferences, and needs. Furthermore, the works in §6 [18, 20, 21, 22, 49] add to the research task of personality acquisition by providing new ways to implicitly acquire personality traits from SNSs.

References

[1] Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. Knowledge and Data Engineering, IEEE Transactions on 17(6), 734–749 (2005)

[2] Back, M.D., Stopfer, J.M., Vazire, S., Gaddis, S., Schmukle, S.C., Egloff, B., Gosling, S.D.: Facebook profiles reflect actual personality, not self-idealization. Psychological science (2010)

[3] Balabanović, M., Shoham, Y.: Fab: content-based, collaborative recommendation. Communications of the ACM 40(3), 66–72 (1997)

[4] Bollen, D., Knijnenburg, B.P., Willemsen, M.C., Graus, M.: Understanding choice overload in recommender systems. In: Proceedings of the fourth ACM conference on RecSys. pp. 63–70. ACM (2010)

[5] Burke, R.: Hybrid recommender systems: Survey and experiments. User Modeling and User-Adapted Interaction 12(4), 331–370 (2002)

[6] Cantador, I., Fernández-tobías, I., Bellogín, A.: Relating Personality Types with User Preferences in Multiple Entertainment Domains. In: Extended Proceedings of the ACM Conference on Recommender Systems (RecSys): Workshop on Emotions and Personality in Personalized Systems (EMPIRE) (2013)

[7] Castells, P., Hurley, N.J., Vargas, S.: Novelty and diversity in recommender systems. Recommender Systems Handbook pp. 881–918 (2015)

[8] Celli, F., Bruni, E., Lepri, B.: Automatic personality and interaction style recognition from facebook profile pictures. In: Proceedings of the ACM International Conference on Multimedia. pp. 1101–1104. ACM (2014)

[9] Chen, L., Wu, W., He, L.: How personality influences users' needs for recommendation diversity? In: CHI'13 Extended Abstracts on Human Factors in Computing Systems. pp. 829–834. ACM (2013)

[10] Chia, P.H., Yamamoto, Y., Asokan, N.: Is this app safe?: a large scale study on application permissions and risk signals. In:

the International World Wide Wide (WWW) Conference. ACM (2012)

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

[12] Ferwerda, B.: The soundtrack of my life: Adjusting the emotion of music. In: Collaborating with Intelligent Machines (workshop), at the ACM Conference on Human Factors in Computing Systems (CHI) (2015)

[13] Ferwerda, B., Schedl, M.: Investigating the relationship between diversity in music consumption behavior and cultural dimensions: A cross-country analysis. In: Extended Proceedings of the 24th International Conference on User Modeling, Adaptation and Personalization (UMAP): 1st Workshop on Surprise, Opposition, and Obstruction in Adaptive and Personalized Systems (SOAP) (2016)

[14] Ferwerda, B., Schedl, M.: Personality-based user modeling for music recommender systems. In: Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (2016)

[15] Ferwerda, B., Graus, M., Vall, A., Tkalcic, M., Schedl, M.: The influence of users' personality traits on satisfaction and attractiveness of diversified recommendation lists. In: Extended Proceedings of the 10th ACM Conference on Recommender Systems (RecSys): 4th Workshop on Emotions and Personality in Personalized Systems (EMPIRE) (2016)

[16] Ferwerda, B., Graus, M., Vall, A., Tkalcic, M., Schedl, M.: How item discovery enabled by diversity leads to increased recommendation list attractiveness. In: In Submission (2017)

[17] Ferwerda, B., Schedl, M.: Enhancing music recommender systems with personality information and emotional states: A proposal. In: Proceedings of the 2nd Workshop on Emotions and Personality in Personalized Services held in conjuction with the 22nd Conference on User Modeling, Adaptation and Personalization. Aalborg, Denmark (2014)

[18] Ferwerda, B., Schedl, M., Tkalcic, M.: To post or not to post: The effects of persuasive cues and group targeting mechanisms

on posting behavior. In: Proceedings of the International Conference on Social Computing (SocialCom) (2014)

[19] Ferwerda, B., Schedl, M., Tkalcic, M.: Personality & emotional states: Understanding users' music listening needs. In: Proceedings of the 23th International Conference on User Modeling, Adaptation and Personalization (UMAP 2015). Dublin, Ireland (2015)

[20] Ferwerda, B., Schedl, M., Tkalcic, M.: Predicting personality traits with instagram pictures. In: Extended Proceedings of the ACM Conference on Recommender Systems (RecSys): Workshop on Emotions and Personality in Personalized Systems (EMPIRE) (2015)

[21] Ferwerda, B., Schedl, M., Tkalcic, M.: Personality traits and the relationship with (non-) disclosure behavior on facebook. In: Proceedings of the 25th International Conference Companion on World Wide Web. pp. 565–568 (2016)

[22] Ferwerda, B., Schedl, M., Tkalcic, M.: Using instagram picture features to predict users' personality. In: Proceedings of the International Conference on MultiMedia Modeling (MMM). Springer (2016)

[23] Ferwerda, B., Vall, A., Tkalcic, M., Schedl, M.: Exploring Music Diversity Needs Across Countries. In: Proceedings of the 24th International Conference on User Modeling, Adaptation and Personalization (UMAP 2016). Halifax, Canada (July 2016)

[24] Ferwerda, B., Yang, E., Schedl, M., Tkalcic, M.: Personality traits predict music taxonomy preferences. In: Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems. pp. 2241–2246. ACM (2015)

[25] Ferwerda, B., Yang, E., Schedl, M., Tkalcic, M.: Personality and taxonomy preferences, and the influence of category choice set size on the user experience for music streaming services. In: Under Review (2016)

[26] Golbeck, J., Robles, C., Edmondson, M., Turner, K.: Predicting personality from twitter. In: Proceedings of the International Conference on Social Computing (SocialCom). pp. 149–156. IEEE (2011)

[27] Gosling, S.D., Gaddis, S., Vazire, S., et al.: Personality impressions based on facebook profiles. Proceedings of the International Conference on Web and Social Media (ICWSM) (2007)

[28] Halko, S., Kientz, J.A.: Personality and persuasive technology: an exploratory study on health-promoting mobile applications. Persuasive technology pp. 150–161 (2010)

[29] Herlocker, J.L., Konstan, J.A., Borchers, A., Riedl, J.: An algorithmic framework for performing collaborative filtering. In: Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval. pp. 230–237. ACM (1999)

[30] Herlocker, J.L., Konstan, J.A., Terveen, L.G., Riedl, J.T.: Evaluating collaborative filtering recommender systems. ACM Transactions on Information Systems (TOIS) 22(1), 5–53 (2004)

[31] Hofstede, G., Hofstede, G.J., Minkov, M.: Cultures and organizations: Software of the mind, vol. 2. Citeseer (1991)

[32] House, R.J., Hanges, P.J., Javidan, M., Dorfman, P.W., Gupta, V.: Culture, leadership, and organizations: The GLOBE study of 62 societies. Sage publications (2004)

[33] Hu, R., Pu, P.: Acceptance issues of personality based recommender systems. In: Proceedings of the ACM Conference on Recommender Systems (RecSys). ACM (2009)

[34] Hudson, N.W., Fraley, R.C.: Volitional personality trait change: Can people choose to change their personality traits? Journal of personality and social psychology 109(3), 490 (2015)

[35] John, O.P., Donahue, E.M., Kentle, R.L.: The big five inventory: Versions 4a and 54, institute of personality and social research. UC Berkeley (1991)

[36] Jung, C.: Psychological Types. No. v. 6 in Bollingen Series No. 20, Princeton University Press (1976)

[37] Kitayama, S., Park, H.: Cultural shaping of self, emotion, and well-being: How does it work? Social and Personality Psychology Compass (2007)

[38] Knijnenburg, B.P., Willemsen, M.C., Gantner, Z., Soncu, H., Newell, C.: Explaining the user experience of recommender systems. User Modeling and User-Adapted Interaction 22(4-5), 441–504 (2012)

[39] Krauskopf, C.J., Saunders, D.R.: Personality and ability: The personality assessment system. Univ Pr of Amer (1994)

[40] Larsen, R.J., Buss, D.M.: Personality psychology. Jastrebarsko: Naklada Slap (2008)

[41] McCrae, R.R., John, O.P.: An introduction to the five-factor model and its applications. Journal of Personality (1992)

[42] McNee, S.M., Riedl, J., Konstan, J.A.: Being accurate is not enough: how accuracy metrics have hurt recommender systems. In: CHI'06 extended abstracts on Human factors in computing systems. pp. 1097–1101. ACM (2006)

[43] Park, G., Schwartz, H.A., Eichstaedt, J.C., Kern, M.L., Kosinski, M., Stillwell, D.J., Ungar, L.H., Seligman, M.E.: Automatic personality assessment through social media language. Journal of personality and social psychology 108(6), 934 (2015)

[44] Quercia, D., Kosinski, M., Stillwell, D., Crowcroft, J.: Our twitter profiles, our selves: Predicting personality with twitter. In: Proceedings of the International Conference on Social Computing (SocialCom). pp. 180–185. IEEE (2011)

[45] Rentfrow, P.J., Gosling, S.D.: The do re mi's of everyday life: the structure and personality correlates of music preferences. Journal of personality and social psychology 84(6), 1236 (2003)

[46] Ricci, F., Rokach, L., Shapira, B.: Introduction to recommender systems handbook. Springer (2015)

[47] Ross, C., Orr, E.S., Sisic, M., Arseneault, J.M., Simmering, M.G., Orr, R.R.: Personality and motivations associated with facebook use. Computers in Human Behavior 25(2), 578–586 (2009)

[48] Scheibehenne, B., Greifeneder, R., Todd, P.M.: What moderates the too-much-choice effect? Psychology & Marketing 26(3), 229–253 (2009)

[49] Skowron, M., Tkalčič, M., Ferwerda, B., Schedl, M.: Fusing social media cues: personality prediction from twitter and instagram. In: Proceedings of the 25th International Conference Companion on World Wide Web (WWW). pp. 107–108 (2016)

[50] Swearingen, K., Sinha, R.: Beyond algorithms: An hci perspective on recommender systems. In: Proceedings of the ACM

Special Interest Group on Information Retrieval (SIGIR): Workshop on Recommender Systems. vol. 13, pp. 1–11. Citeseer (2001)

[51] Tintarev, N., Masthoff, J.: A survey of explanations in recommender systems. In: Data Engineering Workshop, 2007 IEEE 23rd International Conference on. pp. 801–810. IEEE (2007)

[52] Tkalcic, M., Ferwerda, B., Hauger, D., Schedl, M.: Personality correlates for digital concert program notes. In: Proceedings of the Conference on User modeling, Adaptation, and Personalization (UMAP) (2015)

[53] Tkalcic, M., Kunaver, M., Košir, A., Tasic, J.: Addressing the new user problem with a personality based user similarity measure. In: Extended Proceedings of the Conference on User Modeling, Adaptation, and Personalization: Workshop on Decision Making and Recommendation Acceptance Issues in Recommender Systems (DEMRA) (2011)

[54] Trompenaars, F., Hampden-Turner, C.: Riding the Waves of Culture — Understanding Diversity in Global Business. McGraw-Hill Education, 3rd edn. (2012)

[55] Vinciarelli, A., Mohammadi, G.: A survey of personality computing. Affective Computing, IEEE Transactions on 5(3), 273–291 (2014)

[56] Willemsen, M.C., Knijnenburg, B.P., Graus, M.P., Velter-Bremmers, L.C., Fu, K.: Using latent features diversification to reduce choice difficulty in recommendation lists. In: Extended Proceedings of the ACM Conference on Recommender Systems (RecSys): Workshop on Human Decision Making in Recommender Systems. vol. 11, pp. 14–20 (2011)

[57] Winne, J.F., Gittinger, J.W.: An introduction to the personality assessment system. Journal of Community Psychology 1(2), 99–163 (1973)

[58] Ziegler, C.N., McNee, S.M., Konstan, J.a., Lausen, G.: Improving recommendation lists through topic diversification. Proceedings of the Internationa World Wide Web (WWW) Conference p. 22 (2005)