Smartphone Sensing for the Ecologically Valid Assessment of Individual Differences in Music-Listening Behavior

Larissa Sust

München, 2024

Smartphone Sensing for the Ecologically Valid Assessment of Individual Differences in Music-Listening Behavior

Inauguraldissertation

zur Erlangung des Doktorgrades der Philosophie an der

Ludwig-Maximilians-Universität München

vorgelegt von

Larissa Nina Nadine Sust

aus München

München, 2024

Erstgutachter: Prof. Dr. Markus Bühner

Zweitgutachter: Prof. Dr. Felix Schönbrodt

Tag der mündlichen Prüfung: 22.02.2024

Danksagung

An dieser Stelle möchte mich ich bei denjenigen Personen bedanken, die mich im Laufe der letzten Jahre begleitet und unterstützt haben. Zuallererst möchte ich mich bei meinem Doktorvater Markus Bühner für sein großes Vertrauen in meine Person und seine immerwährende Förderung bedanken, die er mir seit meinem Bachelorstudium zu Teil werden lässt. Auch möchte ich mich bei meinem Zweitgutachter Felix Schönbrodt bedanken, der mich als HiWi an den Lehrstuhl holte und so mein Interesse an der Forschungswelt weckte. Darüber hinaus gilt mein besonderer Dank Ramona Schödel, die mir in den vergangenen Jahren als Mentorin und Freundin tatkräftig zur Seite stand und als Sparringspartnerin in zahlreichen Zoom-Calls, als R-Lady oder zu guter Letzt als Korrekturleserin dieser Arbeit fungierte.

Weiterhin möchte ich mich beim gesamten PhoneStudy Team der LMU München bedanken, welches meine Forschung überhaupt erst möglich gemacht hat. Stellvertretend danke ich hier Florian Bemmann, der unsere Studiendurchführung technisch gemanagt hat und bei Serverproblemen stets eine sehr große Hilfe war. Großer Dank gilt an dieser Stelle auch Gayatri Kudchadker für die tolle Zusammenarbeit bei der schier endlos andauernden Datenvorverarbeitung der Song Lyrics aus meiner Arbeit.

Außerdem möchte ich mich bei meinen Lehrstuhl-Kollegen für die wertschätzende Arbeitsatmosphäre und die Kultur der gegenseitigen Unterstützung bedanken. Die gemeinsamen Konferenzbesuche, Sommerfeste und Weinabende machen die Zusammenarbeit unvergesslich. Mein herzlicher Dank gilt dabei Lena-Marei Wiessner, Florian Pargent, Philipp Sckopke, Felix Schönbrodt, Lena Schiestel, Caroline Zygar-Hoffmann, Elisabeth Kraus, Thomas Reiter, und Philipp Sterner. Ebenfalls danken möchte ich meinen externen Kollegen Clemens Stachl und Sanaz Talaifar für die gute Zusammenarbeit und Unterstützung.

Zuletzt geht ein großes Dankeschön an meine Familie und Freunde. Mein allergrößter Dank gilt dabei meiner Mama, die mich in zahllosen (v.a. durch das Verfassen von Abstracts ausgelösten) Krisen emotional aufgebaut hat. Außerdem danke ich meinem Agent R Regina, die mich durch Postkarten, Kinobesuche und Food-Eskalationen motiviert hat, meine Mission zu beenden. Schließlich danke ich meinen Freundinnen Ricarda, Vanessa, Sandra, Stephanie und Lauren dafür, dass sie mir trotz meines chronischen Zeitmangels einen Platz in ihrem Leben freigehalten haben. Zu guter Letzt danke ich der Brauerei Loscher, die mich mithilfe von Club Mate durch so manches Nachmittagstief gebracht hat.

Zusammenfassung

Die Einführung von Smartphones hat zweifellos unser alltägliches Leben revolutioniert, einschließlich der Art und Weise, wie wir Musik konsumieren. Durch ihre kleine Größe und die weite Verbreitung von internetbasierten Streaming-Apps wurden die tragbaren Supercomputer zu beliebten Musikgeräten, die es uns ermöglichen, jederzeit und überall beliebige Songs aus einem riesigen Musikangebot abzuspielen. Damit wurden Smartphones gleichzeitig zu einer digitalen Quelle für Musikhördaten, denn mithilfe speziell entwickelter Forschungsapps lassen sich diverse Nutzungsdaten einschließlich der Musikhör-Historie ihrer Nutzer aufzeichnen. Diese sogenannte „Smartphone Sensing"-Methodik bietet die Möglichkeit, das Musikhörverhalten eines Nutzers über längere Zeiträume und in einem ökologisch validen Umfeld, nämlich in dessen Alltag auf seinen privaten Geräten digital zu beobachten. Die auf diese Weise gewonnenen Daten enthalten die von dem Nutzer abgespielten Songtitel, welche durch Informationen zu ihren musikalischen Merkmalen angereichert und zu Musikpräferenzen aggregiert werden können. Diese Aggregation kann über verschiedene Zeiträume geschehen und erlaubt es sowohl allgemeine Vorlieben als auch die momentane Musikauswahl abzubilden. Dadurch lassen sich Musikpräferenzen nun im echten Leben erfassen, anstatt diese wie bisher in der psychologischen Forschung üblich durch traditionelle Ansätze wie Fragebögen oder experimentelle Settings zu erheben. Neben dieser passiven Form der Datenerhebung erlauben Smartphones auch den Einsatz von wiederholten kurzen Fragebögen, sogenannten „Experience Samplings", die kontextuelle Informationen zum Erleben während des Musikhörens abfragen können.

Die vorliegende Dissertation nutzt diese neuen technischen Möglichkeiten, um inter- und intraindividuellen Unterschiede im natürlichen Musikhörverhalten am Smartphone zu erforschen und mit psychologischen Konstrukten in Beziehung zu setzen. Sie umfasst zwei empirische Studien, die moderne computergestützte Methoden verwenden, um natürliche Musikpräferenzen effizient zu erfassen, numerisch abzubilden und angemessen zu modellieren. Dabei folgt die Dissertation mit einer initialen explorativen Studie und einer darauf aufbauenden konfirmatorischen Studie dem Zyklus empirischer Forschung, um neue Erkenntnisse zu gewinnen und frühere Befunde des Feldes zu konsolidieren.

Die erste Studie untersuchte interindividuelle Unterschiede in allgemeinen Musikpräferenzen und deren Zusammenhang mit stabilen Persönlichkeitseigenschaften. Sie analysierte Musikhördaten von 330 Studienteilnehmern aus einem bestehenden Smartphone Sensing-Datensatz, der über 3 bis 85 Studientage gesammelt wurde. Die durchschnittlichen

Musikpräferenzen der Probanden wurden anhand einer Vielzahl von Audio- und Textmerkmalen ihrer abgespielten Songs quantifiziert. Die technischen Audiomerkmale (z. B. Tempo, Tanzbarkeit) wurden vom Streamingdienst Spotify abgerufen, der diese maschinell aus den Audioaufzeichnungen der jeweiligen Songs extrahierte, und die Textmerkmale (z. B. Emotionswörter der Wut, romantische Themen) wurden mithilfe verschiedener Sprachverarbeitungsalgorithmen aus den Songtexten gewonnen. Zusätzlich wurden einige Indikatoren für Musikhörgewohnheiten einbezogen, die beispielsweise die tägliche Dauer des Musikhörens abbildeten. Insgesamt wurden 844 Variablen extrahiert, die dazu dienten, die Big Five-Dimensionen der Persönlichkeit auf der Ebene der Domänen und Facetten vorherzusagen. In einem Machine Learning-Benchmark wurde die kreuzvalidierte Vorhersagegüte von linearen Elastic Net-Regressionen und nicht-linearen Random Forest-Algorithmen verglichen, welche beide in der Lage sind, Daten zu modellieren, bei denen die Anzahl der Prädiktoren größer ist als die der Beobachtungen. Die Ergebnisse zeigten, dass nur die Big Five-Domäne Offenheit erfolgreich (d. h., signifikant besser als durch Zufall) vorhergesagt werden konnte, während die Domäne Gewissenhaftigkeit sowie mehrere Persönlichkeitsfacetten zwar nicht-signifikante, aber kleine bis moderate Vorhersageleistungen aufwiesen. Dabei waren die nicht linearen Random Forest-Modelle insgesamt leicht überlegen. Da diese Algorithmen nicht per se interpretierbar sind, wurden weiterführenden Analysen durchgeführt, um den individuellen Beitrag von audio- und textbasierten Musikpräferenzen zur Vorhersage von Persönlichkeit zu vergleichen und die Bedeutung einzelner Variablen zu untersuchen. Es zeigte sich, dass Präferenzen für Melodien und Songtexte ähnlich informativ für die Vorhersage von Offenheit waren, während die Präferenzen für Textmerkmale die wichtigere Rolle für die Gewissenhaftigkeits-Modelle spielten. Die relevantesten Prädiktorvariablen der jeweiligen Modelle waren in einer Weise mit den Big Five-Dimensionen verbunden, die mit deren inhaltlichen Definitionen kongruent war.

Die zweite Studie untersuchte inter- und intraindividuelle Unterschiede in momentanen Musikpräferenzen, das heißt in der Musikauswahl von Moment zu Moment, und wie diese mit stabilen Persönlichkeitseigenschaften und fluktuierenden Stimmungszuständen in Zusammenhang stehen. Ihr Ziel bestand darin, verschiedene Hypothesen zu diesen Zusammenhängen (siehe weiter unten) zu überprüfen, die basierend auf theoretischen Überlegungen und der Literatur zu Musikpräferenzen und Emotionsregulation vorab präregistriert wurden. Zu diesen Zweck wurde eine 14-tägige Längsschnittstudie durchgeführt, die 1.631 Musikhörereignisse von 110 Teilnehmern sammelte. Die Studie integrierte aktive und passive Strategien der ambulanten Datenerfassung am Smartphone. Sie zeichnete das

Musikhörverhalten der Probanden automatisch mittels Smartphone Sensing auf und reicherte die Daten mit den selbstberichteten Stimmungszuständen aus zeitlich zugehörigen Experience Samplings an. Auf Grundlage der Songs, welche die Teilnehmer innerhalb des 30-minütigen Zeitfensters um die Stimmungsberichte abspielten, wurden ihre momentanen Musikpräferenzen extrahiert und mithilfe der technische Audiomerkmale Valenz und Energie quantifiziert. Wie bereits in der ersten Studie wurden die Audiomerkmale vom Streamingdienst Spotify abgerufen. Schließlich wurden die beiden Aspekte der momentanen Musikauswahl jeweils in einem Multilevel Regressionsmodell anhand von Persönlichkeitseigenschaften, Stimmungszuständen und deren Interaktionen vorhergesagt. Entsprechend der Hypothesen wurde erwartet, dass die gewählte musikalische Valenz und Energie im Schnitt kongruent zu den inhaltlichen Definitionen der Big Five-Persönlichkeitseigenschaften (z. B. positivere Musik bei höherer Extraversion) und auf momentaner Ebene kongruent zu Valenz und Aktivierungsgrad der aktuellen Stimmung (z. B. positivere Musik bei besserer Stimmung) sind. Dabei sollten Aspekte der Persönlichkeit die Stimmungskongruenz moderieren, da Musik je nach Eigenschaftsausprägungen unterschiedlich für die Emotionsregulation genutzt wird (z. B. weniger Stimmungskongruenz bei höherem Neurotizismus). Die inferenzstatistischen Ergebnisse zeigten jedoch, dass Persönlichkeit und Stimmung nur einen sehr kleinen Teil der Varianz in den Musikpräferenzen erklärten. Insbesondere wurde nur ein einziger signifikanter, aber schwacher Effekt gefunden, nämlich ein positiver Zusammenhang zwischen dem momentanen Aktivierungsgrad der Hörer und ihrer gewählten musikalischen Energie. Dieser Effekt stand in Einklang mit der vorab angenommenen Stimmungskongruenz und wies darauf hin, dass Probanden in stärker aktivierten Stimmungszuständen im Schnitt eher energiegeladene Songs hörten. Darüber hinaus zeigten die Modelle keine der erwarteten Kongruenz- oder Interaktionseffekte für die bevorzugte musikalische Valenz oder Energie. Dieses Bild sollte in Anbetracht der begrenzten Stichprobengröße und der damit einhergehenden geringen statistischen Power jedoch mit Vorsicht interpretiert werden.

Mithilfe neuer computergestützter Methoden lieferten die beiden vorgestellten Studien neue Einblicke in das natürliche Musikhörverhalten am Smartphone und bestätigten einige empirische Befunde aus der Vergangenheit. Allerdings zeigten sie weniger und schwächere Effekte als vorige Studien, in denen Musikpräferenzen mit traditionellen Methoden wie Selbstberichten (z. B. zur Bewertung verschiedener Genres) oder Musikhör-Experimenten (z. B. zur Bewertung experimentell manipulierter Musikstücke) erhoben wurden. Diese Diskrepanz deutet darauf hin, dass das Musikhörverhalten im realen Leben komplexer ist als einmalig abgegebene Präferenzbewertungen. Da es jedoch aus praktischen Gründen lange Zeit kaum möglich war, Verhalten im Alltag zu untersuchen, besteht nach wie vor ein Mangel an natürlichen Verhaltensdaten – sowohl in der Forschung zu Musikpräferenzen als auch in der Persönlichkeitspsychologie allgemein, sodass diese Dissertation nur einen der ersten Schritte in Richtung des Verständnisses natürlichen Musikhörverhaltens darstellt. Der hier vorgestellte methodische Ansatz kann jedoch auf vielfältige Weise ausgebaut werden, um weiterführende Erkenntnisse zu sammeln. Smartphones erwiesen sich hier als nützliche Hilfsmittel zur Anwendung von passiver und aktiver ambulanter Datenerfassung, die sich in Kombination dazu eignen, objektive Daten zum Musikhörverhalten und subjektive kontextuelle Daten zum Hörer und der Hörsituation zu erheben. Während in dieser Dissertation lediglich Personenvariablen im Zusammenhang mit Musikpräferenzen betrachtet wurden, kann der Smartphone-basierte Ansatz auch situative Variablen erfassen, die ebenfalls eine Rolle im Musikhörverhalten spielen könnten. Beispielsweise können Smartphones dank ihrer inhärenten Sensorik sowohl objektive Situationsparameter erheben (z. B. den aktuellen Ort anhand von GPS-Sensoren) als auch deren Wahrnehmung in Experience Samplings abfragen (z. B. die wahrgenommene Geselligkeit). Damit birgt der hier vorgeschlagene methodische Ansatz das Potenzial, die komplette Triade der Persönlichkeit am Beispiel von Musikhören abzubilden, nämlich das Verhalten, die Person und die Situation. Neben dem theoretischen Wert für die Persönlichkeitspsychologie bieten derartige Forschungsansätze möglicherweise auch einen praktischen Nutzen für die Verbesserung von automatisierten Musikempfehlungen durch Einbezug kontextueller Parameter.

Neben den vielfältigen Vorteilen der Smartphone-basierten Untersuchung von Musikhörverhalten unterliegt der hier vorgestellte Ansatz auch einigen Limitationen. Beispielsweise verwenden nach wie vor nicht alle Menschen ihr Smartphone zum Musikhören, sodass derartige Studiendesigns bestimmte Personengruppen (z. B. basierend auf Alter, technischer Affinität oder sozioökonomischen Status) systematisch ausgrenzen und damit zu verzerrten Stichproben führen. Auch nutzen viele Leute neben dem Smartphone noch andere Geräte, um Musik zu hören, wobei systematische Unterschiede im Nutzungsverhalten zwischen den Geräten denkbar sind. Dies könnte unter anderem der Fall sein, falls unterschiedliche Musikgeräte in verschiedenen Situationen benutzt werden (z. B. Smartphones unterwegs und stationäre Geräte daheim). Neben diesen feld-spezifischen Problemen stellt die praktische Durchführung von Smartphone Sensing-Studien generell eine administrative und technische Herausforderung für Forscher in der Psychologie dar. Einerseits erfordern die Entwicklung und Instandhaltung von entsprechenden Forschungsapps mit Sensing-Funktionalität eine intensive interdisziplinäre Zusammenarbeit mit Informatikern. Andererseits produzieren derartige Apps große Mengen unstrukturierter digitaler Daten, deren Vorverarbeitung zeitaufwendig ist und umfangreiche statistische Kenntnisse der Forscher bedarf. Zudem müssen bei jedem dieser Schritte die Datenschutzrechte der Studienteilnehmer gewahrt werden, indem zum Beispiel besonders sensible Informationen (z. B. Inhalte von Textnachrichten) direkt in aggregierter statt in Rohform gespeichert werden und Daten nur auf sicheren Servern abgerufen werden. Ungeachtet dieser Hürden sollten Smartphone Sensing-Studien stets mit den Grundsätzen offener Wissenschaft in Einklang gebracht werden, wie es im Rahmen diese Dissertation bestmöglich versucht wurde. Zum Beispiel sollten geplante Analysen präregistriert werden, selbst wenn konkrete Schritte der Datenverarbeitung aufgrund von Sensing-Problemen nicht vorhersehbar sind und die Registrierung auf einer höheren Abstraktionsebene ablaufen muss. Auch bei der Offenlegung der sensiblen Daten sollten Kompromisse gefunden werden, indem beispielsweise statt Rohdaten zumindest die aggregierten Daten für formale Analysen veröffentlicht werden.

Trotz einiger Herausforderungen zeigten die empirischen Studien dieser Dissertation, dass Smartphones ein nützliches Hilfsmittel bei der Erforschung von natürlichem Musikhörverhalten sind und generell helfen können, den Mangel an natürlichen Verhaltensdaten in der Psychologie schrittweise zu beheben. Die vorgestellten methodischen Ansätze sollten somit weiterverfolgt werden.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ALE .... *Accumulated Local Effects*
API........ *Application Programming Interface*
BERT .... *Bidirectional Encoder Representations from Transformers*
BFI-2-S .... *Big Five Inventory-2 Short Version*
BFSI........ *Big Five Structure Inventory*
CI .... *Confidence Interval*
CV .... *Cross-Validation*
ES .... *Experience Sampling*
GPS........ *Global Positioning System*
ICC .... *Intra-Class Correlation*
iOS........ *iPhone Operating System*
LDA........ *Latent Dirichlet Allocation*
MLM .... *Multilevel Regression Model*
MSE........ *Mean Squared Error*
NLP .... *Natural Language Processing*
NRC........ *National Research Council*
OSF........ *Open Science Framework*
OSM .... *Online Supplemental Material*
$R^2_{(m)}$ .... *Marginal Coefficient of Determination in MLMs*
r_s .... *Spearman Rank Order Correlation Coefficient*
WEIRD........ *Western, Educated, Industrialized, Rich, Democratic*

1 General Abstract

The advent of smartphones has undoubtedly revolutionized our day-to-day lives, including the way we consume music. As popular music devices, smartphones have become a digital source of music-listening data, offering an unprecedented opportunity to investigate music preferences "in the wild" rather than relying on traditional approaches like self-report questionnaires or listening experiments. Taking advantage of this development, the present dissertation investigated inter- and intraindividual differences in natural music-listening behavior. Two empirical studies employed smartphone sensing to collect ecologically valid and longitudinal music-listening records. They extracted music preferences based on participants' played songs and computationally quantified them via intrinsic musical characteristics of melodies and lyrics. The first study focused on overall music preferences to make out-of-sample predictions for the Big Five personality traits on the domain and facet levels using supervised machine-learning algorithms. The second study considered momentary music preferences and modeled them from enduring Big Five domains and concurrent valence and arousal states in multilevel regressions, expecting to replicate trait- and state-congruent associations from literature. With these two studies, the dissertation followed an empirical research cycle integrating the complementary statistical approaches of prediction and inference for an initial exploratory and a follow-up confirmatory analysis. It applied state-of-the-art methods to efficiently collect, numerically represent, and jointly model music-listening data. In doing so, the presented studies provided new insights and corroborated past findings on music preferences. However, the studies obtained fewer and weaker effects for natural music-listening behavior than past research relying on traditional assessments, indicating that real-life behaviors are more complex to understand than one-time preference assessments. Because person variables exhibited only weak associations, other factors like situational aspects may play a role in natural music preferences. The methodological framework proposed here has the potential to explore such novel relations but also presents manifold challenges for researchers, as discussed throughout this dissertation.

2 General Introduction

Smartphones have undoubtedly transformed our daily lives, changing the way we communicate with others, take photos, search for directions, and consume music. In particular, the portable size of smartphones and the rise of Internet-based streaming apps have turned them into popular music devices (IFPI, 2019), which grant users the freedom to play any song anytime and anywhere (Bull, 2005; Krause & North, 2016). This digitalization of music consumption has not only increased the relevance of music in listeners' daily lives (IFPI, 2021; Datta et al., 2018) but also opened up new opportunities for examining individual differences in music listening "in the wild." While personality scientists have repeatedly investigated music preferences through traditional approaches like self-reports (e.g., Greenberg et al., 2022; Rentfrow & Gosling, 2003), smartphones have now become the ideal tool for assessing natural music-listening behavior thanks to their inherent sensing capabilities (Harari et al., 2016; Miller, 2012). Filling the general lack of actual behavioral data in personality psychology (Funder, 2001; Furr, 2009), smartphone sensing previously served to investigate other behaviors such as socializing or app usage (e.g., Harari et al., 2020; Stachl et al., 2017), but its application to music listening is still pending.

To bridge this gap, the present dissertation uses smartphone sensing to study music-listening behavior in an ecologically valid setting and over longer periods of time. For the resulting digital listening records, it outlines a preprocessing pipeline that automatically represents overall and momentary music preferences in terms of intrinsic musical characteristics of the played songs. Finally, the dissertation applies sophisticated modeling approaches to examine the inter- and intraindividual differences in music-listening behavior in relation to personality traits and mood states. With two empirical studies, it integrates tools from computational sciences into personality research to efficiently collect, numerically represent, and adequately model music-listening behavior.

2. 1 Research on Individual Differences in Music Preferences

2.1.1 Rationale and Overview

Streaming platforms like Spotify offer over 100 million songs (Spotify AB, 2023), prompting lay and scientific curiosity about individual music preferences and the factors involved. Accordingly, researchers have spent the past two decades examining the individual differences in music preferences, relating them to various characteristics of listeners, such as their demographics (i.e., age and gender; Anderson et al., 2021; Bonneville-Roussy et al., 2013), cognitive abilities (i.e., intelligence; Bonetti & Costa, 2016; Račevska & Tadinac, 2019), and, most commonly, personality concepts like sensation seeking (Litle & Zuckerman, 1986), the Jungian types (Pearson & Dollinger, 2002), and the most widely established Big Five domains (e.g., Anderson et al., 2021; Delsing et al., 2008; Greenberg et al., 2016, 2022; Rentfrow & Gosling, 2003; Vuoskoski & Eerola, 2011; Qiu et al., 2019). Findings for the Big Five personality domains typically exhibited trait-congruent music preferences, causing researchers to assume the interactionist rationale that music serves to shape auditory environments in a way that aligns with and reinforces their personality (Buss, 1987; Swann, 1987). However, the reported associations were weak across studies (Schäfer & Mehlhorn, 2017), and stable traits could only account for interindividual differences in the music people prefer on average but not for the considerable intraindividual variance in momentary music preferences (see Greb et al., 2019). Momentary music choice, in turn, was repeatedly related to listeners' current mood states in a congruent manner (e.g., Chen et al., 2007; Greb et al., 2019; Taruffi & Koelsch, 2014; Thoma et al., 2012), supporting the notion that music commonly serves mood regulation purposes (e.g., DeNora, 1999; Schäfer et al., 2013). In sum, past research related music preferences to both stable personality traits and fluctuating mood states. However, this line of research suffered from a methodological limitation.

2.1.2 Status Quo of Study Designs

Personality research, in general, and the study of music preferences, in particular, suffer from a lack of behavioral data due to their commonly employed study designs (Baumeister et al., 2007; Funder, 2009). Traditionally, overall and momentary music preferences have been measured through self-report questionnaires (e.g., Bonneville-Roussy et al., 2013; Delsing et al., 2008; Rentfrow & Gosling, 2003; Taruffi & Koelsch, 2014) or listening experiments (e.g., Chen et al., 2007; Greenberg et al., 2016, 2022; Knobloch & Zillman, 2002; Ladinig & Schellenberg, 2012; Nave et al., 2018; Vuoskoski & Eerola, 2011) but both of these approaches may not accurately reflect preferences in natural listening behavior in daily life.

Self-reports, on the one hand, require participants to perform a complex introspection process (Tourangeau et al., 2000), which is prone to different biases like socially desirable responding or memory limitations (Paulhus & Vazire, 2007; Podsakoff et al., 2003). General music preference questionnaires, in particular, ask participants to rate their liking of different musical genres (e.g., Litle & Zuckerman, 1986; Rentfrow & Gosling, 2003). When giving their responses, participants may not necessarily reflect on their past music-listening instances to generate average preferences but instead draw on their self-views to make preference ratings consistent with those (Baumeister, 1982; Paulhus, 1984). In contrast, self-reports of contextual variations in listening behaviors (e.g., in relation to mood states) may be difficult to remember given the high-frequency nature of music consumption (Stein et al., 2013). Supporting these notions, past studies repeatedly reported a mismatch between self-reports and observations for various other types of behavior (e.g., Gosling et al., 1998; Holt & Laury, 2002; Junco, 2013; Kormos & Gifford, 2014).

On the other hand, listening experiments asking participants to rate or choose from different musical excerpts presented without context or after mood induction in laboratory or online settings are generally low in ecological validity (Greenberg & Rentfrow, 2017). That is because the selection of songs provided is highly restricted, consisting of either very popular, artificially manipulated, or unreleased tracks, which are prototypical for certain genres (e.g., Classical, Rock) or musical characteristics (e.g., fast tempo, sad valence) but do not represent the natural music market (Greenberg & Rentfrow, 2017). Beyond that, experimental listening situations lack natural context (e.g., activities while listening to music) and timely dynamics to reveal any information on listening habits (e.g., how long people listen to certain music or music in general).

2.2 Smartphones to the Rescue: The Ambulatory Assessment of Music-Listening Behavior

Collecting behavioral data in the field has long been time-consuming, expensive, intrusive, and, hence, practically infeasible to study natural music-listening behavior (see Baumeister et al., 2007; Furr, 2009). However, the advent of smartphones as popular music-listening devices has opened up new opportunities to study music preferences "in the wild" (IFPI, 2019). As computationally powerful tools, smartphones allow researchers to apply ambulatory assessments, which is an umbrella term comprising a range of technological methods to study people in their natural environments, including in-situ self-reports, behavioral observation, and physiological monitoring (Conner & Mehl, 2015; Trull & Ebner-Priemer, 2014; Wrzus & Mehl, 2015). In particular, smartphones can integrate two forms of ambulatory assessment, namely passive sensing and active sampling, to investigate individual differences in music listening (Conner & Mehl, 2015; Wrzus & Mehl, 2015).

2.2.1 Smartphone Sensing

Smartphone sensing denotes the passive data collection from the system logs (e.g., calling or app usage records) and onboard sensors (e.g., accelerometer, Global Positioning System [GPS], light sensor) of regular off-the-shelf smartphones via designated research apps (Conner & Mehl, 2015; Harari et al., 2016, 2021; Miller, 2012; Schoedel & Mehl, in press). Once installed, sensing apps remain in the background while participants use their smartphones as usual, serving as a digitally mediated behavioral observation that is unobtrusive and low in reactivity (Harari et al., 2016; Miller, 2012; Schoedel & Mehl, in press). In this vein, smartphone sensing can access the songs participants play on their private devices in everyday life, creating longitudinal music-listening records with a high temporal resolution (Harari et al., 2017; Miller, 2012; Schoedel & Mehl, in press; Wrzus & Mehl, 2015). These ecologically valid listening records can be automatically enriched with psychologically meaningful information about the melodies and lyrics of played songs, using tools from Music Information Retrieval or Natural Language Processing (NLP, e.g., Fricke et al., 2018; Qiu et al., 2019). Afterward, listening records can be aggregated over different time spans (e.g., the entire study duration, per day, per listening event) to obtain average or momentary music preferences, allowing researchers to study music preferences across and within persons (Carpenter et al., 2016; Schoedel & Mehl, in press; Harari et al., 2021). Furthermore, behavioral listening records provide insights into general listening habits such as the time spent listening per day (Greenberg & Rentfrow, 2017). While the method has recently gained relevance in psychological research on various behaviors (e.g., Ai et al., 2019; Harari et al., 2020; Montag et al., 2014; Stachl et al., 2017), smartphone sensing has been rarely applied to music-listening behavior and mostly in the field of human-computer interaction to develop music recommender systems (e.g., Gillhofer & Shedl, 2015; Yang & Teng, 2015).

2.2.2 Experience Sampling

Beyond their sensing capabilities, smartphones can also administer experience samplings (ES) and actively ask participants to repeatedly fill out short questionnaires via text messages or notifications throughout the day (Barret & Barett, 2001; Csikszentmihalyi & Larson, 1987; Van Berkel et al., 2017). This form of ambulatory assessment is most suitable for examining subjective experiences like thoughts or feelings (Wrzus & Mehl, 2015), which can only be assessed via self-report and not passive sensing (Paulhus & Vazire, 2007). Here, the in-situ assessment of ES reduces response biases commonly found in self-reports (Lucas et al., 2020; Neubauer et al., 2020). Hence, ES are not ideal for collecting objective behavioral data on music preferences but for investigating contextual factors like listeners' current mood states (see Greb et al., 2019; Randall & Rickard, 2017). In particular, the combination of ES with smartphone sensing allows for event-contingent or event-triggered sampling schedules, whereby questionnaires are triggered whenever music-listening listening behavior is sensed (Conner et al., 2007; Van Berkel et al., 2017).

In sum, smartphones can be used to obtain objective music-listening data, but also subjective contextual self-reports in the field, providing multimethod data for the study of individual differences in natural listening behavior. Using participants' private smartphones for ambulatory assessment not only fosters ecological validity but is also unintrusive, financially economical, and environment-friendly as there is no need to equip participants with special research devices that are burdensome to carry around and expensive to acquire and maintain.

2.3 The Present Dissertation

2.3.1 Rationale

As outlined above, smartphones have laid the groundwork for the ecologically valid assessment of music-listening behavior, pushing the boundaries of music research in personality science. The present dissertation seizes this opportunity and presents two empirical studies that used the research app "PhoneStudy" from LMU Munich to sense digital music-listening records from participants' smartphones. Both studies investigated individual differences in everyday music preferences, aiming to provide new insights and corroborate past findings. More specifically, the first study focused on interindividual differences in *overall* music preferences in relation to personality traits, and the second study additionally considered intraindividual fluctuations in *momentary* music preferences in relation to mood states. For this purpose, the second study combined passive and active ambulatory assessment, enriching smartphone-sensed music-listening records with experience-sampled mood states. In both studies, music preferences were assessed based on automatically extracted attributes of participants' played songs. Thereby, the first study explored an extensive set of *various* audio (e.g., tempo, danceability) and lyrics characteristics (e.g., topic love, angry emotionality), while the second study focused on two *selected* audio characteristics (i.e., valence & energy). For relating these music preferences to listeners' variables, the two studies relied on the complementary statistical approaches of *prediction* and *inference* (Breiman, 2001). The first study applied a supervised machine-learning approach to make out-of-sample predictions for the Big Five personality domains and facets based on overall music preferences and to explore the contribution of different aspects of music listening. In contrast, the second study employed multilevel regression models to predict momentary music preferences from enduring personality traits and concurrent mood states, expecting to replicate trait-congruent associations with the Big Five personality domains and mood-congruent associations with affective valence and arousal. In doing so, the dissertation followed the academic cycle (Yarkoni & Westfall, 2017), starting with an exploratory study and following up with a confirmatory study.

2.3.2 Overview of Papers and Author Contributions

The present dissertation comprises two empirical studies, the first of which has already been published in a peer-reviewed journal and the second of which is being prepared for publication. The creator of this dissertation is the first author and primary contributor to both articles. However, other authors have also made meaningful contributions to these studies as outlined in Table 1.1.

Table 1.1

Author Contributions to the Articles of the Dissertation

Study	Author Contributions
Study 1:	
Sust, L., Stachl, C., Kudchadker, G., Bühner, M., & Schoedel, R. (2023). Personality Computing with Naturalistic Music Listening Behavior: Comparing Audio and Lyrics Preferences. *Collabra: Psychology*, *9*(1). https://doi.org/10.1525/collabra.75214	C.S. contributed initial research idea. **L.S.** designed research. C.S. and R.S. provided data for secondary use. **L.S.** and G.K. preprocessed data. **L.S.** conducted data analysis. **L.S.** wrote and revised the manuscript. C.S., M.B., & R.S. gave feedback to the manuscript. M.B. provided resources.
Study 2:	
Sust, L., & Schoedel, R. (in preparation). Explaining Everyday Music Choice on Smartphones: The Role of Personality Traits and Mood States.	**L.S.** designed research. **L.S.** and R.S. conducted research. R.S. provided code snippets for data preprocessing. **L.S.** preprocessed data. **L.S.** conducted data analysis. **L.S.** wrote and revised the manuscript. R.S. gave feedback to the manuscript.

Note. Contributions of the dissertation's author are in bold.

2.3.3 Open Science Statement

The two studies comprising this dissertation adhere to the principles of open science in the following ways. Study 1 was exploratory and not preregistered because the data had already been used in numerous prior publications. Study 2 was confirmatory and preregistered prior to data preprocessing and analysis. For both articles, the author provides open code and open aggregated data in the respective project repositories on the Open Science Framework (OSF), which are linked throughout this dissertation. However, the raw smartphone-sensing data cannot be made public in order to protect the privacy rights of the participants.

2.4 References

Ai, P., Liu, Y., & Zhao, X. (2019). Big five personality traits predict daily spatial behavior: Evidence from smartphone data. *Personality and Individual Differences, 147*, 285–291. https://doi.org/10.1016/j.paid.2019.04.027

Anderson, I., Gil, S., Gibson, C., Wolf, S., Shapiro, W., Semerci, O., & Greenberg, D. M. (2021). Just the way you are: Linking music listening on Spotify and personality. *Social Psychological and Personality Science, 12*(4), 561-572. https://doi.org/10.1177/1948550620923228

Barrett, L. F., & Barrett, D. J. (2001). An introduction to computerized experience sampling in psychology. *Social Science Computer Review, 19*(2), 175-185. https://doi.org/10.1177/089443930101900204

Baumeister, R. F. (1982). A self-presentational view of social phenomena. *Psychological Bulletin, 91*(1), 3-26. https://doi.org/10.1037/0033-2909.91.1.3

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science*, *2*(4), 396-403. https://doi.org/10.1111/j.1745-6916.2007.00051.x

Bonetti, L., & Costa, M. (2016). Intelligence and musical mode preference. *Empirical Studies of the Arts, 34*(2), 160-176. https://doi.org/10.1177/0276237416628907

Bonneville-Roussy, A., Rentfrow, P. J., Xu, M. K., & Potter, J. (2013). Music through the ages: Trends in musical engagement and preferences from adolescence through middle adulthood. *Journal of Personality and Social Psychology, 105*(4), 703. https://doi.org/10.1037/a0033770

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5-32. https://doi.org/10.1023/A:1010933404324

Bull, M. (2005). No dead air! The iPod and the culture of mobile listening. *Leisure Studies*, *24*(4), 343-355. https://doi.org/10.1080/0261436052000330447

Buss, D. M. (1987). Selection, evocation, and manipulation. *Journal of Personality and Social Psychology, 53*(6), 1214-1221. https://doi.org/10.1037/0022-3514.53.6.1214

Carpenter, R. W., Wycoff, A. M., & Trull, T. J. (2016). Ambulatory assessment: New adventures in characterizing dynamic processes. *Assessment, 23*(4), 414–424. https://doi.org/10.1177/1073191116632341

Chen, L., Zhou, S., & Bryant, J. (2007). Temporal changes in mood repair through music consumption: Effects of mood, mood salience, and individual differences. *Media Psychology*, *9*(3), 695-713. https://doi.org/10.1080/15213260701283293

Conner, T. S., Barrett, L. F., Tugade, M. M., & Tennen, H. (2007). Idiographic personality: The theory and practice of experience sampling. In R. W. Robbins, R. C. Fraley, & R. F. Krieger (Eds.), *Handbook of Research Methods in Personality Psychology* (pp. 79-96). Guilford Press.

Conner, T. S., & Mehl, M. R. (2015). Ambulatory assessment: Methods for studying everyday life. In R. A. Scott & S. M. Kosslyn (Eds.), *Emerging Trends in the Social and Behavioral Sciences* (1^{st} ed., pp. 1-15). Wiley. https://doi.org/10.1002/9781118900772.etrds0010

Csikszentmihalyi, M. & Larson R. (1987) Validity and Reliability of the Experience-Sampling Method. *The Journal of Nervous and Mental Disease, 175*(9), 526-536. https://doi.org/10.1097/00005053-198709000-00004

Datta, H., Knox, G., & Bronnenberg, B. J. (2018). Changing their tune: How consumers' adoption of online streaming affects music consumption and discovery. *Marketing Science*, *37*(1), 5-21. https://doi.org/10.1287/mksc.2017.1051

Delsing, M. J., Ter Bogt, T. F., Engels, R. C., & Meeus, W. H. (2008). Adolescents' music preferences and personality characteristics. *European Journal of Personality, 22*(2), 109-130. https://doi.org/10.1002/per.665

DeNora, T. (1999). Music as a technology of the self. *Poetics*, *27*(1), 31-56. https://doi.org/10.1016/S0304-422X(99)00017-0

Fricke, K. R., Greenberg, D. M., Rentfrow, P. J., & Herzberg, P. Y. (2018). Computer-based music feature analysis mirrors human perception and can be used to measure individual music preference. *Journal of Research in Personality, 75*, 94-102. https://doi.org/10.1016/j.jrp.2018.06.004

Funder, D. C. (2001). Personality. *Annual Review of Psychology, 52*, 197-221. https://doi.org/10.1146/annurev.psych.52.1.197

Funder, D. C. (2009). Persons, behaviors and situations: An agenda for personality psychology in the postwar era. *Journal of Research in Personality, 43*(2), 120-126. https://doi.org/10.1016/j.jrp.2008.12.041

Furr, R. M. (2009). Personality psychology as a truly behavioural science. *European Journal of Personality*, *23*(5), 369-401. https://doi.org/10.1002/per.724

Gillhofer, M., & Schedl, M. (2015). Iron maiden while jogging, debussy for dinner? An analysis of music listening behavior in context. In X. He, S. Luo, D. Tao, C. Xu, J. Yang, M. A. Hasan (Eds.), *Proceedings of the 21st International Conference on MultiMedia Modeling* (pp. 380-391). Springer International Publishing.

Gosling, S. D., John, O. P., Craik, K. H., & Robins, R. W. (1998). Do people know how they behave? Self-reported act frequencies compared with on-line codings by observers. *Journal of Personality and Social Psychology, 74*(5), 1337-1349. https://doi.org/10.1037/0022-3514.74.5.1337

Greb, F., Steffens, J., & Schlotz, W. (2019). Modeling music-selection behavior in everyday life: a multilevel statistical learning approach and mediation analysis of experience sampling data. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.00390

Greenberg, D. M., Kosinski, M., Stillwell, D. J., Monteiro, B. L., Levitin, D. J., & Rentfrow, P. J. (2016). The song is you: Preferences for musical attribute dimensions reflect personality. *Social Psychological and Personality Science, 7*(6), 597-605. https://doi.org/10.1177/1948550616641473

Greenberg, D. M., & Rentfrow, P. J. (2017). Music and big data: A new frontier. *Current Opinion in Behavioral Sciences, 18*, 50-56. https://doi.org/10.1016/j.cobeha.2017.07.007

Greenberg, D. M., Wride, S. J., Snowden, D. A., Spathis, D., Potter, J., & Rentfrow, P. J. (2022). Universals and variations in musical preferences: A study of preferential reactions to Western music in 53 countries. *Journal of Personality and Social Psychology, 122*(2), 286-309. https://doi.org/10.1037/pspp0000397

Harari, G. M., Lane, N. D., Wang, R., Crosier, B. S., Campbell, A. T., & Gosling, S. D. (2016). Using smartphones to collect behavioral data in psychological science: Opportunities, practical considerations, and challenges. *Perspectives on Psychological Science*, *11*(6), 838-854. https://doi.org/10.1177/1745691616650285

Harari, G. M., Müller, S. R., Aung, M. S., & Rentfrow, P. J. (2017). Smartphone sensing methods for studying behavior in everyday life. *Current Opinion in Behavioral Sciences*, *18*, 83-90. https://doi.org/10.1016/j.cobeha.2017.07.018

Harari, G. M., Müller, S. R., Stachl, C., Wang, R., Wang, W., Bühner, M., ... & Gosling, S. D. (2020). Sensing sociability: Individual differences in young adults' conversation, calling, texting, and app use behaviors in daily life. *Journal of Personality and Social Psychology, 119*(1), 204-228. https://doi.org/10.1037/pspp0000245

Harari, G. M., Stachl, C., Müller, S. R., & Gosling, S. D. (2021). Mobile sensing for studying personality dynamics in daily life. In J. Rauthmann (Ed.), *The Handbook of Personality Dynamics and Processes* (pp. 763-790). Academic Press. https://doi.org/10.1016/B978-0-12-813995-0.00029-7

Holt, C.A., & Laury, S. (2002). Risk aversion and incentive effects. *American Economic Review, 92*(5), 1644-1655. https://doi.org/10.1257/000282802762024700

IFPI. (2019). Global Music Listening Report 2019. (n.p.). https://www.ifpi.org/resources/

IFPI. (2021). Engaging with Music. (n.p.). https://www.ifpi.org/resources/

Junco, R. (2013). Comparing actual and self-reported measures of Facebook use. *Computers in Human Behavior, 29(3), 626–631.* https://doi.org/*10.1016/j.chb.2012.11.007*

Knobloch, S., & Zillmann, D. (2002). Mood management via the digital jukebox. *Journal of Communication*, *52*(2), 351-366. https://doi.org/10.1111/j.1460-2466.2002.tb02549.x

Kormos, C., & Gifford, R. (2014). The validity of self-report measures of proenvironmental behavior: A meta-analytic review. *Journal of Environmental Psychology, 40*, 359-371. https://doi.org/10.1016/j.jenvp.2014.09.003

Krause, A. E., & North, A. C. (2016). Music listening in everyday life: Devices, selection methods, and digital technology. *Psychology of Music*, *44*(1), 129-147. https://doi.org/10.1177/0305735614559065

Ladinig, O., & Schellenberg, E. G. (2012). Liking unfamiliar music: Effects of felt emotion and individual differences. *Psychology of Aesthetics, Creativity, and the Arts*, *6*(2), 146-154. https://doi.org/10.1037/a0024671

Litle, P., & Zuckerman, M. (1986). Sensation seeking and music preferences. *Personality and Individual Differences, 7*(4), 575-578. https://doi.org/10.1016/0191-8869(86)90136-4

Lucas, R. E., Wallsworth, C., Anusic, I., & Donnellan, B. (2020). A direct comparison of the day reconstruction method and the experience sampling method. *Journal of Personality & Social Psychology, 120*(3)*,* 816–835. https://doi.org/10.1037/pspp0000289

Miller, G. (2012). The smartphone psychology manifesto. *Perspectives on Psychological Science, 7* (3), 221–237. https://doi.org/10.1177/1745691612441215

Montag, C., Błaszkiewicz, K., Lachmann, B., Andone, I., Sariyska, R., Trendafilov, B., Reuter, M., & Markowetz, A. (2014). Correlating personality and actual phone usage. *Journal of Individual Differences, 35*(3), 158–165. https://doi.org/10.1027/1614-0001/a000139

Nave, G., Minxha, J., Greenberg, D. M., Kosinski, M., Stillwell, D., & Rentfrow, J. (2018). Musical preferences predict personality: Evidence from active listening and Facebook likes. *Psychological Science, 29*(7), 1145-1158. https://doi.org/10.1177/0956797618761659

Neubauer, A. B., Scott, S. B., Sliwinski, M. J., & Smyth, J. M. (2020). How was your day? Convergence of aggregated momentary and retrospective end-of-day affect ratings across the adult life span. *Journal of Personality and Social Psychology, 119*(1), 185–203. https://doi.org/10.1037/pspp0000248

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46*(3), 598-609. https://doi.org/10.1037/0022-3514.46.3.598

Paulhus, D. L., & Vazire, S. (2007). The self-report method. In R. W. Robins, R. C. Fraley & R. C. Kruger (Eds.), *Handbook of Research Methods in Personality Psychology* (pp. 224-239). Guilford Press.

Pearson, J. L., & Dollinger, S. J. (2004). Music preference correlates of Jungian types. *Personality and Individual Differences, 36*(5), 1005-1008. https://doi.org/10.1016/S0191-8869(03)00168-5

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879-903. https://doi.org/10.1037/0021-9010.88.5.879

Qiu, L., Chen, J., Ramsay, J., & Lu, J. (2019). Personality predicts words in favorite songs. *Journal of Research in Personality*, *78*, 25-35. https://doi.org/10.1016/j.jrp.2018.11.004

Račevska, E., & Tadinac, M. (2019). Intelligence, music preferences, and uses of music from the perspective of evolutionary psychology. *Evolutionary Behavioral Sciences, 13*(2), 101-110. https://doi.org/10.1037/ebs0000124

Randall, W. M., & Rickard, N. S. (2017). Personal music listening: A model of emotional outcomes developed through mobile experience sampling. *Music Perception: An Interdisciplinary Journal*, *34*(5), 501-514. https://doi.org/10.1525/mp.2017.34.5.501

Rentfrow, P. J., & Gosling, S. D. (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), 1236-1256. https://doi.org/10.1037/0022-3514.84.6.1236

Schäfer, T., & Mehlhorn, C. (2017). Can personality traits predict musical style preferences? A meta-analysis. *Personality and Individual Differences, 116*, 265-273. https://doi.org/10.1016/j.paid.2017.04.061

Schäfer, T., Sedlmeier, P., Städtler, C., & Huron, D. (2013). The psychological functions of music listening. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00511

Schoedel, R., & Mehl, M. R. (in press). Mobile sensing methods. In H. T. Reis, T. West, & C. M. Judd (Eds.), *Handbook of Research Methods in Social and Personality Psychology* (3rd ed.). Cambridge University Press.

Spotify AB. (2023). About Spotify, *For the Record*. https://newsroom.spotify.com/company-info/

Stachl, C., Hilbert, S., Au, J. Q., Buschek, D., De Luca, A., Bischl, B., Hussmann, H., & Bühner, M. (2017). Personality traits predict smartphone usage. *European Journal of Personality, 31*(6), 701-722. https://doi.org/10.1002/per.2113

Stein, N. L., Ornstein, P. A., Tversky, B., & Brainerd, C. (2013). *Memory for Everyday and Emotional Events*. Psychology Press.

Swann, W. B. (1987). Identity negotiation: Where two roads meet. *Journal of Personality and Social Psychology, 53*(6), 1038-1051. https://doi.org/10.1037/0022-3514.53.6.1038

Taruffi, L., & Koelsch, S. (2014). The paradox of music-evoked sadness: An online survey. *PLoS One*, *9*(10): e110490. https://doi.org/10.1371/journal.pone.0110490

Thoma, M. V., Ryf, S., Mohiyeddini, C., Ehlert, U., & Nater, U. M. (2012). Emotion regulation through listening to music in everyday situations. *Cognition & Emotion*, *26*(3), 550-560. https://doi.org/10.1080/02699931.2011.595390

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The Psychology of Survey Response.* Cambridge University Press. https://doi.org/10.1017/CBO9780511819322

Trull, T. J., & Ebner-Priemer, U. (2014). The role of ambulatory assessment in psychological science. *Current Directions in Psychological Science, 23*(6), 466-470. https://doi.org/10.1177/0963721414550706

Van Berkel, N., Ferreira, D., & Kostakos, V. (2017). The experience sampling method on mobile devices. *ACM Computing Surveys, 50*(6), 1-40. https://doi.org/10.1145/3123988

Vuoskoski, J. K., & Eerola, T. (2011). The role of mood and personality in the perception of emotions represented by music. *Cortex, 47*(9), 1099-1106. https://doi.org/10.1016/j.cortex.2011.04.011

Wrzus, C., & Mehl, M. R. (2015). Lab and/or field? Measuring personality processes and their social consequences. *European Journal of Personality*, *29*(2), 250-271. https://doi.org/10.1002/per.1986

Yang, Y. H., & Teng, Y. C. (2015). Quantitative study of music listening behavior in a smartphone context. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *5*(3), 1-30. https://doi.org/10.1145/2738220

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science, 12*(6), 1100-1122. https://doi.org/10.1177/1745691617693393

3 Study 1: Personality Computing With Naturalistic Music-Listening Behavior

This chapter is an adapted version of the published article “Sust, L., Stachl, C., Kudchadker, G., Bühner, M., & Schoedel, R. (2023). Personality Computing With Naturalistic Music Listening Behavior: Comparing Audio and Lyrics Preferences. *Collabra: Psychology, 9*(1), 75214. https://doi.org/10.1525/collabra.75214”. It was slightly modified in formatting and notation to align with the style of this dissertation. The original published article is under an CC-BY 4.0 license, granting permission to reproduce it here.

3.1 Abstract

It is a long-held belief in psychology and beyond that individuals’ music preferences reveal information about their personality traits. While initial evidence relates self-reported preferences for broad musical styles to the Big Five dimensions, little is known about day-to-day music-listening behavior and the intrinsic attributes of melodies and lyrics that reflect these individual differences. The present study ($N = 330$) proposes a personality computing approach to fill these gaps with new insights from ecologically valid music-listening records from smartphones. We quantified participants’ music preferences via audio and lyrics characteristics of their played songs through technical audio features from Spotify and textual attributes obtained via NLP. Using linear elastic net and non-linear random forest models, these behavioral variables served to predict Big Five personality on domain and facet levels. Out-of-sample prediction performances revealed that – on the domain level – Openness was most strongly related to music listening ($r = .25$), followed by Conscientiousness ($r = .13$), while several facets of the Big Five also showed small to medium effects. Hinting at the incremental value of audio and lyrics characteristics, both musical components were differentially informative for models predicting Openness and its facets, whereas lyrics preferences played the more important role for predictions of Conscientiousness dimensions. In doing so, the models’ most predictive variables displayed generally trait-congruent relationships between personality and music preferences. These findings contribute to the development of a cumulative theory on music listening in personality science and may be extended in numerous ways by future work leveraging the computational framework proposed here.

3.2 Introduction

Music was my first love and it will be the last
Music of the future and music of the past
To live without my music would be impossible to do
In this world of troubles my music pulls me through (Miles, 1976)

Most of us will agree with John Miles' iconic song quote that music plays an important role in our lives. Indeed, we spend nearly one-fourth of our waking time listening to music (Billboard, 2019), and the digitalization of the music market is further increasing these numbers as online streaming services make music more pervasive than ever, with tens of millions of songs accessible anywhere and anytime by over 440 million paid subscribers (IFPI, 2021). This transformation in music consumption has turned streaming platforms and devices into digital sources of music-listening data, creating an unprecedented opportunity to investigate natural music-listening behavior "in the wild" (see Anderson et al., 2021). In doing so, digital listening records provide fine-grained data on various psychologically relevant behavioral outcomes such as music preferences or listening durations (Greenberg & Rentfrow, 2017). Music preferences, in particular, can be automatically represented in terms of the intrinsic properties of the songs played on an everyday basis using tools from computational music information retrieval (e.g., Fricke et al., 2018).

This new ecological validity and granularity in music listening assessment has the potential to push the boundaries of research in personality science, which has long been adopting the interactionist perspective that the music people listen to calibrates their external environments with their personalities and, hence, reflects their individual traits (Greenberg et al., 2020; Rentfrow et al., 2011). Parallel to other types of digital data, such as app usage (Stachl et al., 2017) or social media postings (Schwartz et al., 2013), music-listening records can now be assessed for personality-relevant information via machine-learning algorithms (see Phan & Rauthmann, 2021). The present study adopts this so-called personality computing approach to overcome methodological limitations of the past and model personality from various indicators of natural music listening on smartphones.

3.2.1 Music Listening in Personality Research

Personality researchers have been exploring the associations between music listening and the Big Five personality traits for the past two decades. These studies have mainly focused on individuals' preferences for different styles of music, finding the most robust patterns for the personality dimension of Openness, which correlated positively with preferences for intense (e.g., Rock) and complex (e.g., Classical) musical styles (e.g., Anderson et al., 2021; Bonneville-Roussy et al., 2013; Greenberg et al., 2016; Langmeyer et al., 2012; Nave et al., 2018; Rentfrow & Gosling, 2003).

However, a meta-analysis of the correlation between musical style preferences and personality concluded that the effect sizes for Openness were rather small across studies (r = .12 for intense and r = .21 for complex music), while the remaining Big Five dimensions exhibited average correlations near zero (Schäfer & Mehlhorn, 2017). The studies included in this meta-analysis shared as a limitation tough that they analyzed music preferences via self-reported genre preferences (e.g., Bonneville-Roussy et al., 2013; Rentfrow & Gosling, 2003) or ratings of musical excerpts (e.g., Langmeyer et al., 2012), which may not accurately represent natural music-listening behavior (Greenberg & Rentfrow, 2017). That is because self-reports may suffer from socially desirable responding (e.g., towards music favored by one's peer group; cf. Tarrant et al., 2000) or biased memory recollection (Baumeister et al., 2007), while affective reactions to artificially manipulated or unreleased music excerpts may not reflect preferences displayed on the natural music market.

Only recently, Anderson et al. (2021) overcame this limitation by investigating music-listening behavior exhibited on the streaming service Spotify. They predicted personality in a machine-learning framework and achieved moderate to high performances for the Big Five dimensions, whose predicted and self-reported scores correlated at a range between .26 for Agreeableness and .37 for Emotional Stability. While these findings deviate from those of self-report-based studies with regard to the strength and rank order of associations, these discrepancies cannot be directly attributed to the ecologically valid assessment. That is because Anderson et al. (2021) included not only behavioral music preferences as personality predictors but also streaming behaviors (e.g., the streaming device or the number of artists followed) and participants' demographics (i.e., age and gender). In particular, the demographic predictors, which are known to correlate with personality (see Soto et al., 2011) and improve personality predictions from music preferences (Nave et al., 2018), were among the most predictive variables across all Big Five dimensions except Openness (Anderson et al., 2021). Thus, the current state of literature does not allow for unambiguous conclusions about the personality-relevant information contained in natural music-listening behavior.

3.2.2 Audio vs. Lyrics Characteristics

While previous studies reported important insights into personality correlates in broad musical style or genre preferences, they rarely investigated music preferences on a more granular level, preventing inferences about the intrinsic musical properties underlying these personality associations (Aucouturier & Pachet, 2003; Rentfrow et al., 2011). Non-instrumental songs, in particular, are defined by audio and lyrics characteristics, which may play a distinct role in music preferences and their association with personality. While empirical findings suggest that melodies and lyrics are independently processed when listening to music (Besson et al., 1998; Bonnel et al., 2001) and that both components have a unique impact on the affective listening experience (Ali & Peynircioğlu, 2006; Anderson et al., 2003), they were never compared in a comprehensive analysis in personality psychology.

However, few studies have separately related personality traits to preferences for either audio or lyrics characteristics. For audio characteristics, they found that Openness was correlated to preferences for music with a slow tempo, minor mode, acoustic sounds, and negative valence, while Extraversion was related to preferences for music with major mode, high tones, and positive valence (Dobrota & Reić Ercegovac, 2017; Flannery & Woolhouse, 2021; Fricke & Herzberg, 2017; Vuoskoski & Eerola, 2011). Regarding lyrics characteristics, a pioneering study by Qiu et al. (2019) connected the Big Five personality traits with linguistic style preferences in lyrics, reporting the strongest associations for Conscientiousness, which, for example, correlated positively with a preference for achievement words, and for Emotional Stability, which was related to a preference for positive emotion words in lyrics. These preliminary findings indicate that different aspects of music preferences may be of incremental value for personality prediction.

The sparsity of studies investigating intrinsic musical attributes may be ascribed to a lack of automated extraction tools as researchers had to rely on human labeling to quantify audio and lyrics characteristics (e.g., Dobrota & Reić Ercegovac, 2017; Rentfrow et al., 2011). This approach was not only burdensome and practically infeasible for large collections of songs in natural music-listening records but also at risk of assessing music's subjective experience rather than intrinsic musical properties. However, advances in music information retrieval now enable the automatic extraction of musical characteristics from audio recordings or song lyrics. In particular, technical audio characteristics, ranging from basic physical parameters (e.g., tempo) to more complex aggregated features (e.g., valence) learned via machine-learning algorithms, can now be obtained in a ready-to-use format from external sources such as Spotify (Anderson et al., 2021; Stachl et al., 2020) or via music analysis software (e.g., ESSENTIA; Fricke et al., 2018). To obtain the textual lyrics characteristics, researchers can apply NLP, choosing between closed-vocabulary approaches, which count word usage in a text over pre-defined word categories (see Qiu et al., 2019), and open-vocabulary approaches, which analyze language in a bottom-up manner (e.g., by word clusters). While the closed-vocabulary approaches are often easier to interpret, they are restricted by the word coverage and subjectivity of the underlying dictionaries, which may be why open approaches have proven to be more informative of personality when investigating other sources of written text (e.g., Park et al., 2015; Schwartz et al., 2013). These automated approaches for extracting various musical characteristics open up new possibilities for comparing the contribution of melodies and lyrics when predicting personality from music preferences.

3.2.3 The Present Study

The present study applied a personality computing approach to efficiently collect, computationally represent, and jointly model different aspects of music-listening behavior. For the ecologically valid assessment of music listening, we used smartphones which are currently the most used device for music listening besides radios (IFPI, 2019) and provide granular digital listening records. We analyzed a smartphone-sensing dataset of 330 participants collected over 3 to 85 study days and represented music preferences in terms of intrinsic musical attributes of the songs listened to. Here, we distinguished between preferences quantified via technical audio characteristics from Spotify.com and textual characteristics variables obtained through different natural language models. In addition, we considered habitual listening behaviors that quantified participants' engagement with music (e.g., their listening duration). An extensive set of 844 strictly behavioral variables served us to predict self-reported Big Five personality trait scores on domain and facet level. To counteract overfitting, we applied two machine-learning algorithms suitable for high-dimensional data (i.e., data in which the number of predictors is larger than the number of observations) and evaluated prediction performance in a strict out-of-sample fashion. Finally, we used interpretable machine-learning techniques to compare the independent contribution of audio- and lyrics-based preferences and explored which single music-listening variables were most important in personality predictions.

3.3 Method

We conducted a secondary data analysis based on three smartphone-sensing datasets summarized in Stachl et al., 2020. Since the datasets were previously published, we focus our report here on procedures and decisions relevant to the present study. Additional details on the study procedures are available in the original articles (Schoedel et al., 2019; Schuwerk et al., 2019; Stachl et al., 2017).

This study's design and analyses are purely exploratory and were not pre-registered. However, preliminary (and also exploratory) groundwork provided in a student thesis was preregistered under https://osf.io/as3ze. While this preregistration does not directly pertain to the current study, we still communicate deviations in our Disclosure of Prior Data Uses available in our project's OSF repository under https://osf.io/x7dar/. In this repository, we also provide the code for preprocessing, variable extraction, and predictive modeling, as well as a dataset of aggregated variables used for predictive modeling. However, please understand that the privacy-sensitive nature of the smartphone usage data prevents us from sharing the raw logging data.

3.3.1 Dataset

In the present study, we re-analyzed data from three separate studies conducted within the PhoneStudy project at LMU Munich between 2014 and 2018 (Schoedel et al., 2019; Schuwerk et al., 2019; Stachl et al., 2017). In Table 3.A1 of the Appendix, we provide an overview of the included datasets. The procedures of all three studies were approved by institutional review boards and carried out according to EU laws and ethical standards. All subjects participated willingly and gave informed consent prior to their participation. In all three studies, participants completed a series of self-report questionnaires, including the personality inventory used here. Furthermore, they installed an Android research app on their private smartphones, which logged a variety of smartphone usage behaviors, including music listening, for a period of at least 14 study days. A detailed description of the individual study procedures and all collected measures is available in the respective research articles and in Stachl et al. (2020).

The initial sample was determined by the availability of secondary data and contained logging and self-report data from 684 participants. During pre-processing, we removed participants who had played fewer than five different songs with available lyrics characteristics (see our section on *Song-Level Variables*), resulting in a sample size of 330 participants (54% women) with sufficient music-listening data. We additionally assessed the response validity of our self-report measure but refrained from removing participants based on inconclusive evidence of careless responding (see the Appendix; Curran, 2016; Ward & Meade, 2023). Our final sample was skewed towards younger age (M = 22.42, SD = 4.33, Min = 18, Max = 57) and better education (93% with A-levels and 20% with a university degree).

3.3.2 Personality Measure

All three studies used the German Big Five Structure Inventory (BFSI; Arendasy, 2009) to assess personality based on the well-established Big Five taxonomy: Openness, Conscientiousness, Extraversion, Agreeableness, and Emotional Stability (McCrae & John, 1992). The BFSI consists of 300 items (adjectives and short phrases) and measures the Big Five personality dimensions on five broad domains and 30 more specific facets. Item agreement is stated on a 4-point Likert scale ranging from "untypical for me" to "rather untypical for me" to "rather typical for me" to "typical for me." The BFSI corresponds to the partial credit model (Masters, 1982), which defines an individual's observed item response as a function of their latent trait value (i.e., their person parameter) and the item's latent difficulty thresholds. Correspondingly, we used the person parameters assigned to participants based on their item sum scores as personality estimates in our analyses. Confidence intervals of internal consistencies obtained in our sample are available in Table 3.A2 in the Appendix.

3.3.3 Behavioral Music-Listening Measures

An Android-based research app provided raw sensing data on participants' natural smartphone usage, including their music-listening records. Whenever participants had listened to locally stored or streamed music, the app created time-stamped event logs with the title, artist, and album name of the played song.

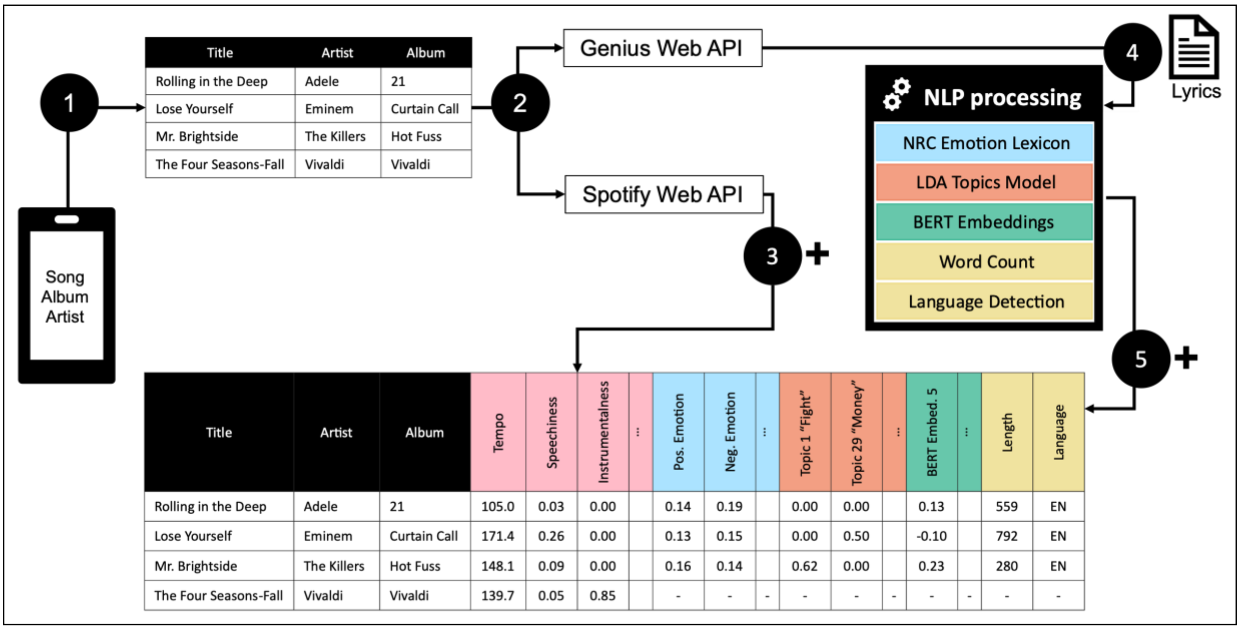
3.3.3.1 Song-Level Variables

To describe the played songs in terms of musical attributes, we enriched the music event logs with audio and lyrics characteristics. Therefore, we retrieved additional song-level data from two external sources. We visualize the data enrichment workflow with exemplary songs in Figure 3.1 and provide further details in the Appendix.

First, we used Spotify's Track Application Programming Interface (API) to retrieve 12 song-level variables provided by Spotify.com (see Table 3.1; Spotify, 2022). These variables contained 11 computationally derived technical audio characteristics (e.g., the songs' "tempo" and "acousticness") based on the songs' audio recordings and one lyrics-based variable indicating the presence of explicit lyrical contents (i.e., strong language or references to sexual or violent behavior).

Figure 3.1

Workflow for Enriching Smartphone-Sensed Music-Listening Records



Note. Sensed music-listening data were enriched with different song-level information. The exemplary songs in the tables demonstrate the face validity of the different audio and lyrics characteristics. Details on the APIs can be found on the respective websites Spotify.com and Genius.com. The enriched musical attributes are defined in Table 3.1. API = Application Programming Interface; NLP = Natural Language Processing.

In addition, we retrieved song lyrics from Genius.com and created meaningful textual variables via a text-mining pipeline combining closed and open vocabulary approaches (see Table 3.1; Genius, 2022). We describe all lyrics analyses at an abstract level here and provide further details in the Appendix. We extracted two stylistic variables representing the lyrics' length and language and applied three natural language models to quantify the content characteristics of the lyrics. First, we detected the emotional content of the lyrics using the Word-Emotion Association Lexicon of the National Research Council (NRC; Mohammad & Turney, 2013). Based on word occurrences, the NRC lexicon assigned each song a score on two sentiments (positive and negative valence) and eight emotion categories (anger, anticipation, disgust, fear, joy, sadness, surprise, and trust). Second, we applied Latent Dirichlet Allocation (LDA; Blei et al., 2003) to obtain the topics covered in the song lyrics. This generative probabilistic model assumes that each document (in our case, song lyrics) in a corpus contains a mixture of latent topics, where each topic is a cluster of co-occurring words. To avoid overfitting the LDA to sample-specific patterns in our lyrics corpus, we pre-trained the model on a large external lyrics corpus (the Million Song Dataset; Bertin-Mahieux et al., 2011). We determined the topic count such that the topic coherence (i.e., the semantic similarity between words within a topic; Chang et al., 2009) was maximized, which resulted in a model with 30

Table 3.1

Description of Song-Level Musical Characteristics

Song-level Variable	Data Source	Description
Audio Characteristics		
Mode	Spotify API	The song's modality (major vs. minor), i.e., the type of scale the song's melodic content is derived from.
Key	Spotify API	The song's melodic key in standard pitch class notation (e.g., 0 = C, 1 = C/D, 2 = D).
Tempo	Spotify API	The song's overall estimated tempo in beats per minute.
Loudness	Spotify API	The song's average loudness in decibels.
Energy	Spotify API	The song's perceived intensity and activity on a scale from 0.0 to 1.0, whereby songs with a high energy feel fast, loud, and noisy. The measure is defined by several elements such as perceived loudness.
Danceability	Spotify API	The song's suitability for dancing on a scale from 0.0 to 1.0, whereby higher values represent more danceable songs. The measure combines several musical elements including tempo and overall regularity.
Acousticness	Spotify API	The song's acousticness (i.e., absence of electronic sounds) on a scale from 0.0 to 1.0, whereby higher values represent an increased confidence that the song is acoustic.
Valence	Spotify API	The musical positiveness conveyed by the song, whereby songs with valence closer to 1.0 sound more positive (e.g., happy) and songs with values closer to 0.0 sound more negative (e.g., sad).
Speechiness	Spotify API	The probability of spoken words in a song, whereby values below 0.33 most likely represent pure music, while higher values represent songs containing both music and speech (e.g., rap music).
Instrumentalness	Spotify API	The probability of vocals in a song, whereby values closer to 1.0 represent a greater likelihood that the song contains no vocal content. Values above 0.5 most likely represent instrumental songs.
Liveness	Spotify API	The probability of an audience in the song's recording, whereby values closer to 1.0 represent a greater likelihood that the song was performed live. Values above 0.8 most likely represent live songs.
Lyrics Characteristics		
Length	Genius API + lyrics NLP	The number of words of the song's lyrics.
Language	Genius API + lyrics NLP	The song's language (i.e., English vs. German vs. Other) derived via language detection from the lyrics.
Explicit content	Spotify API	The presence of explicit words (e.g., swear words) in the song's lyrics.
10 Emotionality scores	Genius API + lyrics NLP	The probability by which the song's lyrics contain words from ten emotion categories of the NRC Emotion Lexicon (e.g., Positivity, Negativity, Sadness, Anger, Joy, Trust).
30 Topics	Genius API + lyrics NLP	The probability by which the song's lyrics belong to each of 30 lyrical topics derived via Latent Dirichlet Allocation.
768 Word embeddings	Genius API + lyrics NLP	The song's value on each of the 768 dimensions in the lyrics embedding space of the BERT-model.

Note. Spotify variable descriptions were derived from Spotify.com. API = Application Program Interface; API calls retrieved ready-to-use variables from Spotify.com and raw song lyrics from Genius.com. NLP = natural language processing; NLP extracted variables from the song lyrics.

topics. This pre-trained topic model assigned each song in our corpus to a score on each of the 30 topics. We provide details on the topic modeling, including coherence metrics (see Table 3.A3) and topic keywords (see Table 3.A4), in the Appendix. Finally, we represented the lyrics as word embeddings using the state-of-the-art Bidirectional Encoder Representations from Transformers (BERT; Devlin et al., 2018). BERT embeddings use a neural network architecture to convert textual data into context-sensitive numerical representations. We employed the pre-trained BERT implementation from the HuggingFace framework (Wolf et al., 2020) to extract one embedding vector for each song's full lyrics. This BERT vector had a length of 768, so each song was assigned a score on 768 embedding dimensions. Again, more details on the BERT modeling are available in the Appendix.

In total, we computed 822 variables quantifying different intrinsic musical characteristics of the songs played in our study (see Table 3.1). These song-level variables were assigned to the respective music events in the logging data. Figure 3.1 illustrates this matching and provides examples of the face validity of song-level variables. However, not all music-listening events could be enriched because some contained non-musical tracks (e.g., audiobooks), had incorrect song information (e.g., typos in the song title), or were not covered by the respective online sources.

3.3.3.2 Person-Level Variables

In the next preprocessing step, we used the song-level enriched music event logs to extract person-level variables capturing music preferences and habitual listening behaviors. Therefore, we first reduced the logs to music events that lasted longer than 20 seconds to exclude skipped songs. Furthermore, we removed music events from the first study day to avoid potential reactivity biases.

We aggregated the distribution of the song-level variables (see Table 3.1) over each participant's played songs via the arithmetic mean (for numeric variables) or percentage scores (for factor variables). We focused on participants' average music preferences to limit our predictor space while also enabling a comparison with past research (e.g., Rentfrow & Gosling, 2003; Nave et al., 2018; Schäfer & Mehlhorn, 2017). The resulting 833 variables covered average music preferences for 1) audio characteristics (e.g., the mean tempo of played songs) and 2) lyrics characteristics represented by a) emotion scores (e.g., the mean negative emotionality of played songs), b) topics (e.g., the mean probability of the topic "love" in played songs), c) word embeddings (e.g., the mean embedding dimension 1 of played songs), and d) other lyrics characteristics (e.g., the percentage of English songs among all played songs). As noted above, the external song-level variables were not available for all tracks, so the music preference variables only covered a portion of participants' played tracks. On average, participants' preferences for Spotify-based variables covered 57% (*SD* = 0.21) of played tracks, while preferences for lyrics-based variables covered 42% (*SD* = 0.19). To account for the limited song coverage, we created an additional validity variable indicating the proportion of participants' songs represented by lyrics-based preference variables.

In addition, we extracted ten variables on habitual listening behaviors by quantifying the extent of participants' music consumption, for example, the total number of played songs, the number of unique artists listened to, or the average daily number of played songs.

In total, we obtained 844 variables capturing participants' music preferences and habitual listening behaviors, which served as predictors in our personality predictions. We provide a list of all person-level variables, including summary statistics, in our repository.

3.3.4 Personality Predictions

3.3.4.1 Machine-Learning Analyses

We trained machine-learning models for the prediction of the five domains and 30 facets of our personality inventory. While we provide a short overview of basic machine-learning concepts relevant to understanding our study here, a more detailed introduction to supervised machine learning can be found in a state-of-the-art tutorial by Pargent et al. (2023).

Models. For each personality outcome, as a benchmark, we compared the predictive performance of elastic net (Zou & Hastie, 2005) and random forest models (Breiman, 2001) with those of a featureless baseline model. The baseline model predicted the mean personality score of a training set for all observations in a respective test set. The elastic net model is an extension of basic linear regression that applies two regularization penalties to encourage simpler models, and the random forest aggregates the output of multiple decision trees to account for non-linear relationships. We chose these models because of their ability to automatically perform a selection of relevant predictors, allowing them to cope with high-dimensional and inter-correlated predictor spaces in small samples. We used the default settings of the models' hyperparameters as specified in their implementation within the mlr3 environment (e.g., Lang et al., 2019).

Resampling Strategy. For a strict separation of training and test data, we estimated the models' expected predictive performance on unseen data using 10-times repeated 10-fold cross-validation (10x10 CV). In this cross-validation scheme, a dataset is randomly split into 10 folds, and each fold serves as an unseen hold-out set for prediction (i.e., the test set) once, while the models are trained on the data of the remaining nine folds (i.e., the training set). Prediction performance is computed separately for each fold of 10x10 CV and then aggregated to the mean across the 100 iterations per model. Such out-of-sample prediction performances have a reduced risk of overfitting sample-specific patterns and provide a more reliable estimate of the models' ability to make predictions in new samples (e.g., Yarkoni & Westfall, 2017).

Performance Evaluation. We evaluated model performances by correlating predicted personality scores with the person-parameter estimates from the self-reported personality trait measure using Spearman rank order correlations (r_s). However, the baseline model produced invariant predictions (i.e., the training set's mean) across all observations, preventing us from calculating this correlation metric. Hence, we additionally determined the mean squared error (MSE) for all models (see the Appendix for the respective formulas). We tested if the MSEs of our prediction models were significantly lower than those of the corresponding featureless baselines. We treated the MSEs of prediction vs. baseline models obtained in the same cross-validation iteration as dependent pairs (due to their shared training set) and compared them across iterations using variance-corrected pairwise student t-tests (one-sided; Bouckaert & Frank, 2004; Nadeau & Bengio, 1999; Stachl et al., 2020). For each personality outcome, we adjusted for multiple comparisons ($n = 2$ models against the common baseline) via Bonferroni correction. Based on this conservative approach, prediction models with a significantly smaller MSE than the baseline were considered predictive as they were consistently successful across resampling iterations.

3.3.4.1 Interpretable Machine Learning

Machine-learning models often lack natural interpretability, so we combined different approaches to gain insights into successful prediction models. First, we grouped our variables by the overarching aspects of music listening (e.g., audio vs. lyrics preferences) they represented and investigated the unique importance of these groups as a whole. We used the settings described above (10x10 CV) and ran additional benchmark analyses with seven different subsets of music-listening variables. More specifically, we compared the independent predictive performance of 1) habitual listening behaviors, 2) preferences for audio characteristics, and 3) preferences for lyrics characteristics, whereby the third group was considered both in aggregation and separately by the types of lyrics information, namely lyrics' a) emotionality, b) topics, c) word embeddings, and d) other lyrics characteristics. As an effect size index of the groups' importance, we considered their individual performance in terms of the Spearman correlation (r_s) metric and computed variance-corrected 95% confidence intervals based on the student t-distribution (Bouckaert & Frank, 2004; Nadeau & Bengio, 1999).

However, we refrained from conducting significance tests for between-group comparisons due to the highly exploratory nature of these analyses.

For insights into the importance of single predictors within the full set of music-listening variables, we applied interpretable machine-learning tools to the full personality prediction models. For random forest models, we computed permutation variable importance, which measures the decrease in a model's prediction performance after randomly permuting one single variable (Casalicchio et al., 2019). Variable importance scores were aggregated across 50 iterations to provide stable estimates. For elastic net models, we considered the model-inherent, non-standardized beta weights known from simple linear regression.

To further explore predictor effects, we extracted the 15 most important variables of the respective models and illustrated their influence on the prediction with accumulated local effects (ALE; Apley & Zhu, 2020). ALE plots visualize the effect of an individual predictor variable by showing how its manifestations, on average, affect the model prediction.

3.3.5 Statistical Software

API calls and NLP analyses were conducted in Python, version 3.7.10 (Python Software Foundation, 2021). We used the libraries MALLET (McCallum, 2002) and gensim (Rehurek & Sojka, 2010) for Latent Dirichlet Allocation, the library NRXLex (Bailey, 2019) for emotion analysis, and the Hugging Face Transformers (Wolf et al., 2020) for extracting BERT embeddings.

All other analyses were conducted with the statistical software R (version 4.0.3 for preprocessing and version 4.2.1 for data analysis; R Core Team, 2022). We used the packages dplyr (version 1.0.7, Wickham et al., 2021) and fxtract (version 0.9.4, Au, 2020) for extracting person-level variables. For predictive modeling, we employed the packages mlr3 (version 0.14.1, Lang et al., 2019), glmnet (version 4.1-6, Friedman et al., 2010), and ranger (version 0.14.1, Wright & Ziegler, 2017). Furthermore, we used iml (version 0.11.1, Molnar et al., 2018) for interpretable machine learning. Finally, the packages ggplot2 (version 3.3.5, Wickham, 2016) and ggwordcloud (version 0.5.0, Le Pennec & Slowikowski, 2019) served for visualizing our results.

3.4 Results

3.4.1 Descriptive Statistics

Across our sample, participants provided between 3 and 85 days of logged smartphone data (M = 43.4, SD = 15.8) and, on average, listened to music on half of these days (M = 47.1%, *SD* = 28.5%). They used an average of 2.3 different music apps (*SD* = 1.4), with Spotify being the most used app (40.6%), followed by Android Music (19.7%) and Google Play Music (9.1%). The number of songs listened to per participant ranged between 5 and 4387 (*M* = 397.6, *SD* = 547.2), and, on average, participants played 9.4 songs per day (*SD* = 12.7) for 31.4 minutes (*SD* = 42.9). Participants' self-reports are summarized in the Appendix (see Table 3.A2). Furthermore, we provide detailed descriptive statistics for behavioral variables, including pairwise Spearman correlations with self-reports, in our OSF project repository.

3.4.2 Personality Predictions

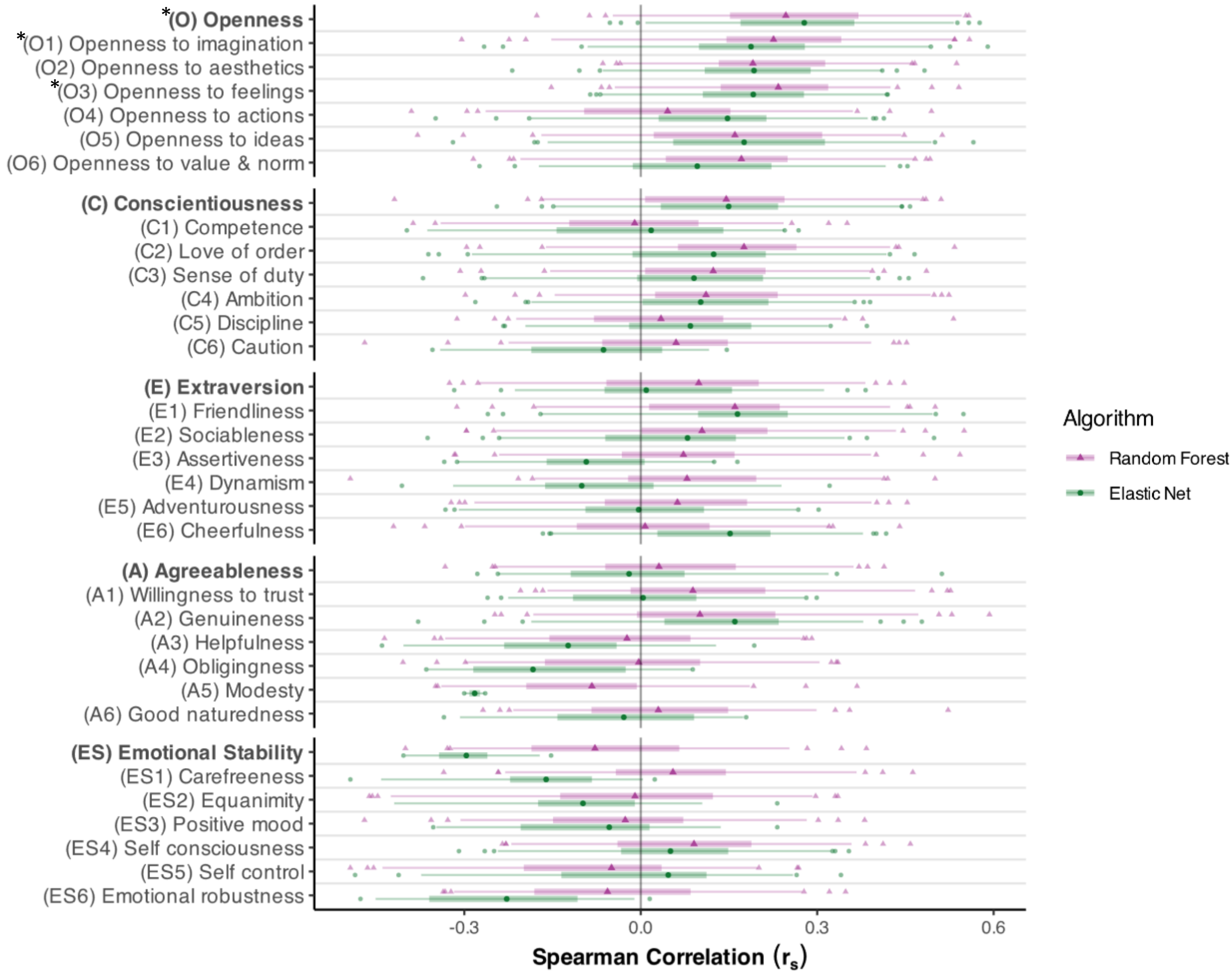
In our main benchmark analysis, we evaluated the performance of two machine-learning algorithms predicting personality from our full spectrum of music-listening variables. In this analysis, the linear elastic net and non-linear random forest models obtained similar prediction performances for most Big Five dimensions (see Figure 3.2). However, the elastic net produced only one instead of three significant models (see Table 3.2) and failed to make variable-based predictions in many of the 100 resampling iterations of the 10x10 CV scheme for several personality dimensions in Figure 3.2 (e.g., the facet Modesty of Agreeableness)[1]. Hence, we focus our reports on the random forest models in the remainder of this article.

The results summarized in Table 3.2 show that the Big Five personality dimension Openness (O) and its facets Openness to imagination (O1) and Openness to feelings (O3) were successfully predicted from our music-listening variables. That means the MSEs of their random forest models across resampling iterations were, on average, significantly lower than those of the featureless baseline model. While the remaining Big Five criteria exhibited no significant reduction in MSEs, the distribution of correlations between predicted and self-reported personality scores in Figure 3.2 reveals promising prediction performances in many resampling iterations for several other personality dimensions. More specifically, 14 outcomes in Table 3.2 exhibited a small- to medium-sized mean correlation on or above a threshold of .10 suggested by Cohen's (1992) effect size conventions (r_s between .10 and .25). Inspection of these selected outcomes suggests that our random forest models worked best for the domain Openness (O, r_s = .25) and its facets Openness to imagination (O1, r_s = .23), followed by Openness to feelings (O3, r_s = .22), Openness to aesthetics (O2, r_s = .21), Openness to ideas (O5, r_s = .15), and Openness to value and norm (O6, r_s = .15). Second best prediction

[1]If predictors contain no relevant information for predicting an outcome, the elastic net shrinks their coefficients to zero and returns intercept-only predictions (i.e., it constantly predicts the training data mean), which are mathematically equivalent to our baseline predictions. The intercept-only predictions produce NAs for the Spearman correlation metric (due to their invariance). Thus, outcomes that produced many intercept-only predictions exhibited low variance in the Spearman correlation metric across iterations in Figure 3.2.

Figure 3.2

Box and Whisker Plots of Prediction Performance From Repeated Cross-Validation for Each Personality Dimension and Algorithm



Note. Prediction performance over the 100 resampling iterations of the cross-validation scheme (10x10 CV). Performance is measured via the Spearman rank correlation between predicted and measured personality scores. The middle symbol represents the median, boxes include values between the 25 and 75% quantiles, and whiskers extend to the 2.5 and 97.5% quantiles. Outliers are depicted by single points. The grey line indicates a correlation of 0.0 between the predicted and self-reported personality scores. Asterisks indicate significantly predictive models.

performances were obtained for the domain Conscientiousness (C, r_s = .13) and its facets Love of order (C2, r_s = .15), followed by Ambition (C4, r_s = .13) and Sense of duty (C3, r_s = .11). In contrast, the remaining facets of Openness and Conscientiousness exhibited correlations close to zero. While the domains Extraversion (E) and Agreeableness (A) obtained correlations below .10, each two of their six facets showed moderate prediction performances, namely Friendliness (E1, r_s = .13) and Sociableness (E2, r_s = .10) as well as Willingness to trust (A1, r_s = .11) and Genuineness (A2, r_s = .11). Only the dimensions of Emotional Stability were completely unrelated to music-listening behavior according to our performance metrics. Please note that all models with moderate prediction performance, including those reaching significance, also

Table 3.2

Mean Prediction Performance per Personality Dimension and Algorithm

	Random Forest			Elastic Net			Baseline
Personality Dimension	r_s	MSE	p_{adj}	r_s	MSE	p_{adj}	MSE
***(O) Openness**	.25	0.50	.041	.27	0.49	.012	0.53
*(O1) Openness to imagination	.23	1.91	.032	.19	1.98	.400	2.04
(O2) Openness to aesthetics	.21	1.64	.151	.18	1.66	.371	1.71
*(O3) Openness to feelings	.22	4.39	.027	.19	4.53	.237	4.64
(O4) Openness to actions	.03	2.23	1	.12	2.16	.726	2.18
(O5) Openness to ideas	.15	2.16	.378	.18	2.13	.166	2.22
(O6) Openness to value & norm	.15	1.09	1	.10	1.09	1	1.08
(C) Conscientiousness	.13	0.55	.987	.14	0.54	.527	0.55
(C1) Competence	-.02	1.45	1	-.01	1.40	1	1.39
(C2) Love of order	.15	2.37	.686	.11	2.39	.998	2.39
(C3) Sense of duty	.11	1.91	1	.09	1.91	1	1.91
(C4) Ambition	.13	2.87	1	.10	2.85	1	2.84
(C5) Discipline	.04	2.23	1	.08	2.17	1	2.17
(C6) Caution	.05	1.92	1	-.09	1.89	1	1.87
(E) Extraversion	.07	0.53	1	.03	0.54	1	0.53
(E1) Friendliness	.13	1.55	.540	.17	1.55	.568	1.57
(E2) Sociableness	.10	2.97	.444	.06	3.03	1	3.03
(E3) Assertiveness	.07	1.90	1	-.08	1.90	1	1.88
(E4) Dynamism	.08	2.56	1	-.07	2.57	1	2.54
(E5) Adventurousness	.06	2.29	1	0	2.28	1	2.26
(E6) Cheerfulness	.01	2.82	1	.13	2.71	.612	2.74
(A) Agreeableness	.04	0.63	1	0	0.63	1	0.62
(A1) Willingness to trust	.11	2.14	.851	0	2.18	1	2.15
(A2) Genuineness	.11	1.01	.898	.13	1	.386	1.01
(A3) Helpfulness	-.03	1.98	1	-.14	1.92	1	1.91
(A4) Obligingness	-.02	1.96	1	-.15	1.94	1	1.92
(A5) Modesty	-.09	1.38	1	-.28	1.31	1	1.31
(A6) Good naturedness	.03	3.52	1	-.03	3.51	1	3.47
(ES) Emotional Stability	-.06	0.55	1	-.30	0.53	1	0.52
(ES1) Carefreeness	.05	1.80	1	-.17	1.78	1	1.77
(ES2) Equanimity	-.01	1.20	1	-.11	1.16	1	1.16
(ES3) Positive mood	-.03	2.24	1	-.08	2.18	1	2.15

Table 3.2 (continued)

	Random Forest			Elastic Net			Baseline
Personality Dimension	r_s	MSE	p_{adj}	r_s	MSE	p_{adj}	MSE
(ES4) Self consciousness	.08	1.40	.905	.05	1.42	1	1.40
(ES5) Self control	-.08	1.03	1	0	0.99	1	0.99
(ES6) Emotional robustness	-.05	1.44	1	-.24	1.40	1	1.39

Note. Performance metrics were first computed separately for each of the 100 iterations of our cross-validation scheme (10x10 CV) and then aggregated to the mean. r_s = Spearman's rank order correlation between predicted and measured personality scores. MSE = Mean squared error. p_{adj} = Bonferroni adjusted p-values of variance corrected one-sided t-tests comparing the MSE measures of prediction models with the baseline. Overarching personality domains are printed in bold font. Significant models (α = .05) are indicated by an asterisk.

contained few resampling iterations with a negative correlation between predicted and self-reported outcomes in Figure 3.2, indicating that the random forests failed to learn systematic patterns in some instances.

3.4.3 Interpretation of Prediction Models

After providing an overview of how well different personality dimensions can be predicted from music-listening variables, we considered what aspects of music listening drove our models' predictions. We applied two interpretation approaches to all random forest models with a minimum mean performance of r_s = .10 listed above.

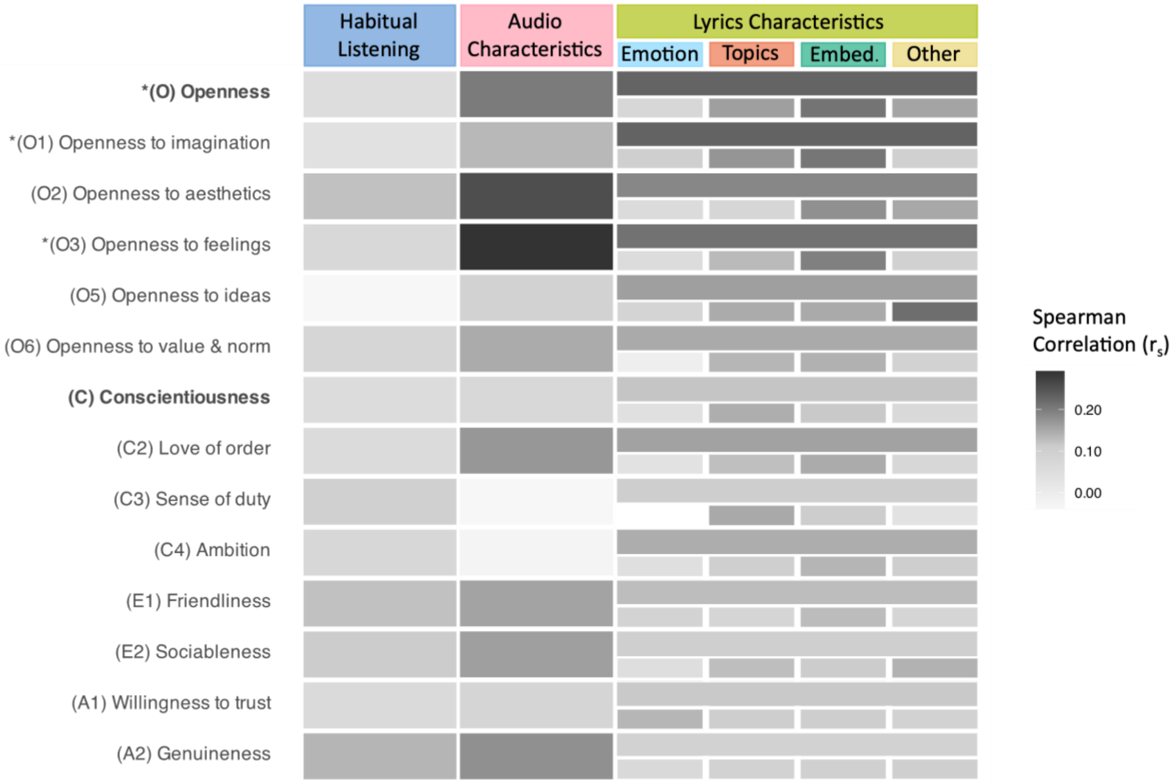
3.4.3.1 Importance of Variable Groups

We conducted an additional benchmark analysis comparing the independent performance of each group of music-listening variables when separately predicting the respective personality scores. We report prediction performances in terms of the average Spearman correlation with 95% confidence intervals across iterations in Table 3.A5 of the Appendix and illustrate them in Figure 3.3. The unique prediction performance represents the relevance of each variable group as a whole (i.e., including all of its variables and their interactions) for our random forest models.

Figure 3.3 shows that, across all personality outcomes, habitual listening behaviors (range r_s = -.04 to .14) were less predictive than music preferences (range r_s = -.04 to .29). In contrast, preferences for audio and lyrics characteristics were relevant for many outcomes. Audio characteristics obtained the highest prediction performances for Openness dimensions (range r_s = .09 to .29) and lowest ones for Conscientiousness facets (e.g., range r_s = -.04 to .17).

Figure 3.3

Heatmap of Prediction Performance by Variable Group for Illustration of Grouped Variable Importance



Note. Prediction performance when using each group of music-listening variables (see columns) separately for predicting personality outcomes (see rows). One benchmark comparing the seven variable groups was conducted for each personality outcome predicted with a minimum performance $r_s \geq .10$ by the full variable set (see Table 3.2). The average Spearman rank correlation (r_s) between predicted and measured personality scores across resampling iterations serves as an indicator of grouped variable importance, whereby higher values indicate greater relevance of the respective variable group. The higher-level group "Lyrics Characteristics" comprised the four lower-level groups Emotionality, Topics, Word Embeddings, and Other Lyrics Characteristics (see Table 3.1).

Lyrics characteristics were also particularly informative about Openness dimensions (range r_s = .15 to .23) but least relevant for facets of Extraversion (range r_s = .10 to .13) and Agreeableness (range r_s = .09 to .12). Among lyrics characteristics, word embeddings were the most relevant group for the largest number of outcomes (8), followed by topics (2), other lyrics characteristics (2), and emotionality (1) – not regarding ties between groups. One may argue that the superiority of word embeddings is related to the large size of this predictor group (i.e., 768 lyrics word embeddings vs. 30 lyrics topics in the second largest group). However, as seen in Figure 3.3, other aspects of lyrics (e.g., topics for Conscientiousness) and also audio characteristics (e.g., for the facet Openness to aesthetics) outweighed word embeddings for several outcomes, indicating that the number of variables per group does not determine its prediction performance.

Looking further into the relevance of music preferences, we can compare their importance for all four personality domains featured in Figure 3.3. For several Openness dimensions, preferences for audio (range r_s = .09 to .29) and lyrics characteristics (range r_s = .15 to .23) were both informative for predictions with differential patterns per dimension. In particular, lyrics characteristics were more relevant for the domain itself (O, r_s = .20 for audio vs. r_s = .23 for lyrics) and its facets Openness to imagination (O1, r_s = .13 for audio vs. r_s = .23 for lyrics) and Openness to ideas (O5, r_s = .09 for audio vs. r_s = .16 for lyrics), while audio characteristics were more important for the facets Openness to aesthetics (O2, r_s = .26 for audio vs. r_s = .19 for lyrics) and Openness to feelings (O3, r_s = .29 for audio vs. r_s = .21 for lyrics). For Openness to value and norm (O6), audio and lyrics preferences were equally predictive (both r_s = .15). Taking a closer look at the different types of lyrics information, word embeddings were most relevant for Openness predictions (range r_s = .15 to .23), followed by topics (range r_s = .07 to .19), other lyrics characteristics (range r_s =.09 to .24) and emotionality (range r_s = -.03 to .10). Only for the facet Openness to ideas, other lyrics characteristics produced best predictions (r_s = .24). For the domain Conscientiousness (C, r_s = .07 for audio vs. r_s = .12 for lyrics) and its facets Sense of duty (C3, r_s = -.04 for audio vs. r_s = .11 lyrics) and Ambition (C4, r_s = -.03 for audio vs. r_s = .15 for lyrics), lyrics characteristics were more informative for prediction models because audio characteristics were (almost) unpredictive. Only for the facet Love of order (C2, r_s = .17 for audio vs. r_s = .16 for lyrics), audio and lyrics characteristics were similarly important. More specifically, lyrics' topics (range r_s = .10 to .16) and word embeddings (range r_s = .11 to .16) were particularly meaningful, while emotionality (range r_s = -.10 to .03) and other lyrics characteristics (range r_s = .02 to .11) were not very predictive. For the Extraversion facets Friendliness (E1, r_s = .16 for audio vs. r_s = .13 for lyrics) and Sociableness (E2, r_s = .16 for audio vs. r_s = .10 for lyrics), audio characteristics were more relevant for predictions compared to lyrics, whose different types of variables were similarly predictive. Finally, lyrics preferences were slightly more predictive for the Agreeableness facet Willingness to trust (A1, r_s = .08 for audio vs. r_s = .12 for lyrics), while audio preferences were more relevant for the facet Genuineness (A2, r_s = .18 for audio vs. r_s = .09 for lyrics). Here, the different aspects of song lyrics were again of comparable relevance.

As seen in the importance measures above, some variable groups performed better on their own than in combination with the remaining variables (see Table 3.2 and Table 3.A5). For example, Openness to feelings (O3) obtained better performances when predicted only from audio characteristics (r_s = .29) compared to the performance of the full predictor set in Table 3.2 (r_s = .22). Please note, however, that these results were obtained for different benchmarks, and that the full variable set performance in the grouped benchmark are reported in Table 3.A5. Such discrepancies highlight the predictive power of single variable groups for the respective outcome and indicate that some of the other groups introduced noise that hindered random forest models from learning systematic patterns.

3.4.3.2 Importance of Single Variables

We also explored which variables – considered individually among the full set of music-listening variables – were most important for predicting each personality dimension. Therefore, we considered the loss in prediction performance after permuting a single variable of the random forest models. In Table 3.3, we present the top ten variables (i.e., those causing the greatest performance loss) for each outcome with some exemplary variable effects in ALE plots. In addition, we provide full lists of variable importance and beta weights for the elastic net models in the online project repository.

Table 3.3

Top 10 Most Important Music-Listening Variables per Personality Model with Selected Accumulated Local Effect Plots

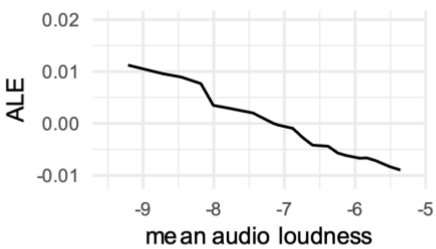
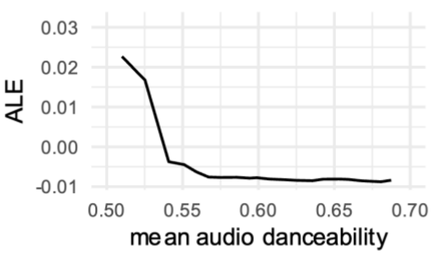
Group	Rank	Top Predictors	r_s	ALE Plot
				***(O) Openness**
L	1	embedding 208	.25	
L	2	embedding 013	.23	
L	3	embedding 092	-.23	
A	4	loudness	-.23	
L	5	embedding 599	-.22	
L	6	embedding 688	-.24	
L	7	embedding 436	-.18	
L	8	embedding 612	.20	
L	9	embedding 315	.22	
L	10	embedding 047	-.21	
				***(O1) Openness to imagination**
L	1	embedding 094	-.23	
L	2	embedding 599	-.22	
L	3	embedding 628	-.18	
A	4	danceability	-.16	
L	5	embedding 767	.16	
L	6	embedding 304	-.15	
L	7	embedding 208	.20	
L	8	embedding 612	.17	
L	9	embedding 315	.21	
L	10	embedding 642	-.17	

Table 3.3 (continued)

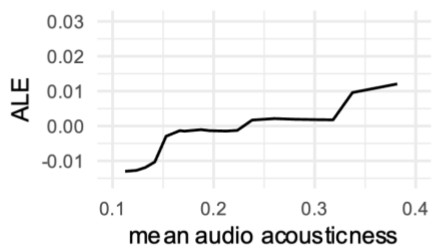
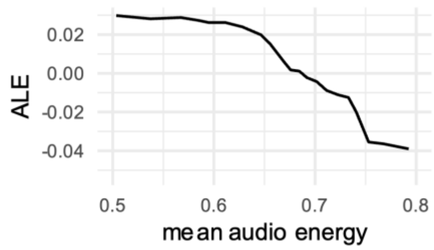
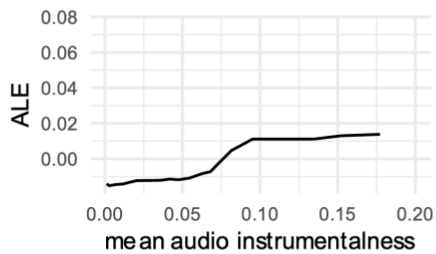
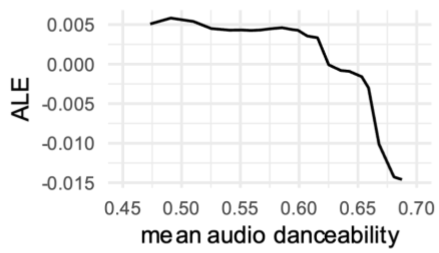
Group	Rank	Top Predictors	r_s	ALE Plot
		(O2) Opennes to aesthetics		
A	1	loudness	-.24	
L	2	embedding 127	.21	
L	3	embedding 144	.22	
L	4	embedding 208	.22	
L	5	embedding 457	.17	
L	6	embedding 047	-.22	
L	7	embedding 690	.15	
L	8	embedding 599	-.19	
L	9	embedding 194	.18	
A	10	acousticness	.17	
		***(O3) Opennes to feelings**		
A	1	loudness	-.23	
L	2	embedding 572	.24	
L	3	embedding 692	-.25	
A	4	acousticness	.27	
A	5	energy	-.27	
L	6	embedding 550	-.13	
L	7	embedding 619	.19	
L	8	embedding 028	.24	
L	9	embedding 189	-.20	
L	10	embedding 077	-.17	
		(O5) Opennes to ideas		
A	1	instrumentalness	.17	
L	2	embedding 639	-.20	
L	3	embedding 434	.20	
L	4	embedding 208	.16	
L	5	embedding 315	.21	
L	6	embedding 703	.25	
L	7	embedding 232	.09	
L	8	embedding 019	-.14	
L	9	embedding 092	-.16	
L	10	embedding 140	-.20	
		(O6) Openness to value and norm		
L	1	embedding 703	.25	
L	2	embedding 599	-.19	
L	3	embedding 085	-.23	
L	4	embedding 173	-.20	
L	5	embedding 051	.23	
L	6	embedding 269	.14	
L	7	embedding 344	-.18	
L	8	embedding 013	.20	
A	9	danceability	-.18	
L	10	embedding 393	.12	

Table 3.3 (continued)

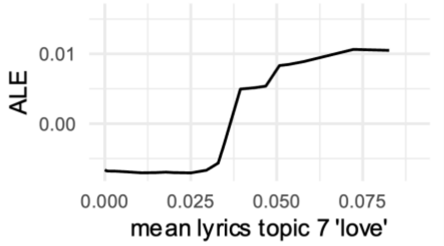
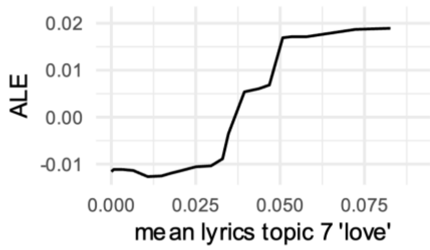
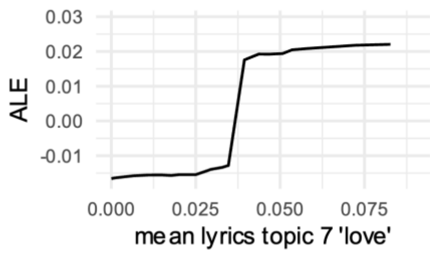
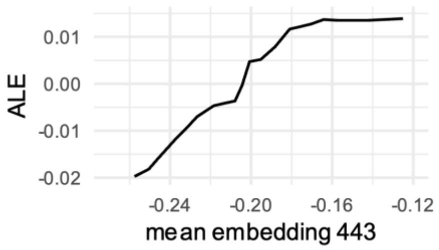
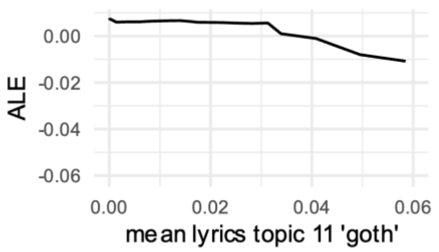
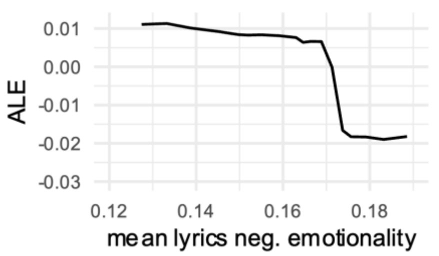
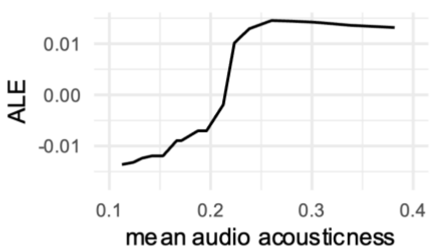
Group	Rank	Top Predictors	r_s	ALE Plot
		(C) Conscientiousness		
L	1	embedding 486	-.23	
L	2	topic 7 “love”	.15	
L	3	embedding 639	-.12	
L	4	embedding 514	.19	
L	5	embedding 208	.11	
L	6	embedding 424	.18	
L	7	embedding 709	-.15	
L	8	embedding 099	.17	
L	9	embedding 420	.19	
L	10	embedding 081	.20	
		(C2) Love of order		
L	1	embedding 486	-.24	
L	2	embedding 001	-.18	
L	3	embedding 038	-.22	
L	4	embedding 547	.14	
L	5	embedding 148	-.18	
L	6	embedding 530	.18	
L	7	embedding 045	-.20	
L	8	embedding 243	.16	
L	9	embedding 478	.18	
L	10	embedding 424	.18	
		(C3) Sense of duty		
L	1	embedding 555	-.21	
L	2	embedding 001	-.19	
L	3	embedding 486	-.22	
L	4	embedding 131	.16	
L	5	embedding 243	.19	
L	6	embedding 514	.15	
L	7	embedding 257	-.17	
L	8	embedding 110	-.17	
L	9	embedding 573	.18	
L	10	topic 7 “love”	.16	
		(C4) Ambition		
L	1	embedding 011	-.18	
L	2	embedding 486	-.21	
L	3	topic 7 “love”	.14	
L	4	embedding 099	.20	
L	5	embedding 163	.22	
L	6	embedding 305	.21	
L	7	embedding 640	-.20	
L	8	embedding 017	.19	
L	9	embedding 148	-.14	
L	10	embedding 393	.20	

Table 3.3 (continued)

Group	Rank	Top Predictors	r_s	ALE Plot
		(E1) Friendliness		
L	1	embedding 443	.20	
L	2	embedding 692	-.20	
L	3	embedding 285	-.12	
L	4	embedding 474	-.17	
L	5	embedding 191	.12	
L	6	embedding 550	-.09	
L	7	embedding 269	.12	
L	8	embedding 387	-.13	
L	9	embedding 082	.12	
L	10	embedding 077	-.17	
		(E2) Sociableness		
L	1	embedding 066	.24	
L	2	embedding 443	.21	
L	3	embedding 474	-.15	
L	4	embedding 191	.16	
L	5	topic 5 "celebration"	.11	
L	6	embedding 302	.15	
L	7	embedding 190	.07	
L	8	topic 11 "goth"	-.12	
L	9	embedding 130	.14	
L	10	embedding 215	.17	
		(A1) Willingness to trust		
L	1	embedding 406	.16	
L	2	embedding 692	-.20	
L	3	negative emotion	-.18	
A	4	energy	-.20	
A	5	acousticness	.19	
L	6	embedding 259	.21	
L	7	embedding 607	-.15	
L	8	embedding 181	.01	
L	9	embedding 197	.06	
L	10	embedding 711	.00	
		(A2) Genuineness		
A	1	acousticness	.20	
L	2	embedding 714	.15	
A	3	energy	-.21	
L	4	embedding 555	-.14	
L	5	embedding 509	-.16	
L	6	embedding 646	-.13	
L	7	embedding 692	-.16	
L	8	embedding 137	-.18	
L	9	embedding 248	.11	
L	10	embedding 374	.11	

Note. The top 10 most important music-listening variables in decreasing order for each personality outcome that was predicted with a minimum performance of $r_s \geq .10$ (see Table 3.2). Significant prediction models are marked with an asterisk. The variables were selected and ranked based on the permutation feature importance extracted from the respective random forest models. Pairwise Spearman correlations between music-listening variables and personality outcomes illustrate the directionality of prediction effects. Colors in the two left-most columns indicate the group membership of each music-listening variable to either ■ (A) Audio Characteristics or ■ (L) Lyrics Characteristics, and, in the latter case, the specific type of Lyrics Characteristics, namely ■ Emotionality, ■ Topics, or ■ Word Embeddings. The two remaining groups of Habitual Listening Behavior and Other Lyrics Characteristics were not represented in the top 10 variables. For visibility, variables based on Lyrics' Word Embeddings, which are non-interpretable but make up most top predictors, are printed in grey font. In the right-most column, exemplary accumulated local effects (ALEs) are presented to illustrate how model predictions changed on average regarding different values in local value areas of the respective predictor. The x-axis differs depending on the variable's scale (see Table 3.1) and ranges between the 10th and 90th percentile of the variable's distribution. ALE values are centered around zero. Further ALE plots are in the Appendix (see Figure 3.A1).

The leftmost column in Table 3.3 shows that, across all outcomes, the majority of the most important variables represented lyrics' characteristics (127), followed by audio characteristics (13), while none of the top predictors captured habitual listening behaviors. This finding confirms the superiority of music preferences over habitual listening behaviors visible in the grouped importance presented earlier (see Figure 3.3). The color-coding in Table 3.3's second leftmost column further indicates that among lyrics characteristics, word embeddings were by far the most relevant group (121), followed by topics (5) and emotionality (1), while other lyrics characteristics were not among the most predictive variables.

For the different outcomes, the variables featured in the top 10 single predictors mostly represent groups identified as most relevant in Figure 3.3. For example, topics were the most predictive group as a whole and were among the most relevant individual variables for Conscientiousness (C) and its facets Sense of duty (C3), and Ambition (C4). However, there were also some discrepancies, where the most relevant individual variables did not (or only sparsely) contain predictors from the most important group. For example, the facet Openness to aesthetics (O2) had only two audio characteristics but eight lyrics characteristics in its top 10 predictors, even though the combined importance of these groups was reversed in Figure 3.3. One possible explanation is that audio characteristics are most predictive when combined as a group. For example, the music's loudness, tempo, and danceability may not be as informative on their own as they are together because only their constellations reveal what a song sounds like (e.g., a fast song vs. a fast, loud, energetic, and danceable song). If that were the case, the random forest models in our grouped benchmarks could have learned interaction effects from audio characteristics, resulting in high grouped prediction performances. In contrast, our single variable importance metric indicates the performance loss after permuting one specific variable and, thus, captures the relevance of a single variable but not its interactions.

For most of the Big Five domains, some individual music-listening variables were repeatedly listed in the top ten predictors across facets, highlighting the relevance of these particular variables in the respective random forest models. While many of these recurring variables were word embeddings (e.g., embedding 315 for Openness, embedding 486 for

Conscientiousness), we refrain from elaborating on them because word embeddings are non-interpretable. For Openness, predictions across several dimensions were higher for people listening to melodies with quieter, less danceable, and more acoustic audio characteristics (see Table 3.3). Similarly, two other audio characteristics representing lower energy and higher instrumentalness of melodies were also relevant for the prediction of one Openness facet each. Providing an exemplary effect interpretation, the ALE plots in Table 3.3 illustrate that random forest models using all variables predicted higher scores in Openness to imagination (O1) for participants listening to music with lower average values on the audio characteristics variable danceability. Regarding Conscientiousness, participants listening to lyrics with more love-related topics (see Figure 3.4 for topic interpretations) received higher predictions on the domain and two of its facets. For Extraversion, none of the top predictors were relevant across both facets inspected in Table 3.3. However, for the facet Sociableness (E2), people obtained higher predicted scores if they listened to lyrics with more celebration- and less goth-themed lyrics. Finally, for Agreeableness, people predicted to score high on the first two facets listened to melodies with less energetic and more acoustic audio characteristics. Furthermore, the models for the facet Willingness to trust (A1) predicted higher scores for participants listening to music with less emotionally negative lyrics, as visible in the ALE plot in Table 3.3.

Figure 3.4

Word Clouds of Most Predictive Lyrics Topics

Topic 5 "celebration" [+ E2] | **Topic 7 "love"** [+ C, C3, C4] | **Topic 11 "goth"** [- E2]

Note. Keywords of lyrics topics, that belong to the most important predictors of personality in random forest models (see Table 3.3). Preference for these topics was predictive for different Big Five scores indicated by square brackets: Topic 5 was relevant for higher Sociableness (E2); topic 7 for higher Conscientiousness (C) and its facets Sense of duty (C3) and Ambition (C4); and topic 11 for lower Sociableness (E2). Topics are part of a model with 30 topics obtained from training Latent Dirichlet Allocation on song lyrics. Keywords of the remaining, non-depicted topics can be found in Table 3.A4 in the Appendix. Word clouds show the 50 most frequent words of each topic. Words occurring in more than 60% of the topics' top 50 words and meaningless fill words (e.g., "yeah", "ooh") were removed for better interpretability. Word size indicates the relative frequency of a word within the topic, whereby larger words are more frequent. Quotation marks contain a post-hoc topic label based on visual inspection of the keywords.

3.5 Discussion

In the present study, we adopted a personality computing approach to explore individual differences in music-listening behavior on smartphones. We extracted an extensive set of variables representing natural music preferences in terms of various audio and lyrics characteristics as well as habitual listening behaviors, which we used to predict the Big Five dimensions on domain and facet level in a machine-learning framework. Afterward, we compared the independent contribution of the aspects of music listening, paying special attention to audio vs. lyrics preferences, and we inspected which single variables were most relevant in personality predictions.

3.5.1 Personality Prediction Based on Music-listening Behavior

To quantify the amount of personality-relevant information in digital music-listening records from smartphones, we assessed out-of-sample predictions of personality based on an extensive set of music-listening variables.

3.5.1.1 Overall Predictability Levels

Our results show that music-listening behavior was moderately predictive of personality with performances of $r_s > .20$ for the significant models, which corresponds to the average reported effect size in personality psychology (Funder & Ozer, 2019). However, we obtained only three significant prediction models and small to moderate effects (r_s between .10 and .21) for 11 other personality outcomes. This limited number and magnitude of effects is in line with the few and weak pooled correlations (six out of 30 coefficients ranging between .10 and .21) obtained between the Big Five domains and self-reported music style preferences in a meta-analysis by Schäfer and Mehlhorn (2017). In contrast, our out-of-sample prediction performances were lower, across domains, than those reported in a similar personality computing study by Anderson et al. (2021), who achieved correlations ranging from .26 to .37. between the Big Five and their predictions based on music-listening behavior on Spotify. While this latter study may seem to provide a fair comparison due to the close proximity in design, the discrepancy in results may be attributed to Anderson et al. 's (2021) significantly larger sample size (N > 5000) or their inclusion of demographic predictor variables (i.e., age and gender), which are known to be related to personality (Soto et al., 2011). Because our personality models used only behavioral predictors, our results seem reasonable, in particular, considering the bandwidth-fidelity dilemma we faced when predicting the Big Five dimensions, which aggregate the entirety of a person's thoughts, feelings, and behaviors, from music listening as one narrow excerpt of human behavior (Cronbach & Gleser, 1957; Rauthmann, 2021). We scratched on the lower range of successful personality prediction performances obtained from diverse behavioral indicators of smartphone usage (*r* =.20 to .40; Stachl et al., 2020) or from digital behaviors explicitly communicating self-views like social media postings (*r* = .28 to .42; Schwartz et al., 2013).

3.5.1.2 Differential Predictability Across Personality Dimensions

In our study, Openness and its facets were most predictable from music-listening behavior compared to the remaining Big Five dimensions. While this pattern is consistent with past findings on musical style and audio preferences (e.g., Dobrota & Reić Ercegovac, 2017; Greenberg et al., 2016; Nave et al., 2018; Rentfrow & Gosling, 2003; Schäfer & Mehlhorn, 2017), it seemingly contradicts Anderson et al.'s (2021) recent finding that Openness only ranked third in predictability from natural music-listening behavior on Spotify. However, their two top-ranking prediction performances for Emotional Stability and Conscientiousness strongly relied on the demographic predictor variable age, while their Openness models were predominantly based on music-listening predictors, so our findings align after all. The pattern of Openness being most strongly related to music listening corroborates the Big Five's conceptualization that more open individuals are generally more interested in different forms of art (DeYoung, 2015).

Albeit not obtaining significant predictions, the dimension of Conscientiousness was second most strongly related to music listening on smartphones. In previous work, Conscientiousness was associated with individuals' favorite song lyrics (Qiu et al., 2019) but not with preferences for musical styles or audio characteristics (e.g., Greenberg et al., 2016; Nave et al., 2018; Schäfer & Mehlhorn, 2017). This pattern was supported by our grouped and single variable importance metrics indicating that lyrics were of greater relevance than audio characteristics when relating music preferences to Conscientiousness.

The dimensions of Extraversion and Agreeableness were not strongly predicted by our music-listening variables, which is in line with a meta-analysis on musical style preferences by Schäfer and Mehlhorn (2017) and findings from music-listening behavior on Spotify (Anderson et al., 2021). As Anderson et al. (2021) noted, privately listening to music does not provide opportunities for social interaction, which, in turn, may suppress the expression of these socially defined traits (Goldberg, 1990). However, music from smartphones may also be used to promote social interactions (e.g., at parties), so associations with Extraversion and Agreeableness may become visible when considering the social listening context, for example, with whom somebody is listening to music.

Emotional Stability was the least predictable personality dimension in our study, which, again, corresponds to previous studies reporting weak relationships with musical style preferences (e.g., Nave et al., 2018; Schäfer & Mehlhorn, 2017). However, our results conflict with Qiu et al. (2019), who successfully related Emotional Stability to lyrics-based music preferences when only investigating participants' favorite songs, whose lyrics may be particularly meaningful compared to those of all played songs. While it seems reasonable that Emotional Stability may be connected to music listening (e.g., the emotionality of song lyrics), which is commonly used for emotion regulation, such relationships may vary intra-individually and be dependent on the emotional context of a music-listening situation (i.e., the listener's mood; e.g., Chamorro-Premuzic et al., 2010).

3.5.2 Importance of Different Aspects of Music-listening Behavior

Beyond disclosing its general predictive power, we applied interpretable machine-learning techniques to explore which granular aspects of natural music-listening behavior were most informative for personality predictions.

3.5.2.1 Variable Groups

Overall, music preferences in terms of audio and lyrics characteristics were both predictive of listeners' personalities (especially the Openness dimension), while habitual listening behaviors played no major role in our models. Among lyrics characteristics, the technically most sophisticated but non-interpretable word embeddings were most informative across outcomes, followed by lyrics' topics (especially for Conscientiousness), while lyrics' emotionality and other aspects (e.g., lyrics length) appeared less relevant. This rank order among natural language models hints at the advantages of open-vocabulary approaches when predicting personality from textual properties, which was previously reported for other text sources (e.g., Park et al., 2015; Schwartz et al., 2013).

At the trait level, preferences for audio and lyrics characteristics exhibited differential prediction performances for most personality dimensions, most notably for Conscientiousness, where lyrics outperformed audio characteristics, and for Extraversion, where audio characteristics outperformed lyrics. These findings may relate to the independent cognitive processing of melodies and lyrics (Besson et al., 1998; Bonnel et al., 2001) and indicate that both audio and lyrics should be considered when investigating music preferences in personality science.

3.5.2.2 *Individual Variables*

When considering individual music-listening variables, the most important (interpretable) predictors were generally congruent with both past findings and the Big Five conceptualization (see DeYoung, 2015; Goldberg, 1990). As an example, that was the case for the positive associations between calm melodies and Openness to feelings, which were previously reported on the domain-level by other studies on audio characteristics (Dobrota & Reić Ercegovac, 2017; Fricke & Herzberg, 2017), or for the positive relations between celebration-themed lyrics and the Extraversion facet of Sociableness, which support previously found associations between the Extraversion domain and positive emotion words in lyrics (Qiu et al., 2019).

Because our study design did not consider causality, these associations may indicate that listeners adjusted their auditory environments to their personalities or vice versa (e.g., Buss, 1987; Bleidorn et al., 2020; Fleeson, 2001; Rauthmann, 2021; Swann, 1987). On the one hand, people with high levels of Openness to feelings may choose calm melodies to accommodate their emotional sensitivity, and those high in Sociability may listen to celebration-themed lyrics to help them experience positive social interactions. On the other hand, repeated exposure to calm melodies may provide opportunities for emotional experiences, which, in turn, accumulate to higher levels of Openness to feelings. Similarly, frequently listening to celebration-themed lyrics may give rise to positive social interactions and, in the long run, cause people to become more extraverted. While most of our variable importance ranking seems plausible in this sense, some findings were surprising, adding potentially new facets to the theoretical trait concepts. For example, the preference for love-related lyrics is rather difficult to reconcile with high levels of Conscientiousness, a trait typically characterized by planning behavior and obedience to norms (Roberts et al., 2004). In sum, these results demonstrate that specific granular aspects of music-listening behavior are distinctly informative about the different Big Five dimensions.

3.5.3 Constraints on Generalizability

We follow the recommendation by Simons et al. (2017) and discuss the generalizability of our empirical findings for different samples, materials, and contexts.

The present study investigated three ad hoc samples of mostly young participants with high education levels, which, given our university recruiting context, suggests that German university students were our proximal population. We are, however, confident that our findings generalize beyond this specific population because the associations we found between personality traits and music preferences generally aligned with those obtained in past studies investigating university students from other countries (e.g., Dobrota & Reić Ercegovac, 2017; Rentfrow & Gosling, 2003; Qiu et al., 2019) or more diverse samples of Facebook users (e.g., Greenberg et al., 2016; Nave et al., 2018). Nevertheless, the young mean age of participants in both our and past music research referenced above may have reduced our sample's variance in personality traits and music-listening behaviors, which both appear to change with age (e.g., Bonneville-Roussy et al., 2013; Lucas & Donnellan, 2011). Hence, we believe our results may not necessarily generalize to samples including older adults, which are currently underrepresented in music-listening research. Furthermore, our and past samples were exclusively representative of WEIRD (i.e., Western, Educated, Industrialized, Rich, Democratic) populations (Henrich et al., 2010). While the Big Five structure of personality (e.g., McCrae & Terracciano, 2005) and its reflection in preferences for Western musical styles were found to generalize across countries (Greenberg et al., 2022), the musical styles actually listened to differ between countries and cultures (e.g., Bello & Garcia, 2021; Park et al., 2019). Thus, natural music-listening behavior and its relation to personality may look differently in non-Western populations. Finally, our specific study's sample was limited to users of Android smartphones due to technical reasons, excluding those owning iPhone Operating System (iOS) devices. However, as previous studies found no meaningful differences in demographic and personality characteristics between Android and iOS users, this bias should not dramatically impact the generalizability of our findings (Götz et al., 2017; Keusch et al., 2020). To summarize, we believe that our findings are representative of young adults in Western societies and recommend that follow-up studies generalize our approach to samples including older adults and other cultures.

While the subject of our study was natural music-listening behavior exhibited on smartphones, we assume that our personality predictions transfer to all forms of private digital music consumption, including all listening instances where participants can freely choose what music to listen to from their own or a very large collection of songs. That should include music listening on any digital device with music storing or streaming functionalities, such as computers or smart TVs, because our data collection took place at a time when music streaming was on the rise, but when some people still listened to locally stored music on their smartphones. In contrast, music listening on more old-fashioned analog devices such as record players may differ from that on smartphones due to the restricted availability of contemporary songs in the respective formats. This may, in turn, introduce systematic differences in music preferences between non-digital and digital devices (e.g., playing only oldies on the record player but more modern hits on digital devices) and hinder replication of our study. Furthermore, our personality patterns may not generalize to individuals' full spectrum of music-listening behavior when including instances where music is not self-chosen, such as music listening on the radio or at a café.

The most important aspect of our procedure was that we assessed music-listening behavior with high ecological validity and in an unobtrusive and objective manner via smartphone sensing. To replicate our findings, future studies should also assess digital music-listening records, either obtained from listening devices or directly from streaming services such as Spotify (see Anderson et al., 2021). This procedure, however, excludes populations currently not listening to music digitally, such as older people and people in developing countries with very low smartphone penetration. In contrast, when assessing music listening in a more intrusive way (e.g., in laboratory settings), participants may adapt their behaviors in a socially desirable manner (e.g., based on assumptions about researcher goals), so replication is not guaranteed. Similarly, we do not expect our findings to fully generalize to self-reported music-listening behaviors, even though they exhibited some overlap with studies on self-reported music preferences (see Schäfer & Mehlhorn, 2017). Finally, to replicate our procedure, it is important to represent music in terms of the intrinsic properties of its melodies and lyrics instead of broad musical styles or genres. The automatic approaches for extracting these musical characteristics can be transferred to all samples of music worldwide and is, thus, widely applicable.

Beyond the considerations outlined above, we currently have no reason to believe that our results depended on other characteristics of the participants, materials, or procedure.

3.5.4 Limitations and Future Directions

The present study has several limitations. First, the relatively small sample size may have prevented our machine-learning algorithms from detecting stable patterns that transfer from training to test sets in our cross-validated resampling scheme. Thus, our low prediction performances represent a rather conservative estimate of how well personality may be predicted from music-listening behavior in larger samples. Second, careless or insufficient effort responding to our lengthy self-report measure (300 items) may have further attenuated associations between music listening and personality traits (Curran, 2016; Ward & Meade, 2023). While most of our self-reports appeared plausible, different post-hoc response validity analyses identified few participants suspicious of careless responding (see the Appendix; Curran, 2016). However, our random forest algorithms are rather robust to outliers and should, thus, not have been impacted too dramatically by the inclusion of potentially careless responses

(Breiman, 2001). Third, the music preference variables extracted from participants' song records depended on the availability of external song-level information (i.e., Spotify's audio metrics and Genius' lyrics), possibly resulting in an underrepresentation of uncommon songs and restricted prediction performances for participants with an exotic taste in music. Fourth, our lyrics-based variables may not necessarily represent conscious preferences for song lyrics because we could not confirm that our German participants had fully understood the mainly English lyrics they were listening to. While most young Germans speak English fluently[2], personality patterns in lyrics preferences may be even more pronounced when considering only lyrics in the sample's mother tongue. Fifth, we could not distinguish instances where participants played the music on their smartphones themselves from those where others (e.g., friends or children) initiated music-listening events, which may have introduced noise to participants' music-listening metrics.

Our study demonstrates the potential of smartphone sensing for music-listening research in personality psychology and beyond. As popular music-listening devices, smartphones allowed us to collect digital records of participants' day-to-day listening habits and music preferences over time (Greenberg & Rentfrow, 2017). However, our rather traditional approach to investigating average music-listening metrics captures only a small proportion of the information in these longitudinal data. For example, if a person listens to either very calm or very energetic melodies, their average score cannot accurately represent their music-listening behavior. Thus, to seize the full potential of digital music-listening records, future studies should analyze variations in single listening events over time instead of aggregating them. When investigating listening events nested within persons, personality traits may exhibit relations to intra-individual variations in music listening. For example, the trait Openness, which was previously associated with more diverse average music preferences (Bansal et al., 2020), may be even more predictable from variations within individuals' music-listening events than from aggregated scores. Beyond stable personality traits, future research may also include momentary aspects such as mood states or situation perceptions to explain intra-individual variance in music-listening behavior, which remains largely unexplored to this date. Smartphone-based ambulatory assessment has laid the foundation for this kind of research because it enables the simultaneous collection of objective music-listening and other contextual data (e.g., places where music listening occurred; see Schoedel et al., 2023) via smartphone sensing and self-reported subjective experiences in-situ via the experience-sampling methodology (van Berkel et al., 2017). The combination of passive smartphone sensing with

[2] Due to compulsory English language schooling from Kindergarden onwards.

active ES is quite novel but provides great opportunities for personality research in general (Schoedel et al., 2023). In sum, smartphones open up ample possibilities for investigating the interplay of various enduring and fluctuating variables, which will broaden our understanding of music-listening behavior.

3.6 Conclusion

The present study demonstrates that smartphone sensing is a promising method to investigate natural music-listening behavior and its association with personality. Overcoming self-report assessments of broad musical style preferences, we introduced a personality computing framework for predicting the Big Five dimensions from preferences for intrinsic musical properties and habitual listening behaviors extracted from digital music-listening records. Machine-learning models revealed that only the personality dimension of Openness was successfully predicted from our music-listening variables, corroborating past findings that out of the Big Five, Openness is most strongly related to music listening. In contrast, Conscientiousness and several personality facets showed non-significant but small to moderate prediction effects in our models. Furthermore, our study compared the contribution of audio and lyrics characteristics for relating music preferences to personality, finding that they are both distinctly predictive and that the associations between specific music preference variables and certain personality traits were generally in line with the Big Five's theoretical conceptualization. In sum, our findings provide new insights into personality patterns in natural music-listening behavior, which may be extended in numerous ways using the methodological framework proposed here.

3.7 Acknowledgements

We thank the entire PhoneStudy team at LMU Munich for their continuous and diligent work on the PhoneStudy app, making our research possible. In particular, we thank Tobias Schuwerk for his role in providing parts of the data. We also thank Schuhfried GmbH for providing the Big Five Structure Inventory as a digital version. We further thank Florian Pargent and Felix Schönbrodt for giving insightful modeling advice. Special thanks go to Monika Wintergerst for advising on the language modeling in our research project.

3.8 Appendix

3.8.1 Supplemental Method

Dataset

Table 3.A1

Description of Datasets Used in the Study

Data set	N	Collection period	Logging days	% Music days	Music events	% Music events with lyrics	Reference
Study 1	67 (137)	09/2014 - 08/2015	47 - 60	43.64 %	30654	38.33 %	Stachl et al., 2017
Study 2	105 (245)	08/2016 - 08/2017	22 - 38	50.03 %	24582	44.57 %	Schuwerk et al., 2019
Study 3	158 (316)	10/2017 - 01/2018	3 - 85	46.55 %	75970	42.31 %	Schoedel et al., 2019
This study	330	09/2014 - 01/2018	3 - 85	47.06 %	131206	42.22 %	

Note. N indicates the size of the sample from the respective study after the application of our inclusion criteria. The total sample size per study is given in parentheses. Logging days indicate the minimum and the maximum number of days with smartphone-sensing data within the study sample. % Music days indicates the sample's average percentage of study days where music listening occurred on the smartphone out of all logging days. Music events indicate the total amount of logged music-listening activities that lasted longer than 20 seconds (i.e., unskipped music events) across all participants and study days. % Music events with lyrics indicate the percentage of all music events that could be matched with external lyrics data.

Response Validity Analyses. With its extensive length of 300 items, the BFSI (Arendasy, 2009) has an increased risk of triggering careless or insufficient effort (C/IE) responding (Curran, 2016; Ward & Meade, 2022). Hence, we conducted a medium-level analysis of response validity as recommended by Ward and Meade (2022). Please note that we performed these analyses post-hoc (i.e., after our machine-learning benchmarks) to estimate the impact of careless responding on our predictions and not to remove participants beforehand. To detect different forms of careless responding, we combined a) multivariate outlier analysis via Mahalanobis Distance, invariance analysis via b) the long-string index, and c) intraindividual response variability (IRV), and d) consistency analysis via the even-odd index (Johnson, 2005; Meade & Craig, 2012). All analyses were conducted with the careless package in R (Yentes & Wilhelm, 2021) and the respective code is available in our project repository. Mahalanobis

Distance revealed no multivariate outliers, indicating that none of our participants exhibited aberrant responses across all items. Long-string analysis detected 32 cases with over 30 identical responses in consecutive items two of which even lasted for over 50 items. Similarly, IRV identified seven participants whose intra-individual standard deviation across items was more than two standard deviations below the sample's mean. While these two indices seem to flag some participants as careless, invariability should not be over-interpreted in the context of the BFSI, which contains 60 items assessing the same Big Five dimension and consists only of adjectives with the same directionality and intensity (Dunn et al., 2018). Finally, the even-odd consistency was critically low (i.e., below the recommended cutoff of .30) for only one participant indicating a lack of consistent responses within the BFSI's sub-scales. In sum, we cannot rule out that our analyses also included instances of invalid data produced through careless responding. However, we refrained from removing these participants for a lack of appropriate and unambiguous evidence and discuss the limited response validity instead (Curran, 2016).

Personality Measure

Table 3.A2

Descriptive Statistics of Big Five Personality and Demographic Variables

Variable	M	SD	Median	Min	Max	Range	alpha CI95%
Gender	1.46	0.50	1	1	2	1	-
Age	22.42	4.33	21	18	57	39	-
Education	4.10	0.58	4	2	5	3	-
(O) Openness	-0.01	0.73	-.011	-2.00	2.12	4.12	[0.93, 0.95]
(O1) Openness to imagination	1.33	1.43	1.30	-2.30	5.33	7.62	[0.84, 0.89]
(O2) Openness to aesthetics	0.43	1.30	0.29	-2.79	4.61	7.40	[0.84, 0.88]
(O3) Openness to feelings	2.06	2.15	2.00	-3.86	6.04	9.90	[0.91, 0.94]
(O4) Openness to actions	1.42	1.47	1.18	-2.18	5.42	7.60	[0.84, 0.88]
(O5) Openness to ideas	1.69	1.49	1.29	-1.63	5.52	7.15	[0.82, 0.87]
(O6) Openness to value & norm	0.98	1.04	0.94	-3.54	4.86	8.41	[0.74, 0.81]
(C) Conscientiousness	-0.06	0.74	-0.10	-2.17	2.29	4.46	[0.95, 0.96]
(C1) Competence	0.91	1.18	0.72	-2.10	5.66	7.76	[0.74, 0.81]
(C2) Love of order	1.17	1.54	1.16	-4.34	5.67	10.01	[0.87, 0.90]
(C3) Sense of duty	1.97	1.38	1.78	-1.59	5.50	7.10	[0.78, 0.84]
(C4) Ambition	1.82	1.68	1.58	-2.39	5.86	8.25	[0.85, 0.89]
(C5) Discipline	1.48	1.47	1.50	-3.61	5.75	9.36	[0.82, 0.87]
(C6) Caution	1.59	1.37	1.45	-2.51	5.75	8.26	[0.80, 0.85]
(E) Extraversion	0	0.72	0.03	-2.18	1.96	4.14	[0.95, 0.96]
(E1) Friendliness	1.43	1.25	1.37	-1.70	5.41	7.11	[0.77, 0.83]

Table 3.A2 (continued)

Variable	M	SD	Median	Min	Max	Range	alpha CI95%
(E2) Sociableness	1.26	1.74	1.28	-4.50	5.64	10.14	[0.89, 0.92]
(E3) Assertiveness	0.45	1.37	0.47	-3.35	4.41	7.75	[0.84, 0.88]
(E4) Dynamism	1.25	1.59	1.15	-2.92	5.94	8.86	[0.85, 0.89]
(E5) Adventurousness	0.46	1.50	0.55	-4.40	5.27	9.67	[0.87, 0.91]
(E6) Cheerfulness	2.05	1.65	1.85	-1.88	6.09	7.96	[0.85, 0.89]
(A) Agreeableness	-.03	0.79	-0.11	-1.94	2.64	4.58	[0.93, 0.95]
(A1) Willingness to trust	0.42	1.47	0.31	-5.34	5.42	10.76	[0.85, 0.89]
(A2) Genuineness	1.02	1.01	0.86	-1.56	4.25	5.81	[0.65, 0.74]
(A3) Helpfulness	1.70	1.38	1.67	-2.02	6.04	8.06	[0.77, 0.83]
(A4) Obligingness	1.24	1.38	1.05	-2.10	5.55	7.64	[0.81, 0.86]
(A5) Modesty	0.73	1.14	0.73	-2.68	5.11	7.79	[0.78, 0.84]
(A6) Good naturedness	2.19	1.86	2.09	-2.06	6.40	8.46	[0.86, 0.90]
(ES) Emotional Stability	0.04	0.72	0.05	-2.19	2.52	4.71	[0.93, 0.95]
(ES1) Carefreeness	0.23	1.33	0.27	-4.30	4.30	8.60	[0.82, 0.87]
(ES2) Equanimity	0.65	1.07	0.70	-2.61	5.02	7.63	[0.78, 0.84]
(ES3) Positive mood	1.09	1.46	1.08	-5.78	5.60	11.38	[0.84, 0.89]
(ES4) Self consciousness	0.69	1.18	0.78	-3.57	3.90	7.47	[0.82, 0.87]
(ES5) Self control	0.69	0.99	0.74	-3.55	3.96	7.51	[0.73, 0.81]
(ES6) Emotional robustness	0.73	1.18	0.79	-3.32	5.53	8.85	[0.79, 0.85]

Note. N = 330; alpha CI95% = 95% confidence intervals for Cronbach alpha coefficients.

Behavioral Music-Listening Measures: Song-Level Variables

Retrieval of External Song-Level Data. Whenever participants had played music on their smartphones, our custom Android research application created a time-stamped event log containing the title, artist, and album name of the played track. To obtain additional information about the played tracks, we conducted API calls from Spotify.com and Genius.com using the combination of the tracks' logged title-artist-album triples. However, in some cases, the logged track information returned no match, for example, if a song had incorrect titles (e.g., spelling mistakes) or additional information (e.g., "[Bonus CD]") in the title. To still retrieve external data in these cases, we used several heuristics. First, we removed special characters (e.g., punctuation) and "tags" (i.e., bracketed characters, e.g., "[Bonus CD]"). If this was still unsuccessful, we subsequently tried further modifications, such as searching without an album or with a "split" artist field (sometimes artist data incorrectly contained a list of artists). These modified search titles served to retrieve audio features from Spotify.com and song lyrics from Genius.com. After enriching the music-listening records with these external data, we verified the match between the modified search titles and the original logged track titles. We automatically checked for perfect matches via the Levenshtein similarity and manually validated all cases with a similarity score under one to keep only close matches (s., the validated song lists in our repository). This procedure improved the coverage of successfully matched tracks without compromising on the quality. In one additional step, we filtered all correct matches for tracks with a Spotify *speechiness* value greater than .60 to remove non-musical entries (e.g., audiobooks; Spotify, 2022). Starting with a total of 78.165 tracks played in our study, we obtained external Spotify data for 55% and lyrics data for 37% of all tracks. Unmatched tracks either contained non-musical tracks, had incorrect song information, or were not covered by the respective online sources.

Analysis of Song Lyrics. After obtaining the lyrics of the songs played in our study, we applied a text-mining pipeline to describe them in terms of objectively computed characteristics. We provide the full code for these analyses in our project repository. First, we cleaned the lyrics from formatting issues (e.g., annotations like "chorus" or missing text repetitions indicated by "2x"). Then, we combined two language detection algorithms (Joulin et al., 2016; Lui, 2016) to create a variable indicating whether a song was in English, German, or another language. We filtered for English and German lyrics (96 % of all songs), which were most likely understood by our German student sample, to ensure that our variables reflected conscious lyrics preferences. Finally, to enable natural language modeling, we translated German lyrics into English with the neural-network-based translation software DeepL (DeepL GmbH, n.d.). We describe the lyrics preprocessing in greater detail in the repository. Next, we applied different natural language models, including LDA. We trained and evaluated LDA models on a separate corpus of over 180.000 lyrics from the Million Song Dataset (MSD; Bertin-Mahieux et al., 2011) to avoid overfitting the topic distributions to sample-specific patterns. After preprocessing the MSD corpus in parallel to our original lyrics file (s. details in the repository), we fit 12 topic models differing in the choice of priors and the number of topics. We tested a fixed default value of .01 for the priors alpha and beta against an optimization of the prior alpha that allows some topics to be more prominent than others. These two settings were paired with six choices for the number of topics: 15, 30, 50, 60, 75, 120. We evaluated the resulting models and chose the winner settings such that the topic coherence measure *u_mass* was maximized (Rehurek & Sojka, 2010). The winner model with fixed priors and 30 topics (s., Table 3.A3) served to infer topic distributions on our initial lyrics corpus. In Table 3.A4, we provide the keywords for the final topic solution. Another language model we applied to our lyrics corpus was a pre-trained implementation of BERT by Wolf et al. (2020). BERT provided one embedding vector for each word in a song's lyrics, plus one additional [CLS]-

token embedding, which we used as a condensed numerical representation of the entire lyrics. However, 10% of our lyrics exceeded BERT's maximum input of 512 words per sequence, so we needed a cutting heuristic. When the excess words made up less than half of a song's lyrics, we followed the standard procedure to cut off the lyrics' tail which often contained fade-outs or chorus repetitions. In the remaining cases, we parsed the lyrics into multiple chunks, extracted separate embedding vectors for each chunk, and averaged them later. The reproducible code for all lyrics analyses is available in the OSF repository. Single functions have been adapted from other authors and highlighted within the code for lyrics preprocessing (McKew, 2020), LDA modeling (Konrad, 2016), and song-level variable extraction (Bertin-Mahieux, 2011; The Hugging Face Team, 2020).

Table 3.A3

Tuning Results for Different LDA Specifications

Number of topics	Alpha	Beta	Number of iterations	Optmization interval	Topic coherence	Training time
15	0.01	0.01	1000	OFF	-1.675	09 min 19 sec
15	-	0.03	1000	ON-100	-2.410	11 min 12 sec
30	0.01	0.01	1000	OFF	-1.661*	15 min 33 sec
30	-	0.02	1000	ON-100	-2.707	09 min 00 sec
50	0.01	0.01	1000	OFF	-1.691	21 min 00 sec
50	-	0.01	1000	ON-100	-2.934	09 min 46 sec
60	0.01	0.01	1000	OFF	-1.702	19 min 31 sec
60	-	0.01	1000	ON-100	-3.017	09 min 58 sec
75	0.01	0.01	1000	OFF	-1.689	29 min 40 sec
75	-	0.01	1000	ON-100	-3.084	21 min 57 sec
120	0.01	0.01	1000	OFF	-1.738	31 min 16 sec
120	-	0.01	1000	ON-100	-3.309	20 min 13 sec

Note: The prior alpha governs whether the documents in LDA contain an exclusive (smaller values) or broad (larger values) range of topics. The prior beta determines whether the words in LDA belong to many (larger values) or only a few topics (smaller values). For the models with optimization "ON", the prior alpha was optimized for each topic separately. Hence the absence of an overall alpha value for the respective models. The number of iterations determines how many iterations the LDA uses to find the optimal topic-document and word-topic distributions, whereby 1000 iterations are the default value. The optimization interval sets the length of the search interval of hyperparameter optimization. It was set to 100 for all models. Models without hyperparameter optimization have no corresponding value. *The lowest coherence score is marked by an asterisk.

Table 3.A4

Top 20 Highest-Frequency Words per Topic in the Winner LDA Model

Topic	Top 20 Words
0	love, heart, baby, fall, dream, always, hold, believe, find, long, everything, leave, cry, wait, nothing, tonight, alone, stay, something, change
1	kill, die, lie, nothing, man, people, head, fight, god, end, f*ck, hate, face, everything, believe, stand, dead, change, inside, war
2	F*ck, man, sh*t, people, n*gga, money, hate, die, put, rock, head, stop, run, kill, talk, face, god, street, gun, start
3	god, soul, die, burn, blood, lord, death, heart, light, lie, fire, dark, end, love, fear, man, dead, fall, pain, word
4	love, baby, heart, long, home, dream, happy, hold, sing, cry, little, find, always, believe, light, wait, everything, song, remember, shine
5	christmas, man, lyrics, shake, good, baby, boy, girl, love, little, run, home, town, gal, rock, play, sing, dance, ride, fun
6	love, baby, good, girl, little, man, nobody, dance, talk, woman, move, everybody, blue, home, work, boy, walk, shake, long, really
7	love, heart, dream, cry, home, find, long, wait, leave, always, hold, baby, good, alone, fall, tonight, lonely, light, stay, shine
8	love, heart, song, sing, friend, little, long, always, word, hold, dream, nothing, hear, find, home, something, fall, wait, alone, walk
9	love, home, walk, man, run, little, long, head, dream, find, light, heart, leave, always, friend, town, sun, place, people, nothing
10	n*gga, sh*t, man, f*ck, rock, stop, b*tch, boy, put, hit, big, money, y'all, girl, baby, ass, real, beat, show, n*ggaz
11	light, heart, love, fall, dream, burn, soul, dark, god, die, sun, fire, sky, face, end, inside, pain, lie, hear, wait
12	love, baby, heart, girl, believe, hold, cry, good, find, always, true, long, little, stay, someone, really, hurt, please, somebody, thing
13	baby, love, girl, little, good, boy, tonight, hot, man, crazy, rock, ready, everybody, dance, talk, alright, body, really, stop, show
14	love, nothing, always, believe, heart, find, dream, something, change, wait, alone, walk, leave, fall, inside, lost, hold, everything, left, end
15	love, everything, heart, long, find, dream, wait, change, fall, run, leave, nothing, always, good, baby, something, walk, believe, move, alone
16	n*gga, f*ck, sh*t, b*tch, man, n*ggaz, y'all, ass, money, hit, put, real, game, big, h*e, f*ckin, check, street, head, boy
17	love, god, heart, dream, die, lord, soul, dead, death, blood, light, burn, fall, end, dark, lie, fire, save, stand, fear
18	love, baby, girl, dance, tonight, good, heart, little, stop, really, shake, boy, hold, move, kiss, true, lover, real, sweet, stay
19	baby, love, little, girl, rock, man, roll, boom, home, good, dance, long, stop, heart, boy, play, bye, tonight, talk, touch
20	head, run, dead, man, die, face, light, heart, dream, fall, end, love, sky, inside, lie, burn, walk, leave, watch, black

Table 3.A4 (continued)

Topic	Top 20 Words
21	love, baby, girl, good, alright, little, tonight, wait, please, heart, long, home, everything, mine, really, man, dance, boy, stay, hold
22	love, run, find, nothing, wait, heart, lie, inside, leave, fall, change, hold, long, everything, something, always, end, face, believe
23	love, heart, fall, light, dream, find, wait, inside, hold, nothing, end, leave, face, breath, die, believe, lie, alone, rain, run
24	love, home, little, man, long, ride, sing, song, good, blue, play, light, rock, girl, roll, boy, town, hear, sun, music
25	man, girl, big, boy, rock, n*gga, sh*t, f*ck, put, ride, roll, head, good, baby, hit, money, play, party, little, people
26	love, heart, baby, always, hold, find, something, believe, leave, nothing, lie, wait, cry, dream, alone, girl, everything, break, wrong, fall
27	god, love, burn, heart, fire, soul, light, lord, die, heaven, dark, jesus, fall, man, holy, blood, sky, dream, beautiful, angel
28	man, dead, head, f*ck, little, kill, die, people, boy, hell, blood, god, city, good, lyrics, black, face, big, love, play
29	man, money, people, f*ck, work, little, good, put, kid, play, run, love, friend, head, sh*t, hit, move, big, beat, boy

Note. The 20 most frequent words for each topic of our lyrics-based LDA model. Words occurring in more than 60% of the topics' high-frequency words and meaningless fill words (e.g., "yeah") were removed for better interpretability. The word ranking is based on the word's relative frequency within a certain topic and goes from left (high) to right (low).

Personality Predictions: Performance Evaluation

Prediction Performance Metrics. We used the following two metrics for evaluating the performance of our personality prediction models:

Spearman correlation: $$(r_s) = \frac{\sum_{i=1}^{n}\left(\left(R(\gamma_i) - \overline{R(\gamma)}\right)\left(R(\hat{\gamma}_i) - \overline{R(\hat{\gamma})}\right)\right)}{\sqrt{\sum_{i=1}^{n}\left(R(\gamma_i) - \overline{R(\gamma)}\right)^2 \sum_{i=1}^{n}\left(R(\hat{\gamma}_i) - \overline{R(\hat{\gamma})}\right)^2}}$$

Mean Squared Error: $$(MSE) = \frac{1}{n}\sum_{i=1}^{n}\left(\gamma_i - \hat{\gamma}_i\right)^2$$

Where:

$R(\gamma_i)$: rank of the true criterion value for the observation *i*

$\overline{R(\gamma)}$: mean rank of the criterion value

$R(\hat{\gamma}_i)$: rank of the predicted criterion value for the observation *i*

$\overline{R(\hat{\gamma})}$: mean rank of the predicted criterion value

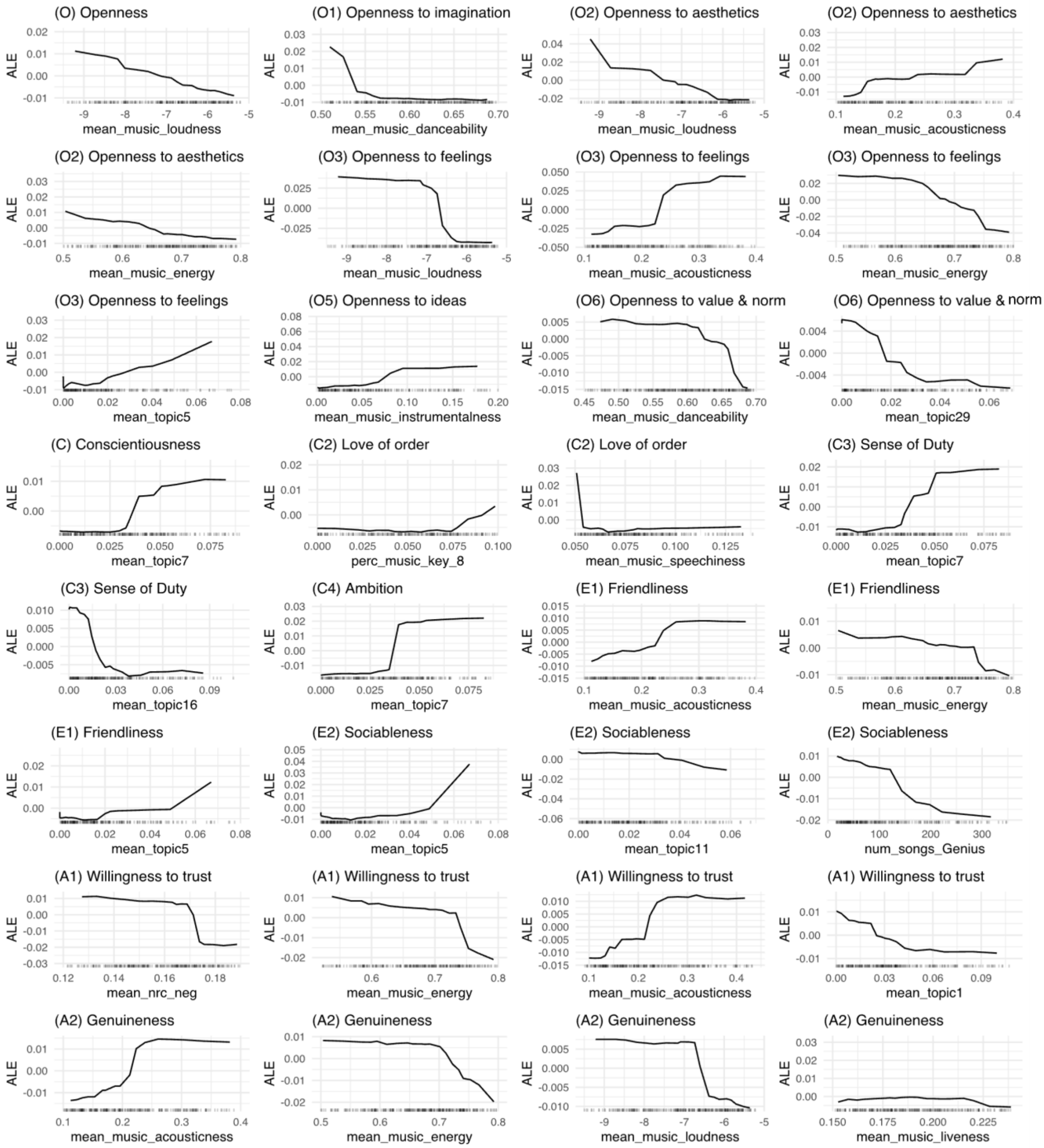
γ_i : true criterion value for the observation *i*

$\hat{\gamma}_i$: predicted criterion value for the observation *i*

3.8.2 Supplemental Results

Figure 3.A1

Selected Accumulated Local Effect Plots for Random Forest Models



Note. Accumulated local effects (ALEs) for random forest models with a minimum prediction accuracy of r_s=.10 (see Table 3.2). ALEs are presented for the 15 most important predictors (i.e., those with the highest permutation feature importance). Predictors based on the group of Lyrics' Word Embeddings are excluded as they are non-interpretable. ALE plots indicate how model predictions changed on average with regard to different values in local value-areas of the predictor. Values on the x-axis range between the 10th and 90th percentile of the respective variable's distribution and vertical stripes on the x-axis show the data distribution. ALE values are centered around zero. The audio characteristic loudness is measured in dB, while all other audio characteristics and lyrics topics and emotionality range between 0 and 1 (see Table 3.1).

Table 3.A5

Mean Prediction Performance per Personality Dimension and Variable Group for Random Forest Models

Personality Dimension	Full Predictor Set	Habitual Listening Behavior	Audio Characteristics	Lyrics Characteristics	Lyrics Emotionality	Lyrics Topics	Lyrics Word Embeddings	Other Lyrics Characteristics
(O) Openness	0.24 [0.11,0.37]	0.05 [-0.06,0.16]	0.20 [0.09,0.31]	0.23 [0.10,0.36]	0.06 [-0.05,0.18]	0.17 [0.06,0.28]	0.23 [0.11,0.36]	0.17 [0.06,0.27]
(O1) O. to imagination	0.24 [0.13,0.36]	0.04 [-0.08,0.16]	0.13 [0.02,0.25]	0.23 [0.12,0.35]	0.10 [0.00,0.20]	0.19 [0.07,0.30]	0.23 [0.11,0.34]	0.09 [-0.01,0.20]
(O2) O. to aesthetics	0.20 [0.09,0.32]	0.12 [0.02,0.23]	0.26 [0.15,0.37]	0.19 [0.08,0.30]	0.05 [-0.06,0.15]	0.07 [-0.02,0.16]	0.19 [0.08,0.31]	0.16 [0.05,0.27]
(O3) O. to feelings	0.23 [0.12,0.34]	0.07 [-0.03,0.17]	0.29 [0.19,0.39]	0.21 [0.11,0.32]	0.04 [-0.07,0.16]	0.14 [0.03,0.25]	0.21 [0.10,0.33]	0.09 [-0.03,0.21]
(O4) O. to actions	0.03 [-0.08,0.14]	0.09 [-0.02,0.21]	0.05 [-0.06,0.16]	0.03 [-0.08,0.13]	0.04 [-0.08,0.15]	0.00 [-0.12,0.11]	0.03 [-0.08,0.13]	0.13 [0.02,0.25]
(O5) O. to ideas	0.17 [0.05,0.29]	-0.04 [-0.16,0.08]	0.09 [-0.02,0.20]	0.16 [0.05,0.28]	0.08 [-0.04,0.19]	0.16 [0.05,0.27]	0.16 [0.04,0.27]	0.24 [0.14,0.34]
(O6) O. to value & norm	0.15 [0.04,0.26]	0.08 [-0.03,0.19]	0.15 [0.04,0.26]	0.15 [0.04,0.26]	-0.03 [-0.15,0.08]	0.14 [0.04,0.24]	0.15 [0.04,0.26]	0.09 [-0.02,0.20]
(C) Conscientiousness	0.12 [0.00,0.24]	0.06 [-0.05,0.17]	0.07 [-0.05,0.19]	0.12 [0.00,0.24]	0.03 [-0.10,0.15]	0.15 [0.03,0.27]	0.12 [0.00,0.24]	0.05 [-0.06,0.17]
(C1) Competence	0.01 [-0.11,0.12]	-0.01 [-0.12,0.10]	0.04 [-0.07,0.15]	0.01 [-0.10,0.13]	-0.03 [-0.15,0.09]	0.09 [-0.03,0.20]	0.01 [-0.10,0.12]	0.03 [-0.08,0.14]
(C2) Love of order	0.16 [0.04,0.27]	0.06 [-0.05,0.16]	0.17 [0.07,0.27]	0.16 [0.04,0.28]	0.02 [-0.09,0.13]	0.13 [0.02,0.24]	0.16 [0.04,0.27]	0.06 [-0.05,0.18]
(C3) Sense of duty	0.10 [-0.01,0.21]	0.10 [-0.01,0.20]	-0.04 [-0.15,0.07]	0.11 [0.00,0.22]	-0.10 [-0.20,0.00]	0.16 [0.06,0.26]	0.11 [0.00,0.22]	0.02 [-0.10,0.13]
(C4) Ambition	0.14 [0.01,0.27]	0.08 [-0.02,0.17]	-0.03 [-0.13,0.08]	0.15 [0.02,0.27]	0.03 [-0.09,0.15]	0.10 [0.01,0.19]	0.14 [0.02,0.27]	0.11 [0.00,0.21]
(C5) Discipline	0.04 [-0.07,0.14]	0.01 [-0.09,0.11]	0.01 [-0.11,0.12]	0.05 [-0.06,0.15]	0.09 [-0.02,0.20]	0.10 [-0.01,0.20]	0.04 [-0.06,0.15]	0.13 [0.03,0.22]
(C6) Caution	0.05 [-0.05,0.16]	0.12 [0.02,0.21]	0.00 [-0.11,0.12]	0.06 [-0.05,0.17]	0.10 [-0.01,0.21]	0.08 [-0.04,0.19]	0.05 [-0.06,0.16]	-0.01 [-0.14,0.11]
(E) Extraversion	0.06 [-0.06,0.18]	0.06 [-0.06,0.18]	0.09 [-0.02,0.21]	0.06 [-0.06,0.18]	0.02 [-0.09,0.13]	0.10 [-0.02,0.22]	0.06 [-0.07,0.18]	0.12 [0.03,0.22]
(E1) Friendliness	0.13 [0.00,0.25]	0.12 [0.02,0.23]	0.16 [0.05,0.27]	0.13 [0.00,0.26]	0.08 [-0.04,0.20]	0.07 [-0.04,0.19]	0.13 [0.01,0.26]	0.07 [-0.04,0.19]
(E2) Sociableness	0.11 [0.00,0.22]	0.11 [0.00,0.22]	0.16 [0.04,0.28]	0.10 [-0.01,0.22]	0.03 [-0.08,0.15]	0.13 [0.03,0.24]	0.11 [0.00,0.22]	0.15 [0.03,0.27]
(E3) Assertiveness	0.06 [-0.06,0.18]	-0.11 [-0.22,0.00]	-0.01 [-0.12,0.09]	0.07 [-0.05,0.20]	-0.04 [-0.16,0.07]	0.10 [0.00,0.20]	0.05 [-0.06,0.17]	0.17 [0.07,0.28]
(E4) Dynamism	0.08 [-0.03,0.19]	-0.03 [-0.14,0.09]	0.04 [-0.08,0.17]	0.08 [-0.03,0.19]	0.05 [-0.06,0.16]	0.13 [0.02,0.24]	0.07 [-0.03,0.18]	0.15 [0.03,0.26]
(E5) Adventurousness	0.05 [-0.07,0.17]	0.15 [0.04,0.26]	0.04 [-0.07,0.15]	0.05 [-0.07,0.17]	0.00 [-0.12,0.12]	0.10 [0.00,0.21]	0.05 [-0.07,0.17]	0.09 [-0.03,0.21]
(E6) Cheerfulness	0.01 [-0.12,0.15]	-0.02 [-0.15,0.10]	-0.07 [-0.18,0.04]	0.02 [-0.11,0.15]	0.01 [-0.09,0.11]	-0.06 [-0.18,0.06]	0.03 [-0.11,0.16]	0.07 [-0.03,0.17]
(A) Agreeableness	0.04 [-0.07,0.15]	0.12 [-0.01,0.24]	0.09 [-0.02,0.20]	0.03 [-0.08,0.14]	-0.03 [-0.14,0.07]	0.02 [-0.10,0.14]	0.03 [-0.08,0.15]	0.01 [-0.11,0.13]
(A1) Willingness to trust	0.11 [-0.01,0.24]	0.06 [-0.03,0.16]	0.08 [-0.04,0.20]	0.12 [-0.01,0.24]	0.14 [0.04,0.25]	0.10 [0.00,0.21]	0.10 [-0.02,0.23]	0.09 [-0.03,0.20]

Table 3.A5 (continued)

Personality Dimension	Full Predictor Set	Habitual Listening Behavior	Audio Characteristics	Lyrics Characteristics	Lyrics Emotionality	Lyrics Topics	Lyrics Word Embeddings	Other Lyrics Characteristics
(A2) Genuineness	0.10 [-0.01,0.22]	0.14 [0.02,0.25]	0.18 [0.08,0.28]	0.09 [-0.02,0.21]	0.06 [-0.05,0.16]	0.09 [-0.03,0.20]	0.10 [-0.02,0.21]	0.10 [-0.01,0.21]
(A3) Helpfulness	-0.05 [-0.16,0.07]	0.06 [-0.06,0.17]	0.02 [-0.10,0.13]	-0.06 [-0.17,0.06]	-0.02 [-0.14,0.09]	-0.02 [-0.12,0.08]	-0.05 [-0.17,0.06]	-0.06 [-0.18,0.05]
(A4) Obligingness	0.00 [-0.12,0.11]	0.18 [0.07,0.29]	0.01 [-0.11,0.13]	-0.04 [-0.17,0.08]	-0.08 [-0.20,0.03]	0.01 [-0.12,0.14]	-0.04 [-0.16,0.08]	-0.08 [-0.20,0.03]
(A5) Modesty	-0.08 [-0.18,0.02]	0.03 [-0.09,0.14]	-0.03 [-0.15,0.09]	-0.06 [-0.17,0.04]	0.00 [-0.11,0.11]	-0.07 [-0.18,0.04]	-0.06 [-0.17,0.05]	-0.05 [-0.17,0.08]
(A6) Good naturedness	0.06 [-0.06,0.18]	0.09 [-0.02,0.21]	0.11 [-0.02,0.23]	0.05 [-0.07,0.16]	0.01 [-0.11,0.13]	0.03 [-0.07,0.14]	0.05 [-0.07,0.17]	0.08 [-0.04,0.20]
(ES) Emotional Stability	-0.06 [-0.19,0.07]	0.01 [-0.10,0.11]	0.04 [-0.07,0.15]	-0.06 [-0.19,0.06]	-0.08 [-0.20,0.04]	-0.10 [-0.23,0.02]	-0.06 [-0.19,0.06]	0.12 [0.02,0.22]
(ES1) Carefreeness	0.06 [-0.05,0.18]	-0.02 [-0.12,0.08]	0.09 [-0.02,0.19]	0.05 [-0.07,0.17]	-0.09 [-0.20,0.03]	-0.08 [-0.18,0.03]	0.06 [-0.07,0.18]	0.07 [-0.03,0.17]
(ES2) Equanimity	-0.01 [-0.10,0.08]	0.08 [-0.04,0.19]	0.04 [-0.08,0.16]	-0.01 [-0.11,0.08]	0.01 [-0.09,0.12]	-0.07 [-0.18,0.04]	-0.01 [-0.10,0.08]	0.02 [-0.09,0.13]
(ES3) Positive mood	-0.03 [-0.13,0.07]	-0.06 [-0.18,0.05]	-0.04 [-0.14,0.07]	-0.03 [-0.13,0.07]	-0.01 [-0.12,0.10]	-0.06 [-0.17,0.04]	-0.02 [-0.12,0.09]	0.05 [-0.06,0.16]
(ES4) Self consciousness	0.08 [-0.05,0.20]	0.03 [-0.07,0.13]	0.02 [-0.08,0.13]	0.08 [-0.04,0.20]	0.01 [-0.09,0.11]	0.12 [0.02,0.22]	0.08 [-0.04,0.19]	0.09 [-0.02,0.20]
(ES5) Self control	-0.11 [-0.23,0.01]	0.10 [-0.03,0.22]	-0.13 [-0.24,-0.02]	-0.10 [-0.22,0.02]	-0.03 [-0.16,0.09]	0.05 [-0.07,0.17]	-0.10 [-0.23,0.02]	-0.03 [-0.14,0.08]
(ES6) Emo. robustness	-0.04 [-0.16,0.07]	-0.12 [-0.24,0.00]	-0.06 [-0.17,0.04]	-0.03 [-0.15,0.08]	-0.07 [-0.17,0.04]	-0.04 [-0.16,0.08]	-0.02 [-0.13,0.10]	0.15 [0.03,0.26]

Note. Prediction performance when using all music-listening variables (i.e., the Full Predictor Set) vs. each group of variables separately for predicting personality scales. One benchmark was conducted for each personality outcome that was previously predicted with a minimum accuracy $r_s \geq .10$ based on the full predictor set (see Table 3.2). The Spearman rank order correlation (r_s) between predicted and measured personality scores was determined for each of the 100 resampling iterations of the cross-validation scheme (10x10 CV). This table provides means and 95% variance-corrected t-confidence intervals (CIs) across iterations. Please note, that CIs are purely for descriptive purposes and that they do not consider multiple testing. The group of "Lyrics Characteristics" comprised the four lower-level groups of Lyrics' Emotionality, Topics, Word Embeddings, and other Lyrics Characteristics (see Table 3.1).

3.9 References

Ali, S. O., & Peynircioğlu, Z. F. (2006). Songs and emotions: Are lyrics and melodies equal partners? *Psychology of Music*, *34*(4), 511-534. https://doi.org/10.1177/0305735606067168

Anderson, C. A., Carnagey, N. L., & Eubanks, J. (2003). Exposure to violent media: The effects of songs with violent lyrics on aggressive thoughts and feelings. *Journal of Personality and Social psychology, 84*(5), 960-971. https://doi.org/10.1177/0093650203258281

Anderson, I., Gil, S., Gibson, C., Wolf, S., Shapiro, W., Semerci, O., & Greenberg, D. M. (2021). "Just the way you are": Linking music listening on Spotify and personality. *Social Psychological and Personality Science*, *12*(4), 561-572. https://doi.org/10.1177/1948550620923228

Apley, D. W., & Zhu, J. (2020). Visualizing the effects of predictor variables in black box supervised learning models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *82*(4), 1059-1086. https://doi.org/10.1111/rssb.12377

Arendasy, M. (2009). *BFSI: Big-Five Struktur-Inventar (Test & Manual).* SCHUHFRIED GmbH.

Au, Q. (2020). *fxtract: Feature Extraction from Grouped Data* (Version 0.9.4.) [R package]. https://rdrr.io/cran/fxtract/

Aucouturier, J. J., & Pachet, F. (2003). Representing musical genre: A state of the art. *Journal of New Music Research, 32*(1), 83-93. https://doi.org/10.1076/jnmr.32.1.83.16801

Bailey, M. (2019). *NRCLex* (Version 3.0.0) [Python package]. https://pypi.org/project/NRCLex/

Bansal, J., Flannery, M. B. & Woolhouse, M. H. (2021). Influence of personality on music-genre exclusivity. *Psychology of Music, 49*(5), 1356-1371. https://doi.org/10.1177/0305735620953611

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science*, *2*(4), 396-403. https://doi.org/10.1111/j.1745-6916.2007.00051.x

Bello, P., & Garcia, D. (2021). Cultural divergence in popular music: The increasing diversity of music consumption on Spotify across countries. *Humanities and Social Sciences Communications*, *8*(1), 1-8. https://doi.org/10.1057/s41599-021-00855-1

Bertin-Mahieux, T. (2011, March 21). Lyrics to bow. *Million Song Dataset, GitHub repository*. https://github.com/tbertinmahieux/MSongsDB/blob/master/Tasks_Demos/Lyrics/ lyrics_to_bow.py

Bertin-Mahieux, T., Ellis, D. P.W., Whitman, B., & Lamere, P. (2011). The Million Song Dataset. In A. Klapuri, & C. Leider (Eds.), *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011)* (pp. 1-6). University of Miami. https://doi.org/10.7916/D8NZ8J07

Besson, M., Faita, F., Peretz, I., Bonnel, A. M., & Requin, J. (1998). Singing in the brain: Independence of lyrics and tunes. *Psychological Science*, *9*(6), 494-498. https://doi.org/10.1111/1467-9280.00091

Billboard. (2019, September 12). *Weekly time spent listening to music in the United States from 2015 to 2019 (in hours)*. Statista. https://www.statista.com/statistics/828195/time-spent-music/

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research, 3*(1), 993-1022.

Bleidorn, W., Hopwood, C. J., Back, M. D., Denissen, J. J., Hennecke, M., Jokela, M., Kandler, C., Lucas, R. E., Luhmann, M., Orth, U., Roberts, B. W., Wagner, J., Wrzus, C., & Zimmermann, J. (2020). Longitudinal experience–wide association studies - A framework for studying personality change. *European Journal of Personality, 34*(3), 285-300. https://doi.org/10.1002/per.2247

Bonnel, AM., Faita, F., Peretz, I., & Besson, M. (2001). Divided attention between lyrics and tunes of operatic songs: Evidence for independent processing. *Perception & Psychophysics*, *63*(7), 1201-1213. https://doi.org/10.3758/BF03194534

Bonneville-Roussy, A., Rentfrow, P. J., Xu, M. K., & Potter, J. (2013). Music through the ages: Trends in musical engagement and preferences from adolescence through middle adulthood. *Journal of Personality and Social Psychology, 105*(4), 703. https://doi.org/10.1037/a0033770

Bouckaert, R. R., & Frank, E. (2004). Evaluating the replicability of significance tests for comparing learning algorithms. In H. Dai, R. Srikant, & C. Zhang (Eds.), *Lecture notes in computer science: Vol. 3056. Advances in knowledge discovery and data mining* (pp. 3–12). Springer. https://doi.org/10.1007/978-3-540-24775-3_3

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5-32. https://doi.org/10.1023/A:1010933404324

Buss, D. M. (1987). Selection, evocation, and manipulation. *Journal of Personality and*

Social Psychology, 53(6), 1214-1221. https://doi.org/10.1037/0022-3514.53.6.1214

Casalicchio, G., Molnar, C., & Bischl, B. (2019). Visualizing the feature importance for black box models. In M. Berlingerio, F. Bonchi, T. Gärtner, N. Hurley, G. Ifrim (Eds.), *Lecture notes in computer science: Vol. 11051. Machine learning and knowledge discovery in databases* (pp. 655-670). Springer. https://doi.org/10.1007/978-3-030-10925-7_40

Chamorro-Premuzic, T., Fagan, P., & Furnham, A. (2010). Personality and uses of music as predictors of preferences for music consensually classified as happy, sad, complex, and social. *Psychology of Aesthetics, Creativity, and the Arts, 4*(4), 205-213. https://doi.org/10.1037/a0019210

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J. L., & Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. In Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, & A. Culotta (Eds.), *Advances in Neural Information Processing Systems (Vol. 22)* (pp. 288-296). Curran Associates, Inc.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159. https://doi.org/10.1037/0033-2909.112.1.155

Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology, 66*, 4-19. https://doi.org/10.1016/j.jesp.2015.07.006

Cronbach, L. J., & Gleser, G. C. (1957). *Psychological tests and personnel decisions.* University of Illinois Press.

DeepL GmbH. (n.d.). *DeepL API* [computer software]. https://www.deepl.com/de/docs-api/

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). *Bert: Pre-training of deep bidirectional transformers for language understanding.* arXiv. https://doi.org/10.48550/arXiv.1810.04805

DeYoung, C. G. (2015). Openness/intellect: A dimension of personality reflecting cognitive exploration. In M. Mikulincer, P. R. Shaver, M. L. Cooper, & R. J. Larsen (Eds.), *APA handbook of personality and social psychology: Vol. 4. Personality processes and individual differences* (pp. 369–399). American Psychological Association. https://doi.org/10.1037/14343-017

Dobrota, S., & Reić Ercegovac, I. (2015). The relationship between music preferences of different mode and tempo and personality traits – Implications for music pedagogy. *Music Education Research, 17*(2), 234-247. https://doi.org/10.1080/14613808.2014.933790

Flannery, M. B., & Woolhouse, M. H. (2021). Musical Preference: Role of personality and music-related acoustic features. *Music & Science*, *4*, https://doi.org/10.1177/20592043211014014

Fleeson, W. (2001). Toward a structure-and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology, 80*(6), 1011-1027. https://doi.org/10.1037/0022-3514.80.6.1011

Fricke, K. R., Greenberg, D. M., Rentfrow, P. J., & Herzberg, P. Y. (2018). Computer-based music feature analysis mirrors human perception and can be used to measure individual music preference. *Journal of Research in Personality, 75*, 94-102. https://doi.org/10.1016/j.jrp.2018.06.004

Fricke, K. R., & Herzberg, P. Y. (2017). Personality and self-reported preference for music genres and attributes in a German-speaking sample. *Journal of Research in Personality*, *68*, 114-123. https://doi.org/10.1016/j.jrp.2017.01.001

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software, 33*(1), 1-22.

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science, 2*(2), 156-168. https://doi.org/10.1177/2515245919847202

Genius. (2022). Getting Started, *Genius*. https://docs.genius.com/#/getting-started-h1

Götz, F. M., Stieger, S., & Reips, U. D. (2017). Users of the main smartphone operating systems (iOS, Android) differ only little in personality. *PloS ONE, 12*(5), e0176921. https://doi.org/10.1371/journal.pone.0176921

Goldberg, L. R. (1990). An alternative" description of personality": The big-five factor structure. *Journal of Personality and Social Psychology, 59*(6), 1216-1229. https://doi.org/10.1037/0022-3514.59.6.1216

Greenberg, D. M., Kosinski, M., Stillwell, D. J., Monteiro, B. L., Levitin, D. J., & Rentfrow, P. J. (2016). The song is you: Preferences for musical attribute dimensions reflect personality. *Social Psychological and Personality Science*, *7*(6), 597-605. https://doi.org/10.1177/1948550616641473

Greenberg, D. M., Matz, S. C., Schwartz, H. A., & Fricke, K. R. (2020). The self-congruity effect of music. *Journal of Personality and Social Psychology, 121*(1), 137-150. https://doi.org/10.1037/pspp0000293

Greenberg, D. M., & Rentfrow, P. J. (2017). Music and big data: A new frontier. *Current Opinion in Behavioral Sciences, 18*, 50-56. https://doi.org/10.1016/j.cobeha.2017.07.007

Greenberg, D. M., Wride, S. J., Snowden, D. A., Spathis, D., Potter, J., & Rentfrow, P. J. (2022). Universals and variations in musical preferences: A study of preferential reactions to Western music in 53 countries. *Journal of Personality and Social Psychology, 122*(2), 286-309. https://doi.org/10.1037/pspp0000397

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, *466*(7302), 29-29. https://doi.org/10.1038/466029a

IFPI. (2019). Global Music Listening Report 2019. (n.p.). https://www.ifpi.org/resources/

IFPI. (2021). Engaging with Music. (n.p.). https://www.ifpi.org/resources/

Johnson, J. A. (2005). Ascertaining the validity of individual protocols from web-based personality inventories. *Journal of Research in Personality, 39*(1), 103-129. https://doi.org/10.1016/j.jrp.2004.09.009

Joulin, A., Grave, E., Boianowski, P., & Mikolov, T. (2016). *Bag of tricks for efficient text classification.* arXiv. https://doi.org/10.48550/arXiv.1607.04606

Keusch, F., Bähr, S., Haas, G. C., Kreuter, F., & Trappmann, M. (2020). Coverage error in data collection combining mobile surveys with passive measurement using apps: Data from a German national survey. *Sociological Methods & Research*. https://doi.org/10.1177/0049124120914924

Konrad, M. (2016, June 17). Creating a sparse document term matrix for topic modeling via lda. *WZB Data Science Blog*. https://datascience.blog.wzb.eu/2016/06/17/creating-a-sparse-document-term-matrix-for-topic-modeling-via-lda/

Lang, M., Binder. M., Richter. J., Schratz, P., Pfisterer, F., Coors, S., Au, Q., Casalicchio, G., Kotthoff, L., & Bischl, B. (2019). mlr3: A modern object-oriented machine learning framework in R. *Journal of Open Source Software, 4*(44), 1903-1905. https://doi.org/10.21105/joss.01903

Langmeyer, A., Guglhör-Rudan, A., & Tarnai, C. (2012). What do music preferences reveal about personality? *Journal of Individual Differences, 33*(2). https://doi.org/10.1027/1614-0001/a000082

Le Pennec, E., & Slowikowski, K. (2019). *ggwordcloud: A word cloud geom for 'ggplot2'* (Version 0.5.0) [R package]. https://CRAN.R-project.org/package=ggwordcloud

Lucas, R. E., & Donnellan, M. B. (2011). Personality development across the life span: Longitudinal analyses with a national sample from Germany. *Journal of Personality and Social Psychology, 101*(4), 847-861. https://doi.org/10.1037/a0024298

Lui, M. (2016). langid (Version 1.1.6) [Python package]. https://pypi.org/project/langid/

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149-174. https://doi.org/10.1007/BF02296272

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*(3), 437–455. https://doi.org/10.1037/a0028085

McCallum, A. K. (2002). *MALLET: A Machine Learning for Language Toolkit* (Version 2.0.8) [Python package]. http://mallet.cs.umass.edu

McCrae, R., R., & John, O., P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality, 60*(2), 175–215. https://doi.org/10.1111/j.1467-6494.1992.tb00970.x

McCrae, R. R., & Terracciano, A. (2005). Universal features of personality traits from the observer's perspective: Data from 50 cultures. *Journal of Personality and Social Psychology*, *88*(3), 547-561. https://doi.org/10.1037/0022-3514.88.3.547

McKew, J. (2020, May 29). Translating Text in Python - DeepL Translator. *Jack McKew's Blog*. https://jackmckew.dev/translating-text-in-python.html

Miles, J. (1976). Music. On *Rebel*. Decca.

Mohammad, S. & Turney, P. (2013). Crowdsourcing a word-emotion association lexicon, *Computational Intelligence*, 29 (3), 436-465. https://doi.org/10.1111/j.1467-8640.2012.00460.x

Molnar, C., Bischl, B., & Casalicchio, G. (2018). iml: An R package for interpretable machine learning. *Journal of Open Source Software, 3*(26). https://doi.org/10.21105/joss.00786

Nadeau, C., & Bengio, Y. (2003). Inference for the generalization error. *Machine Learning*, *52*(3), 239-281. https://doi.org/10.1023/A:1024068626366

Nave, G., Minxha, J., Greenberg, D. M., Kosinski, M., Stillwell, D., & Rentfrow, J. (2018). Musical preferences predict personality: Evidence from active listening and Facebook likes. *Psychological Science, 29*(7), 1145-1158. https://doi.org/10.1177/0956797618761659

Pargent, F., Schoedel, R., & Stachl, C. (2022). *Best Practices in Supervised Machine Learning: A Tutorial for Psychologists.* PsyArXiv. https://doi.org/10.31234/osf.io/89snd

Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., Ungar, L. H., & Seligman, M. E. (2015). Automatic personality assessment through social media language. *Journal of Personality and Social Psychology, 108*(6), 934-952. https://doi.org/10.1037/pspp0000020

Park, M., Thom, J., Mennicken, S., Cramer, H., & Macy, M. (2019). Global music streaming data reveal diurnal and seasonal patterns of affective preference. *Nature Human Behaviour*, *3*(3), 230-236. https://doi.org/10.1038/s41562-018-0508-z

Phan, L. V., & Rauthmann, J. F. (2021). Personality computing: New frontiers in personality assessment. *Social and Personality Psychology Compass, 15*(7), e12624. https://doi.org/10.1111/spc3.12624

Python Software Foundation. (2021). *Python: A dynamic, open source programming language* (Version 3.7.10). Python Software Foundation. https://www.python.org

Qiu, L., Chen, J., Ramsay, J., & Lu, J. (2019). Personality predicts words in favorite songs. *Journal of Research in Personality*, *78*, 25-35. https://doi.org/10.1016/j.jrp.2018.11.004

R Core Team (2022). R: A language and environment for statistical computing [Computer Software]. R Foundation for Statistical Computing. https://www.r-project.org

Rauthmann, J., F. (2021). Capturing interactions, correlations, fits, and transactions: A person-environment relations model. In J. F. Rauthmann (Ed.), *The handbook of personality dynamics and processes* (pp., 427-522). Elsevier.

Rehurek, R., & Sojka, R. (2010). Software framework for topic modelling with large corpora. In R. Witte, H. Cunningham, J. Patrick, E. Beisswanger, E. Buyko, U. Hahn, K. Verspoor, & A. R. Coden (Eds.), *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks* (pp. 45-50). University of Malta.

Rentfrow, P. J., Goldberg, L. R., & Levitin, D. J. (2011). The structure of musical preferences: A five-factor model. *Journal of personality and Social Psychology, 100*(6), 1139-1157. https://doi.org/10.1037/a0022406

Rentfrow, P. J., & Gosling, S. D. (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), 1236-1256. https://doi.org/10.1037/0022-3514.84.6.1236

Roberts, B. W., Bogg, T., Walton, K. E., Chernyshenko, O. S., & Stark, S. E. (2004). A lexical investigation of the lower-order structure of conscientiousness. *Journal of Research in Personality, 38*(2), 164–178. https://doi.org/10.1016/S0092-6566(03)00065-5

Schäfer, T., & Mehlhorn, C. (2017). Can personality traits predict musical style preferences? A meta-analysis. *Personality and Individual Differences, 116*, 265-273. https://doi.org/10.1016/j.paid.2017.04.061

Schoedel, R., Au, Q., Völkel, S. T., Lehmann, F., Becker, D., Bühner, M., Bischl., B., Hussmann, H., & Stachl, C. (2019). Digital footprints of sensation seeking. *Zeitschrift für Psychologie, 226*(4). https://doi.org/10.1027/2151-2604/a000342

Schoedel, R., Kunz, F., Bergmann, M., Bemmann, F., Bühner, M., & Sust, L. (2022). *Snapshots of Daily Life: Situations Investigated Through the Lens of Smartphone Sensing*. PsyArXiv. https://doi.org/10.31234/osf.io/f3htz

Schuwerk, T., Kaltefleiter, L. J., Au, J. Q., Hoesl, A., & Stachl, C. (2019). Enter the wild: Autistic traits and their relationship to mentalizing and social interaction in everyday life. *Journal of Autism and Developmental Disorders, 49*(10), 4193-4208. https://doi.org/10.1007/s10803-019-04134-6

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Seligman, M. E. P., & Ungar, L. H. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS ONE, 8*(9), e73791. https://doi.org/10.1371/journal.pone.0073791

Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on generality (COG): A proposed addition to all empirical papers. *Perspectives on Psychological Science*, *12*(6), 1123-1128. https://doi.org/10.1177/1745691617708630

Spotify. (2022). Get Track's Audio Features, *Spotify for Developers*. https://developer.spotify.com/documentation/web-api/reference/#/operations/get-several-audio-features

Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2011). Age differences in personality traits from 10 to 65: Big Five domains and facets in a large cross-sectional sample. *Journal of Personality and Social Psychology, 100*(2), 330. https://doi.org/10.1037/a0021717

Stachl, C., Au, Q., Schoedel, R., Gosling, S. D., Harari, G. M., Buschek, D., Völkel, S. T., Schuwerk, T., Oldemeier, M., Ullmann, T., Hussmann, H., Bischl, B., & Bühner, M. (2020). Predicting personality from patterns of behavior collected with smartphones. *Proceedings of the National Academy of Sciences (PNAS)*, *117*(30), 17680-17687. https://doi.org/10.1073/pnas.1920484117

Stachl, C., Hilbert, S., Au, J. Q., Buschek, D., De Luca, A., Bischl, B., Hussmann, H., & Bühner, M. (2017). Personality traits predict smartphone usage. *European Journal of Personality, 31*(6), 701-722. https://doi.org/10.1002/per.2113

Swann, W. B. (1987). Identity negotiation: Where two roads meet. *Journal of Personality and Social Psychology, 53*(6), 1038-1051. https://doi.org/10.1037/0022-3514.53.6.1038

Tarrant, M., North, A. C., & Hargreaves, D. J. (2000). English and American adolescents' reasons for listening to music. *Psychology of Music*, *28*(2), 166-173. https://doi.org/10.1177/0305735600282005

The Hugging Face Team. (2020). Transformers. *Hugging Face*. https://huggingface.co/docs/transformers/index

Van Berkel, N., Ferreira, D., & Kostakos, V. (2017). The experience sampling method on mobile devices. *ACM Computing Surveys, 50*(6), 1-40. https://doi.org/10.1145/3123988

Vuoskoski, J. K., & Eerola, T. (2011). The role of mood and personality in the perception of emotions represented by music. *Cortex, 47*(9), 1099-1106. https://doi.org/10.1016/j.cortex.2011.04.011

Ward, M. K., & Meade, A. W. (2023). Dealing with careless responding in survey data: Prevention, identification, and recommended best practices. *Annual Review of Psychology, 74*, 577-596. https://doi.org/10.1146/annurev-psych-040422-045007

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (Version 3.3.5) [R package]. Springer-Verlag. https://ggplot2.tidyverse.org

Wickham, H., François, R., Henry, L., & Müller, K. (2021). *dplyr: A Grammar of Data Manipulation* (Version 1.0.4) [R package]. https://CRAN.R-project.org/package=dplyr

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., ... & Rush, A. M. (2020). HuggingFace's Transformers: State-of-the-Art Natural Language Processing. In Q. Liu, & D. Schlangen (Eds.), *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing* (pp. 38-45). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.emnlp-demos.6

Wright, M., & Ziegler, A. (2017). Ranger: A fast implementation of random forests for high dimensional data in C++ and R, *Journal of Statistical Software, 77*(1), 1-17. https://doi.org/10.18637/jss.v077.i01

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science, 12*(6), 1100-1122. https://doi.org/10.1177/1745691617693393

Yentes, R. D., & Wilhelm, F. (2021). *careless: Procedures for computing indices of careless Responding* (Version 1.2.1) [R package].

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology, 67*(2), 301-320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

4 Study 2: The Role of Personality and Mood in Everyday Music Choices

This chapter has not yet been published but will eventually be submitted as "Sust, L., & Schoedel, R. Explaining Everyday Music Choice on Smartphones: The Role of Personality Traits and Mood States."

4.1 Abstract

Digitalization has created an unparalleled freedom of choice in music consumption, pronouncing inter- and intraindividual differences in everyday listening behavior. To shed light on the factors involved in natural music choices, the present study collected 1,631 music-listening events from 110 participants over 14 consecutive days using smartphones for both active and passive ambulatory assessments. More specifically, we obtained mobile-sensed music-listening records and experience-sampled mood states from participants' smartphones, as well as their Big Five personality traits via traditional surveys. Using multilevel regressions, we predicted momentary music preferences in terms of the musical valence and energy of played songs from enduring personality traits, concurrent mood states, and their respective interactions. As preregistered, we expected to replicate past empirical findings of trait- and state-congruent music choice and theory-based interaction effects. However, our models showed that personality and mood accounted only for a small fraction of variance in music choices, with only one significant effect indicating that people in more activated mood states chose more energetic songs. Beyond that, our models failed to show any congruency or interaction effects for chosen musical valence and energy. We discuss methodological differences between past and present studies as potential reasons for our lack of results and outline future avenues for research on everyday music choices, expanding on our smartphone-based assessment approach.

4.2 Introduction

From stationary record players to portable cassette-, CD-, or MP3-players, and now to Internet-based streaming on smartphones – technological advancements have revolutionized the way people consume music. This digitalization has gradually lifted restrictions on mobility and choice, allowing listeners to play any song anytime and anywhere (Bull, 2005; Krause & North, 2016; North et al., 2004). With mobile music listening, people can create "auditory bubbles" (Bull, 2005, p. 344) in any situation to focus on themselves and enjoy a very personal music experience (Heye & Lamont, 2010; Kuch & Wöllner, 2021). This new freedom has enhanced the quantity and diversity of music consumption within listeners (Datta et al., 2018), sparking psychologists' interest in the factors involved in daily music-choice behavior (e.g., Greb et al., 2019; Randall & Rickard, 2017). Research on the uses and gratifications of music indicates that people actively seek out songs that fulfill their individual needs (Gantz et al., 1978; Katz et al., 1973). In particular, music commonly serves listeners for the goals of self-expression and mood regulation (e.g., DeNora, 1999; Laiho, 2004; Lonsdale & North, 2011; Schäfer et al., 2013), suggesting that both personality traits and mood states may contribute to the music preferences people exhibit on a daily basis. Taking advantage of the digital era of music listening, the present study investigates natural listening records from smartphones to model momentary music choice in everyday life jointly from enduring personality traits and fluctuating mood states.

4.2.1 Personality Traits and Music Preferences

Personality science has long focused on interindividual differences in the music people prefer on average and how they relate to the stable characteristics of listeners. Trait theories suggest that individuals exhibit enduring dispositions that describe and explain how they differ in their thoughts, feelings, and behaviors over time and across situations, and that can be organized into a taxonomy such as the widely used Big Five Model of personality (e.g., Allport, 1927; McCrae & John, 1992).

According to theories of person-environment transactions, music-listening behavior may be understood as a navigation mechanism by which individuals create or modify auditory environments that reflect and reinforce aspects of their personality (Buss, 1987; Rauthmann, 2021; Swann, 1987). Thereby, different types of music seem to serve specific uses, which help listeners fulfill their personality-related needs (Chamorro-Premizic & Furnham, 2007; Vella & Mills, 2017). Empirical studies have repeatedly supported the idea of trait-congruent music choice by linking the Big Five personality domains to preferences for broad musical styles and granular audio characteristics that match the respective traits' conceptualization (see de Raad, 2000; Goldberg, 1990; John & Srivastava, 1999) and trait-typical uses of music (see Chamorro-Premizic & Furnham, 2007; Vella & Mills, 2017). For example, individuals who score high on the trait Openness, which is associated with creativity, curiosity, and intellect, often prefer sophisticated musical styles (e.g., Classic, Jazz, Anderson et al., 2021; Bonneville-Roussy et al., 2013; Greenberg et al., 2022; Nave et al., 2018; Rentfrow & Gosling, 2003) and songs with negative valence, slow tempo, and low danceability and energy (Anderson et al., 2021; Dobrota

& Reić Ercegovac, 2015; Ladinig & Schellenberg, 2012; Sust et al., 2023; Vuoskoski & Eerola, 2011). In a trait-congruent manner, this type of music seems to help more open listeners achieve an intellectually stimulating listening experience (Chamorro-Premuzic & Furnham, 2007; Vella & Mills, 2007). In contrast, the Big Five domain Conscientiousness, which is conceptualized by duty, task orientation, self-discipline, and obedience to norms, was often related to an aversion towards intense musical styles (e.g., Rock, Punk) and a preference for unpretentious music (e.g., Pop, Country; Anderson et al., 2021; Bonneville-Roussy et al., 2013; Greenberg et al., 2022; Rentfrow & Gosling, 2003) and emotionally positive songs (Anderson et al., 2021; Dobrota & Reić Ercegovac, 2015; Flannery & Woolhouse, 2017; Qiu et al., 2018). Such music is socially conformist and may foster productivity in highly conscientious people. Furthermore, the domain of Extraversion, characterized by positive affect, increased energy, and high sociability levels, was repeatedly related to an affinity for both unpretentious and contemporary (e.g., R&B, Rap) musical styles (Anderson et al., 2021; Bonneville-Roussy et al., 2013; Greenberg et al., 2022; Nave et al., 2018; Rentfrow & Gosling, 2003) and, accordingly, positive valence and fast tempo in music (Dobrota & Reić Ercegovac, 2015; Flannery & Woolhouse, 2021; Ladinig & Schellenberg, 2012; Qui et al., 2018; Sust et al., 2023; Vuoskoski & Eerola, 2011). Illustrating the concept of trait-congruence, songs with these characteristics may be particularly stimulating and suitable to support social interactions of extraverted individuals. For the trait Agreeableness, which is defined as being kind, trustful, and cooperative, past studies found associations with preferences for unpretentious music (Anderson et al., 2021; Bonneville-Roussy et al., 2013; Greenberg et al., 2022; Rentfrow & Gosling, 2003) and songs with positive valence and low energy (Anderson et al., 2021; Dobrota & Reić Ercegovac, 2015; Sust et al., 2023; Vuoskoski & Eerola, 2011). Such songs are widely popular and have a low potential for conflict, which suits the definition of Agreeableness (e.g., John & Srivastava, 1999). Finally, listeners high in the domain of Neuroticism, characterized by the experience of negative affect and emotional instability, often prefer intense music genres (Anderson et al., 2021) and songs with negative valence (Anderson et al., 2021; Dobrota & Reić Ercegovac, 2015; Qui et al., 2018; Vuoskoski & Eerola, 2011). This type of music appears to be most suitable to support mood regulation in those high in Neuroticism (Chamorro-Premuzic & Furnham, 2007; Vella & Mills, 2017).

While these individual studies have repeatedly linked overall music preferences to the Big Five domains of personality, a meta-analysis in this field revealed only small effect sizes and inconsistent patterns across studies (Schäfer & Mehlhorn, 2017), indicating that factors beyond personality contribute to listeners' preferences. Additionally, personality traits can only account for overall (i.e., average) music preferences, while momentary preferences (i.e., music choices) in everyday life vary not only between but also within individuals. Indeed, past studies have shown that between-person differences only explain 20% of the variance in daily music choices, suggesting that, beyond stable traits, fluctuating listener states play an important role in natural music-listening behavior (Greb et al., 2019; Greb, Steffens & Schlotz, 2018).

4.2.2 Mood States and Music Preferences

One state-level factor that could be linked to momentary music preferences on a moment-to-moment basis is listeners' mood. In contrast to fully-fledged emotions, mood refers to longer-lasting affective states of lower intensity that are most pronounced as a shift in subjective feelings but not necessarily accompanied by physiological responses (Gross, 1998; Larsen, 2000). According to Russell's (1980) circumplex model, mood (along with other affective states) is defined by a valence (i.e., pleasantness) and an arousal (i.e., activation) dimension which together determine affective categories like happiness (i.e., positive valence and high arousal) or sadness (i.e., negative valence and low arousal; Watson & Tellegen, 1985).

With its ability to express and elicit various affective states (e.g., Eerola & Vuoskoski, 2012; Juslin & Laukka, 2004; Lundqvist et al., 2009), music is a popular tool to modify or maintain the valence and arousal of listeners' mood states (e.g., DeNora, 1999; Laiho, 2004; Lonsdale & North, 2011; Saarikallio & Erkkilä, 2007; Schäfer et al., 2013; Sloboda & O'Neill, 2001). Hence, the music people choose to play may reflect their current mood, sparking researchers' interest in mood-dependent music choices. Thereby, past studies mainly focused on music-choice behavior during negative mood states positioned at the lower end of the valence dimension. Inspired by Mood Management Theory (Zillmann, 2015), researchers assumed that individuals in negative states would be driven to seek out pleasant experiences and would, therefore, choose songs with positive valence. From a mood regulation perspective (see Naragon-Gainey et al., 2017), such mood-incongruent music choices may act as a disengagement strategy (e.g., distraction, avoidance) if the music shifts listeners' attention away from the negative mood (Miranda & Claes, 2009; Saarikallio & Erkkilä, 2007; Sakka & Juslin, 2018). While a few studies found support for mood-incongruent music choices (Knobloch & Zillmann, 2002; Tahlier et al., 2013), the majority of studies in this field reported mood-congruent momentary preferences in negative mood states (e.g., Chen et al., 2007; DeMarco et al., 2015; Lee et al., 2013; Taruffi & Koelsch, 2014). Several studies also found mood-congruent music choices at both ends of the valence spectrum and for the arousal dimension (Friedman et al., 2012; Greb et al., 2019; Kinghorn, 2021; Randall & Rickard, 2017; Thoma et al., 2012; Xue et al., 2018; Yang & Liu, 2013). In light of Mood Management Theory, mood-congruent music choice is theoretically more challenging to explain because negative music can increase negative mood states (e.g., Hunter et al., 2010; ter Bogt et al., 2021) instead of maximizing pleasure (Zillmann, 2015). However, negative music may serve various mood-regulation functions (see Naragon-Gainey et al., 2017 for an overview of regulatory strategies). People in negative states may choose matching songs to enact adaptive engagement strategies such as problem-solving, acceptance, positive reappraisal, or as a proxy for social support (Chin & Rickard, 2012; Miranda & Claes, 2009; Saarikallio & Erkkilä, 2007; Sakka & Juslin, 2018). For example, the lyrics of negative-valence songs may provide information relevant to solving a distressing situation (Saarikallio & Erkkilä, 2007; Van den Tol & Edwards, 2015) or offer a sense of social sharing akin to interpersonal relationships (Lee et al., 2013; Taruffi & Koelsch, 2014; Van den Tol & Edwards, 2015). However, listening to mood-congruent negative music may also serve non-adaptive regulation strategies like venting or aversive cognitive perseveration (e.g., rumination), preventing listeners from disengaging from negative states (Miranda & Claes, 2009; Saarikallio & Erkkilä, 2007; Sakka & Juslin, 2018).

4.2.3 Personality as Moderator of Mood-Congruent Music Preferences

Because mood regulation provides a theoretical framework for both mood-congruent and -incongruent momentary music preferences (e.g., Miranda & Claes, 2009; Saarikallio & Erkkilä, 2007; Sakka and Juslin, 2018), the contradictory results of past studies on mood-based music choice presented above may be related to the use of different regulatory strategies. In turn, the different styles of mood regulation and, more specifically of coping (i.e., self-regulatory attempts to reduce stress; Lazarus, 1966), were previously associated with the Big Five personality traits (Agbaria & Mokh, 2022; Barańczuk, 2019; Carver & Connor-Smith, 2010; Connor-Smith & Flachsbart, 2007; Watson & Hubbard, 1996). According to these studies, individuals scoring high in the domains of Openness, Conscientiousness, Extraversion, and Agreeableness are more likely to use adaptive engagement strategies such as problem-solving, positive reappraisal, or seeking social support, which, as laid out above, may require mood-congruent music. In contrast, the personality domain Neuroticism seems to be positively related to employing disengagement strategies like distraction (Baranczuk, 2019; Carver & Connor-Smith, 2010; Connor-Smith & Flachsbart, 2007; Watson & Hubbard, 1996), which should call for mood-incongruent music choices. Following this rationale, personality traits may moderate the relationship between mood states and momentary music preferences.

4.2.4 Assessing Momentary Music Preferences

Music-listening behavior fluctuates dynamically throughout the day, making it methodologically challenging to assess music choices. Most past studies on overall or momentary music preferences relied on self-report questionnaires (e.g., Rentfrow & Gosling, 2003; Taruffi & Koelsch, 2014) or reactions to musical excerpts presented without context or after mood induction in laboratory settings (e.g., Chen et al., 2007; Flannery & Woolhouse, 2021; Greenberg et al., 2022; Knobloch & Zillman, 2002). However, self-reports of music preferences are prone to biases like socially desirable responding or memory limitations, which can be especially problematic when reporting on contextual variations of a high-frequency behavior like music choices (Baumeister et al., 2007; Stein et al., 2013). Preferences among musical excerpts, on the other hand, are restricted by the sample of songs provided by researchers, which is typically limited to very popular, artificially manipulated, or unreleased tracks and, hence, provides less choice than the natural music market (Greenberg & Rentfrow, 2017).

For an ecologically valid assessment of music preferences, researchers must investigate music choices as they naturally occur in listeners' everyday lives. While collecting behavioral data in the field has long been practically infeasible (see Furr, 2009), the digitalization of music consumption has turned digital devices like smartphones into the ideal tool to investigate music-listening behavior "in the wild." Besides radios, smartphones are the most widely used device for playing music (IFPI, 2019). Furthermore, they can collect real-time data on people's thoughts, feelings, behaviors, or environments through active and passive ambulatory assessments (see Conner & Mehl, 2015; SAA, 2018).

On the one hand, smartphones can administer the experience-sampling methodology and *actively* ask participants to fill out short and identical questionnaires on repeated occasions throughout the day (Larson & Csikzentmihalyi, 2014; van Berkel et al., 2017). Previous studies have implemented this form of in-situ self-report assessment to explore momentary music choices in relation to contextual factors like mood states (Greb et al., 2019; Kinghorn, 2021; Randall & Rickard, 2017). They repeatedly asked participants to rate the musical properties of the song they were currently listening to. These ES were either randomly triggered (Greb et al., 2019), which, however, is not very efficient as people are exposed to music only in 40% of randomly sampled moments throughout the day (Juslin et al., 2008; North et al., 2004; Sloboda et al., 2001), or whenever participants used a specially-developed music player app (Kinghorn, 2021; Randall & Rickard, 2017). However, this approach only captured participants' subjective experiences, which may not always align with the objective characteristics of their selected music because the perception of musical emotion, in turn, depends on listeners' personality traits and current mood states (Hunter et al., 2011; Vuoskoski & Eerola, 2011).

On the other hand, smartphones provide a more objective way to collect music-listening data in the field via smartphone sensing, that is, the *passive* collection of smartphone usage data via custom research apps (Harari et al., 2016, 2017; Wrzus & Mehl, 2015). Smartphone-sensing apps can access a phone's systems logs, including the music-listening records, and unobtrusively collect music choices in everyday life, serving as a digitally-mediated behavioral observation. In contrast to experience-sampled self-reports, smartphones' digital listening records provide continuous and more granular information on selected songs. In addition, tools from music information retrieval (see Downie, 2003) allow researchers to automatically represent songs from those listening records in terms of various intrinsic musical characteristics based on their audio recordings (e.g., Flannery & Woolhouse, 2021; Sust et al., 2023; Yang & Liu, 2013). These technical audio characteristics range from basic physical parameters (e.g., tempo, pitch) to more complex aggregated features (e.g., valence, energy) learned via machine-learning algorithms, which can represent the emotionality of music in a valid manner (e.g., Eerola et al., 2009; Laurier et al., 2009). While first studies have started to objectively assess and represent overall preferences displayed in natural music-listening behavior on smartphones (Stachl et al., 2020; Sust et al., 2023) or streaming platforms (Anderson et al., 2021), they considered only summary metrics, ignoring the potential of longitudinal listening records for uncovering intra-individual fluctuations in daily music choices.

4.2.5 The Present Study

In this naturalistic study, we investigated music choices made in everyday life and their relation to enduring personality traits and fluctuating mood states. We applied an intensive longitudinal sampling design and collected 1,631 music-listening events from 110 participants over 14 consecutive days. Using smartphone sensing, we obtained the natural music-listening records from participants' private smartphones. We extracted their momentary music preferences based on the songs they played at a given moment and represented their music choices in terms of the two computationally-derived audio characteristics of musical valence and energy obtained from Spotify.com. In addition, we administered event-triggered ES to capture the valence and arousal of participants' mood states during the respective music-listening events and an online survey to assess participants' Big Five personality traits. To account for the inter- and intraindividual fluctuations in these multi-method data, we analyzed them in a multilevel-regression framework and predicted both the valence and energy of momentary music choices from personality traits, mood states, and their respective interactions. Based on the theoretical reasoning and past empirical findings presented throughout our introduction, we tested the following preregistered assumptions for our music choice models.

Our first set of hypotheses concerned the role of personality traits. We expected to replicate the findings of trait-congruent overall music preferences for momentary music choices and translated the most consistent associations from the past to our two technical audio characteristics (**H1**). Specifically, we assumed that Openness would be negatively related to chosen musical valence (*H1.1a*) and musical energy (*H1.1b*), that Conscientiousness would be positively related to chosen musical valence (*H1.2a*) and negatively related to musical energy (*H1.2b*), that Extraversion should be positively related to chosen musical valence (*H1.3a*) and musical energy (*H1.3b*), that Agreeableness would be positively related to chosen musical valence (*H1.4a*) and negatively related to musical energy (*H1.4b*), and, finally, that Neuroticism would be negatively related to chosen musical valence (*H1.5a*) and positively related to chosen musical energy (*H1.5b*).

Our second research question concerned the role of mood states. We expected to replicate the mood-congruent music choices found in the majority (but not all) of past studies for both dimensions of mood and based on natural music-listening behavior (**H2**). In particular, we assumed that mood valence should be positively related to chosen musical valence (*H2a*) and that mood arousal would be positively related to chosen musical energy (*H2b*).

Finally, our third set of hypotheses targeted the interaction between enduring personality traits and momentary mood states. Based on mood-regulation research, we believed that personality traits would moderate the relationship between mood states and music choices in everyday life (**H3**). More specifically, we assumed that the relationship between mood valence and musical valence would be stronger (i.e., more positive) for those with higher levels of Openness (*H3.1a*), Conscientiousness[3] (H*3.2a*), Extraversion (H*3.3a*), and Agreeableness (H*3.4a*), but weaker (i.e., less positive or negative) for those higher in Neuroticism (*H3.5a*). Complementary to this, we assumed that the relationship between mood arousal and musical energy would be stronger (i.e., more positive) for those scoring higher in Openness (*H3.1b*), Conscientiousness[3] (H*3.2b*), Extraversion (H*3.3b*), and Agreeableness (H*3.4b*), but weaker (i.e., less positive or negative) for those higher in Neuroticism (*H3.5b*). By testing these

[3] Please note that the subset of hypotheses concerning Conscientiousness (H3.2a & H3.2b) was accidentally not preregistered, although we mentioned the domain in the corresponding rationale in our registration. We provide further details on all deviations from the preregistration in our OSF project.

hypotheses, our study aims to bring clarity to the conflicting literature on personality- and mood-based music choices based on ecologically valid real-life data.

4.3 Method

The present study was conducted within the interdisciplinary PhoneStudy research project at LMU Munich. It was part of a larger student project on another research question carried out during a seminar and a Bachelor thesis. Throughout this project, we collected various self-reported and behavioral data by combining three data collection modalities, namely online surveys, ES, and smartphone sensing. Not to go beyond the scope of this manuscript, we limit our report here to the procedures relevant to the present research question and give a full account of all collected measures in the online supplemental material (OSM) in our OSF repository https://osf.io/d25et.

All study procedures have received ethical approval and adhered to the General Data Protection Regulation. The hypotheses and procedures reported in this manuscript were preregistered before conducting data preprocessing or analysis under https://osf.io/7j5e3. We had to make some changes to our preregistered protocol to accommodate practical hurdles encountered during data preprocessing and communicate all deviations from the preregistration throughout the methods section and, in greater detail, in our OSF project. While the privacy-sensitive nature of the mobile-sensing data prevents us from sharing raw logging data, we provide a dataset of aggregated variables and all code for preprocessing and multilevel analyses in our OSF project.

4.3.1 Procedures

Our data collection took place between May and November 2020 in Munich, Germany. We recruited participants with the help of students during the course of a seminar and a thesis, using university mailing lists, social media, and personal contacts. To be eligible for participation, subjects had to be over 18 years old, be fluent in German, and, for technical reasons, be the sole user of a smartphone running on the Android operating system. As for compensation, participants received an individual personality profile in addition to either 10 Euro or 4h of course credit, unless they decided to donate their data.

Participants were first invited to an onboarding survey where they received information about the aim and scope of the study. After providing informed consent during the onboarding, they installed our Android-based sensing app PhoneStudy on their private smartphones and went through a second round of informed consent within the app (i.e., granting permission access). For the following 14 days, the PhoneStudy app unobtrusively logged various smartphone usage data, including participants' music-listening records (see Chapter S1 of our OSM for details). In addition, the app administered ES, asking participants to report their current mood states. ES were scheduled in an event-triggered manner and appeared with a five-second delay each time participants opened a music app – as defined by an app categorization by Stachl et al. (2020) – on their smartphones. We chose this procedure to create a timely contingency between mood reports and music-listening behavior. In a concluding reactivity check, participants reported, on a scale from 1 to 5, that their music-listening behavior had been *not at all* to *barely* influenced by the event-triggered ES on average (M = 1.38, SD = 0.62). Additional ES were administered each morning, but these were not relevant to the current research question (see Chapter S1 of OSM for details). Beyond these two forms of ambulatory assessment, participants filled out two online surveys – one at the beginning and one at the end of the 14-day study period – which included demographic questions, a personality inventory, and other psychological measures that we list in our OSM (see Table S1.2).

4.3.2 Sample

Formal power estimation in multilevel data structures (e.g., where ES are nested within persons) requires an a priori estimation of fixed and random (co-)variance components. Because these parameters were difficult to determine given the unexplored nature of natural music-listening behavior (particularly at the state level), we aimed to collect the largest possible sample size given the time constraints of the corresponding student research project, lasting from May to November 2020.

The resulting convenience sample initially contained mobile-sensing data from 476 participants. However, not all of them had listened to music on their smartphones and participated in our ES (either due to non-compliance or a misconception of our sampling schedule[4]), as discerned in Table 4.A1 in the Appendix. Hence, during pre-processing, we had to remove 363 participants with fewer than four valid *music-listening events*. We defined valid music-listening events as completed ES instances (i.e., where both mood items were answered) surrounded by a 30-minute window where a) music was played for at least one minute and b) at least one played song had available song-level information from Spotify.com (see our section on "Data Enrichment" for details). Furthermore, we removed two participants with zero

[4] The PhoneStudy app triggered an ES whenever participants opened a music app on their smartphone. However, many music apps provide a banner with controls (e.g., << ▶ >>) on the smartphone's lock screen, which allowed participants to start, skip, and stop songs without actively opening the corresponding music app and, hence, without triggering an ES.

variance in either of our two mood items across the ES and one participant who had not completed the personality measure. These exclusions resulted in a final sample of 110 participants with sufficient data in all measures relevant to our hypotheses. Please note that we adapted some of our preregistered exclusion criteria to preserve a reasonably large dataset without risking a loss of data quality. We report more details on the data exclusion pipeline in the Appendix (see Table 4.A1).

The final sample comprised 75 women (67%) and 35 men (33%). Participants' age ranged between 18 and 57, with an average of 23 years ($SD = 6.5$), and the sample was skewed towards better education (78% with A-levels and 16% with a university degree), which reflects the university context of our data acquisition. The demographic variables served for descriptive purposes only in this study.

4.3.3 Impact of COVID-19

Because our data collection took place during the COVID-19 pandemic, we checked the containment measures implemented by the German government during the corresponding time period to get an impression of possible implications on the collected data. We inspected the so-called Stringency Index, a composite measure that subsumes different restriction indicators, such as workplace closures, prohibition of public events, and travel bans (Hale et al., 2021). The index shows that restrictions in Germany were relatively loose during our study period between May (22nd) and November (5th) 2020, compared to the onset and later (winter) stages of the pandemic (see the dashboard presentation by Mathieu et al., 2020). However, the Stringency Index – with a range of 0 (= *no restriction*) to 100 (= *full restriction)* – still exhibited values from 50 to 63, which are much higher than the value of 15 obtained in December 2022 when life was mostly back to normal. Thus, we cannot rule out that the COVID-19 restrictions in place still affected daily life during our study. For example, restrictions such as *working (or studying) remotely* or *meeting only a limited number of persons at the same time in public* (see Steinmetz et al., 2022) might have affected mobility patterns and socializing behavior and, in turn, changed the occasions for music listening on the smartphone. Alternatively, the restrictions (or the pandemic background itself) may have had a (dampening) effect on participants' mood states (see Ammar et al., 2020; Charles et al., 2021; Zacher & Rudolph, 2020). In sum, the pandemic situation potentially influenced our state-level data prompting us to interpret our findings against the pandemic background in the Discussion.

4.3.4 Measures

4.3.4.1 Self-Report Measures

Personality Traits. We used the German adaption (Rammstedt et al., 2020) of the short form of the Big Five Inventory-2 (BFI-2-S; Soto and John, 2017) to assess participants' personality traits during the survey at the beginning of our study. The BFI-2-S captures the Big Five domains of personality (i.e., *Openness (or Open-Mindedness), Conscientiousness, Extraversion, Agreeableness,* and *Neuroticism (or Negative Emotionality)*) with six items each (i.e., 30 items total). The items comprise short self-descriptive phrases (e.g., "I am full of energy and drive."), where agreement is indicated on a 5-point Likert scale ranging from *strongly disagree* (1) to *strongly agree* (5). For each personality domain, ratings were averaged across its six items, with higher scores indicating higher trait levels. We report confidence intervals of internal consistencies for the domain scores obtained in our sample in Table 4.1. Consistencies for the domains Extraversion and Agreeableness were rather low with McDonald's omega point estimates of .67 and .73, respectively, which, however, is not surprising given the short questionnaire length. Overall, all consistency estimates obtained in our study were in the same range or slightly above those reported by Rammstedt and colleagues (2020).

Table 4.1

Descriptive Statistics of State- and Trait-Level Measures

	M	*SD*	min	max	ICC	ω [CI$_{95\%}$]
Music Choice						
Musical Valence	0.47	0.18	0.03	0.97	.22	-
Musical Energy	0.65	0.17	0.00	0.99	.32	-
Mood States						
Mood Valence	4.69	0.97	1.00	6.00	.29	-
Mood Arousal	4.09	1.24	1.00	6.00	.21	-
Personality Traits						
Openness	3.72	0.75	1.83	5.00	-	.81 [.74, .86]
Conscientiousness	3.53	0.67	1.67	4.83	-	.80 [.73, .85]
Extraversion	3.14	0.62	1.50	4.67	-	.67 [.55, .78]
Agreeableness	3.86	0.60	2.50	5.00	-	.73 [.64, .80]
Neuroticism	2.96	0.87	1.17	5.00	-	.86 [.80, .90]

Note. N_{L1} = 1,631 observations from N_{L2} = 110 participants. Music choice was obtained from music-listening behavior on smartphones and coded on a continuous scale from 0 to 1. Mood states and personality traits were assessed via self-reports on a five-point (personality) and six-point (mood) five- Likert-type response scale. For both musical and mood valence, higher values indicate more positive valence. Intra-class correlation coefficients (ICCs) reflect the proportion of variance of state-level measures attributable to their grouping within persons. The reliability coefficient ω refers to McDonald's omega total, calculated with the MBESS package (Kelley, 2016). The square brackets contain the 95% confidence intervals (CI) for omega coefficients.

Mood States. The PhoneStudy app captured participants' mood states via ES. In line with previous studies (e.g., Kushlev & Heintzelman, 2018; Schoedel et al., 2023), we used two single-item measures to assess participants' mood in terms of valence and arousal – the two dimensions of the circumplex model of affect (Russell, 1980). The items asked participants to report their current emotionality and level of activity at the time of the ES. Responses were made on a bipolar six-point Likert scale, ranging from *very negative* (1) to *very positive* (6) for valence and from *very inactive* (1) to *very activated* (6) for arousal.

4.3.4.1 Behavioral Music-Listening Measures

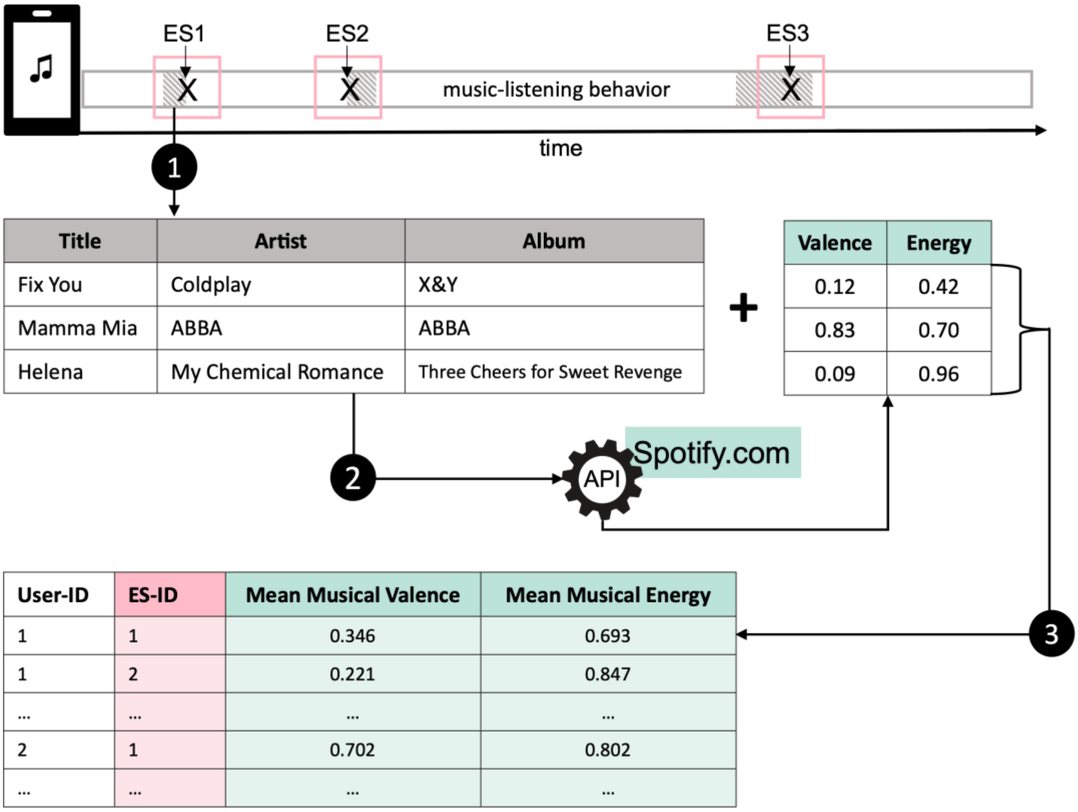
The PhoneStudy app provided mobile-sensing data on a wide range of smartphone usage behaviors (see Table S1.1 for an overview), including participants' music-listening records. The app created time-stamped logs whenever participants listened to locally stored or streamed music on their smartphones. To extract momentary music choice variables from these music logs, we administered a sophisticated preprocessing pipeline (see Figure 4.1).

Data Enrichment. The sensed music logs specified the title, artist, and album of participants' played songs, but lacked psychologically meaningful information about their intrinsic musical attributes. Hence, to describe participants' song choices, we enriched the raw music logs with song-level information using Spotify's Track API[5]. We visualize this workflow in steps 1 and 2 of Figure 4.1. For each song, we retrieved two audio characteristics that Spotify derived computationally from the respective song's audio recording (Spotify, 2023). These audio characteristics reflected the musical valence and energy of the songs in our music logs. According to Spotify, the musical valence captures the positiveness conveyed by a song, whereby songs with values closer to 1.0 sound more positive (e.g., cheerful), and songs with values closer to 0.0 sound more negative (e.g., sad, angry). In contrast, the musical energy represents the perceived intensity of a song, whereby values range from 0.0 to 1.0 and songs with higher values sound faster, louder, and noisier. We assigned both audio characteristics to the respective songs in our music logs but were not able to enrich all entries because some contained non-musical tracks (e.g., podcasts), had incorrect song information (e.g., typos in the song title), or were not covered by Spotify.com. We provide further details on the song-level data enrichment in Chapter S3 of our OSM.

[5] https://developer.spotify.com/documentation/web-api/reference/#/operations/get-several-audio-features

Figure 4.1

Preprocessing Workflow for Extracting Momentary Music Choice Variables from Smartphone-Sensed Music-Listening Records



Note. The PhoneStudy app logged song records whenever music was listened to and administered event-triggered experience samplings (ES) whenever a music app was opened. We enriched the raw song records with two song-level variables via the Spotify Tracks Application Programming Interface (API). The exemplary songs in the grey table demonstrate the face validity of the audio characteristics valence and energy (green table), whereby for musical valence, higher values represent more positive-sounding songs. More details on the enrichment are available in Chapter S3 of our OSM. Finally, we aggregated the song-level variables across all songs listened to within a 30-minute window surrounding an ES instance to represent momentary music choice in relation to the ES (bottom table).

Variable Extraction. To capture participants' momentary music choices in a timely contingency to their self-reported mood states, the song-level enriched music logs had to be matched with the ES instances. As the PhoneStudy app triggered ES questionnaires whenever participants opened a music app, we had preregistered to aggregate music choice (i.e., the audio characteristics of played songs) over a 30-minute window *following* the event-triggered ES. However, this rationale failed because participants often started songs through the controls banner on their smartphone's lock-screen[4], not opening their music app at all or only later to search for a specific song or to stop the music. Hence, many ES instances were preceded but not followed by music-listening behavior. To accommodate this data structure, we adapted our preregistered extraction strategy and defined music-listening events as 30-minute time windows *surrounding* an ES instance (i.e., 15 minutes before and after an ES). This definition was sufficiently broad to capture some music-listening behavior but narrow enough to still assume a timely contingency between music choices and self-reported mood states. Within these music-listening events, we removed songs played for less than 20 seconds (i.e., skipped songs) and aggregated the two audio characteristics over all unskipped songs using the weighted arithmetic mean based on the songs' playtime. The resulting variables represented participants' momentary music choices in terms of the musical valence and energy of their played songs. However, as noted above, Spotify's audio characteristics were not available for all tracks in our music logs, so the music choice variables covered only a portion of participants' played tracks. In 74% of music-listening events, the audio characteristics were available for all songs listened to, while in the remaining events, the availability of song-level data ranged between 8% and 94% of played songs (M = 69%, SD = 17%).

4.3.5 Data Analysis

For our regression analyses, we first adjusted extreme outliers (> |M +/- 3SD|) in our music choice variables to the value three standard deviations from the respective mean (see Winsorizing according to Ghosh & Vogt, 2012) to eliminate the influence of possible inaccuracies in our logging data. Apart from that, we did not exclude or adapt any outliers or influential cases.

To account for the hierarchical two-level structure of our data (i.e., music-listening events nested within participants), we applied multilevel regression modeling (MLM) to test our hypotheses (Bates, Mächler et al., 2015; Kuznetsova et al., 2017). As preregistered, we computed one model for each of the two audio characteristics representing participants' chosen songs (i.e., musical valence and energy). Both models simultaneously estimated the between-person effects of personality traits (H1), the within-person effects of mood states (H2), and the cross-level interaction effects between mood and personality (H3) on momentary music choices (see Barr et al., 2013; Bates, Kliegl et al., 2015). More precisely, the MLMs contained both dimensions of mood (i.e., valence and arousal) and their interaction as Level-1 predictors, the Big Five personality domains as Level-2 predictors, and the cross-level interaction between each personality domain and the mood state focal to the respective hypothesis (i.e., mood valence for the musical valence model in H3.1-5a, mood arousal for the musical energy model in H3.1-5b). Because mood states could manifest within as well as between persons, they were within-person centered, and their person means were reintroduced as additional Level-2 predictors as recommended by Enders and Tofighi (2007). All Level-2 predictors (i.e., personality traits and aggregated mood states) were grand-mean centered. After initially running random-intercept-random-slope models, we removed the random slopes to avoid problems of singular fit (i.e., variance estimates near zero; Bates, Kliegl et al., 2015). We provide our final model equations in Chapter S4 of our OSM.

To estimate the effect size of the combined predictors, we determined the models' marginal $R^2_{(m)}$, which indicates the proportion of the criterium variance explained by all fixed effects (Rights & Sterba, 2019). Furthermore, we computed fully standardized versions of the two MLMs specified above to obtain standardized regression coefficients as effect size estimates of the single predictors (Lorah, 2018). In these models, we z-standardized all variables and – after standardizing – person-mean centered Level-1 predictors.

Beyond the preregistered MLMs, we additionally computed beta-distributed generalized linear mixed models to account for the bounded scale of our criteria variables, ranging from 0.00 to 1.00 (Fournier et al., 2012). The models were run with the same specifications as our unstandardized MLMs.

We applied the conservative level of α = .005 to determine the significance of our hypotheses. We derived this alpha level by correcting the default of .05 for multiple testing. Applying the Bonferroni correction, we divided alpha by the maximum number of tests necessary for any of our higher-order hypotheses, which were 10 tests for H1 and H3. Because we had preregistered the directionality of our hypotheses, we used one-tailed p-values to determine the statistical significance of the predicted effects[6]. P-values for all other effects that were not part of our hypotheses were purely exploratory and, hence, reported in a two-tailed manner. The p-values presented in our results tables below are tagged accordingly.

4.3.6 Statistical Software

The API call from Spotify.com was conducted in Python (version 3.8.6, Python Software Foundation, 2021), while all other steps of analysis were conducted in the statistical software R – version 4.2.1 for data preprocessing on an RStudio Server and version 4.1.2 for descriptive and multilevel analysis in a local R environment (R Core Team, 2022). For preprocessing our raw logging data on the RStudio Server, we used the packages *tidyr* (version 1.2.1, Wickham et al., 2022) and *dbplyr* (version 2.2.1, Wickham et al., 2023). We also provide an Excel file listing all R packages installed on the RStudio server in our OSF project. For the locally conducted multilevel modeling, we employed the packages *lme4* (version 1.1-31, Bates,

[6] Only the interaction effects regarding Conscientiousness were tested with two-tailed p-values because the corresponding hypotheses (H3.2a & H3.2b) were not formally preregistered by accident (see footnote 3).

Mächler et al., 2015) and *lmerTest* (3.1-3, Kuznetsova et al., 2017) as well as *r2mlm* (version 0.3.2, Shaw et al., 2020) to determine marginal squared *R*s and *glmmADMB* (version 0.8.3.3, Fournier et al., 2012) to run the beta-distributed models. For reproducibility, we used the package management tool *renv* (version 0.16.0, Ushey, 2022) for our local data analyses and provided a complete list of all installed R packages in a renv.lock file in our OSF project.

4.4 Results

4.4.1 Descriptive Statistics

Across our 110 participants, we sampled a total of 1,631 music-listening events, i.e., experience-sampled mood states surrounded by music-listening behavior. More precisely, participants provided between four and 55 music-listening events, with an average of 14.83 events (SD = 10.36) across persons. The number of available music-listening events was negatively related to age (r = -.24, $CI_{95\%}$ [-.33, -.15]) but otherwise unrelated to demographics or personality traits. During the 30-minute window of our music-listening events, participants, on average, played 4.31 songs (SD = 2.38) for 11.76 minutes (SD = 5.99), using mostly the music app “Spotify” (89%), followed by Google Play Music (3%), and Amazon Music (3%). We report descriptive statistics on music choices, mood states, and personality traits in Table 4.1 and their intercorrelations in Table 4.2. We visualize these intercorrelations and present correlations with demographic variables in Chapter S5 of our OSM.

The intra-class correlation coefficients (ICCs) in Table 4.1 indicate that 22% (for musical valence) and 32% (for musical energy) of the total variance of music choice measures was attributed to the grouping of music-listening events within persons. At the same time, music choices varied substantially within individuals (see Figure 4.2), confirming the multilevel structure of our data.

4.4.2 Multilevel Models

We regressed participants’ momentary music choices in terms of musical valence and energy on their enduring personality traits and concurrent mood states using multilevel modeling. The results of our MLMs and beta-distributed generalized linear mixed models showed considerable overlap, so we focus our report below on the preregistered MLMs (see Table 4.3 & Table S6.1) and report the results of the beta models in our OSM (see Table S6.2). The MLMs may have worked well despite the bounded criterium variables because most data points of the criteria fell in the middle of their scales (i.e., between 0.1 and 0.9), not close to the bounds (i.e., 0.0 and 1.0). Further confirming the MLMs’ fit, residuals at both levels

Table 4.2

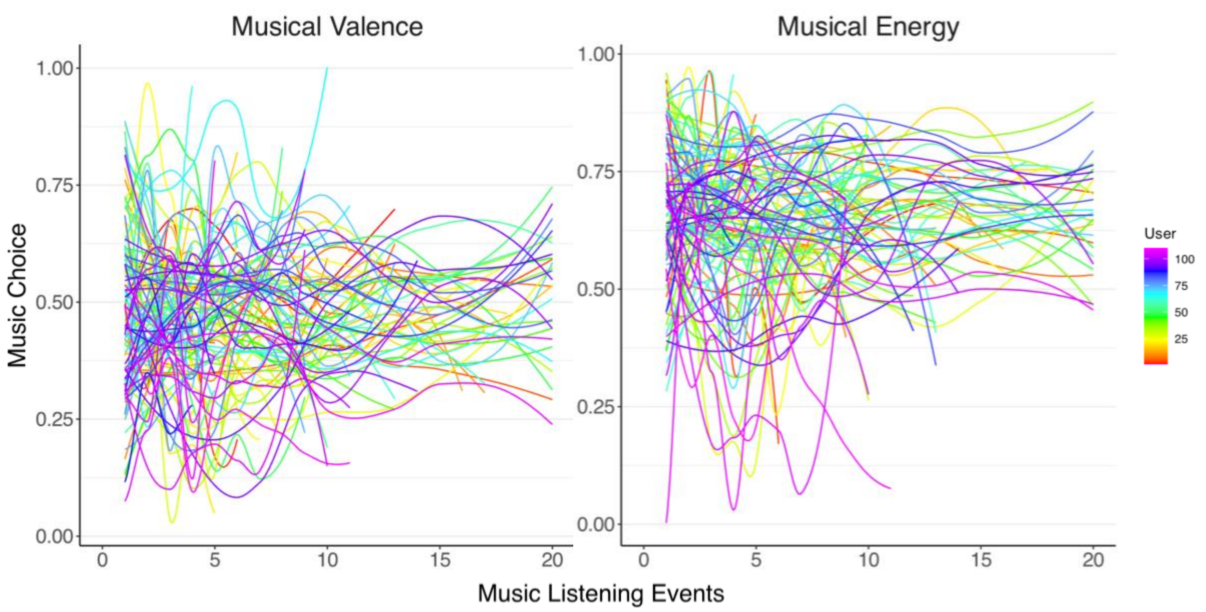
Within- and Between-Person Correlations Between State- and Trait-Level Measures

	Music Choice		Mood States		Personality Traits				
	Valence	Energy	Valence	Arousal	O	C	E	A	N
Music Choice									
Musical Valence	-	**.51** **[.36, .70]**	.05 [-.21, .31]	.05 [-.15, .26]	-	-	-	-	-
Musical Energy	**.42** **[.34, .50]**	-	.05 [-.17, .27]	-.03 [-.22, .17]	-	-	-	-	-
Mood States									
Mood Valence	-.01 [-.09, .07]	.05 [-.04, .14]	-	**.48** **[.34, .63]**	-	-	-	-	-
Mood Arousal	.02 [-.05, .10]	.06 [-.03, .15]	**.41** **[.32, .52]**	-	-	-	-	-	-
Personality Traits									
Openness	.01 [-.21, .21]	.06 [-.12, .26]	**.20** **[.02, .43]**	.16 [-.04, .35]	-	-	-	-	-
Conscientiousness	.08 [-.09, .26]	-.09 [-.28, .09]	.12 [-.06, .32]	**.25** **[.09, .44]**	.02 [-.17, .21]	-	-	-	-
Extraversion	.03 [-.18, .25]	-.05 [-.22, .15]	**.36** **[.22, .53]**	**.31** **[.14, .52]**	**.31** **[.12, .50]**	.15 [-.04, .32]	-	-	-
Agreeableness	.08 [-.11, .27]	-.07 [-.25, .12]	.11 [-.08, .31]	**.18** **[.01, .39]**	-.01 [-.21, .18]	.17 [-.03, .35]	.16 [-.03, .35]	-	-
Neuroticism	-.04 [-.29, .21]	.06 [-.16, .26]	**-.48** **[-.65, -.35]**	**-.28** **[-.47, -.08]**	-.11 [-.32, .06]	-.17 [-.38, .02]	**-.47** **[-.62, -.34]**	**-.20** **[-.39, -.02]**	-

Note. Each cell contains Pearson correlation coefficients (first row) and their 95% bootstrapped confidence intervals (CI; second row). For correlations among state measures (i.e., music choice & mood states), coefficients below the diagonal are means of within-person correlations (with Fisher's z-transformation used for pooling), and coefficients above the diagonal (highlighted in gray) are between-person correlations (i.e., correlations between person-means of the respective states). Coefficients in bold font represent correlations whose CIs do not contain zero. For musical and mood valence, higher values indicate more positive valence.

Figure 4.2

Variation in Music Choice Across Music-Listening Events



Note. Average musical valence and energy across the songs played in each music-listening event by each participant (N = 110). The number of collected music-listening events varied between four and 55 (*M* = 14.83, *SD* = 10.36) but the x-axis was cut at 20 events for clarity. Musical features represent automatically extracted audio characteristics of the songs from music logs of participants' smartphones (see Spotify.com). For musical valence, higher values indicate more positive-sounding songs.

exhibited no severe deviations from the model assumptions (see Snijders & Bosker, 2012). Only the homoscedasticity in our musical energy model was slightly violated for residuals at Level-1, but fixed effects in MLMs were previously shown to be robust to such distributional violations (Schielzeth et al., 2020).

For musical valence, Table 4.3 shows that none of the 15 included (and 10 preregistered) associations were statistically significant at the level of $p < .005$, leading us to reject all hypotheses for this criterium variable. In more detail, neither personality traits (see H1.1a-1.5a), mood states (see H2a), nor their interactions (see H3.1a-3.5a) appeared to be related to the musical valence of songs listened to. For the level-2 personality predictor Neuroticism, the regression coefficient even pointed in the reverse direction compared to our hypotheses. In line with this lack of effects, the overall proportion of variance explained by the MLM's fixed effects on musical valence was very small as suggested by a marginal $R^2_{(m)}$ of .01.

For musical energy, Table 4.3 indicates that only one of the 15 included (and 10 preregistered) associations was statistically significant at $p < .005$. The Level-1 predictor *mood arousal* was positively related to musical energy ($b = 0.01$, $p = .001$), which is consistent with the directionality expected in hypothesis 2 (see H2b). The unstandardized regression

Table 4.3

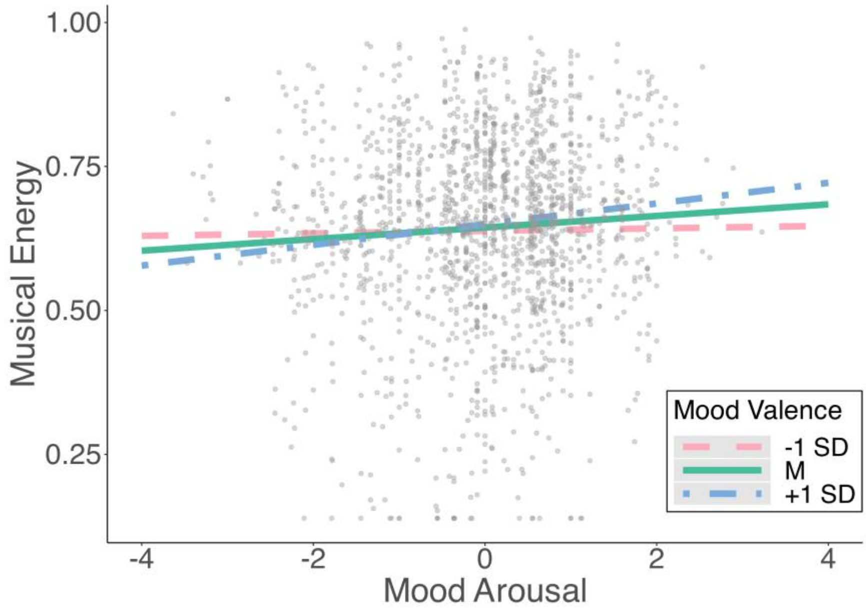
Fixed Effects for Multilevel Regression Models Predicting Music Choice

	b	*SE*	$CI_{95\%}$	β	*p*
Musical Valence					
Intercept	0.471	0.010	[0.45, 0.49]	-.01	.000
Mood Valence[1]	0.001	0.005	[-0.01, 0.01]	.00	.452^
Mood Arousal[1]	0.006	0.004	[0.00, 0.01]	.04	.107
Mean Mood Valence[2]	0.006	0.021	[-0.04, 0.05]	.02	.774
Mean Mood Arousal[2]	0.004	0.018	[-0.03, 0.04]	.01	.826
Openness[2]	0.000	0.014	[-0.03, 0.03]	.00	.507^
Conscientiousness[2]	0.006	0.015	[-0.02, 0.04]	.02	.337^
Extraversion[2]	0.001	0.019	[-0.04, 0.04]	.00	.476^
Agreeableness[2]	0.009	0.016	[-0.02, 0.04]	.03	.289^
Neuroticism[2]	0.003	0.014	[-0.02, 0.03]	.01	.576^
Mood Valence[1] x Mood Arousal[1]	0.006	0.004	[0.00, 0.01]	.04	.199
Mood Valence[1] x Openness[2]	0.008	0.007	[-0.01, 0.02]	.03	.142^
Mood Valence[1] x Conscientiousness[2]	0.012	0.007	[0.00, 0.03]	.04	.075
Mood Valence[1] x Extraversion[2]	-0.012	0.010	[-0.03, 0.01]	-.04	.883^
Mood Valence[1] x Agreeableness[2]	-0.009	0.008	[-0.02, 0.01]	-.03	.869^
Mood Valence[1] x Neuroticism[2]	0.003	0.006	[-0.01, 0.01]	.01	.697^
Musical Energy					
Intercept	0.644	0.010	[0.62, 0.66]	-.01	.000
Mood Valence[1]	0.007	0.005	[0.00, 0.02]	.04	.117
Mood Arousal[1]	**0.011**	**0.003**	**[0.00, 0.02]**	**.08**	**.001**^
Mean Mood Valence[2]	0.023	0.022	[-0.02, 0.07]	.08	.302
Mean Mood Arousal[2]	-0.005	0.019	[-0.04, 0.03]	-.02	.785
Openness[2]	0.010	0.014	[-0.02, 0.04]	.04	.747^
Conscientiousness[2]	-0.013	0.016	[-0.04, 0.02]	-.05	.210^
Extraversion[2]	-0.007	0.020	[-0.05, 0.03]	-.03	.644^
Agreeableness[2]	-0.009	0.017	[-0.04, 0.03]	-.03	.307^
Neuroticism[2]	0.007	0.014	[-0.02, 0.04]	.04	.311^
Mood Valence[1] x Mood Arousal[1]	0.010	0.004	[0.00, 0.02]	.07	.009
Mood Arousal[1] x Openness[2]	0.007	0.005	[0.00, 0.02]	.04	.066^
Mood Arousal[1] x Conscientiousness[2]	0.001	0.004	[-0.01, 0.01]	.01	.796
Mood Arousal[1] x Extraversion[2]	-0.011	0.006	[-0.02, 0.00]	-.05	.960^
Mood Arousal[1] x Agreeableness[2]	0.008	0.005	[0.00, 0.02]	.04	.047^
Mood Arousal[1] x Neuroticism[2]	0.002	0.004	[-0.01, 0.01]	.01	.695^

Note. N_{L1} = 1,631 observations from N_{L2} = 110 participants. *b* = unstandardized coefficients of multilevel models with person-mean centered Level-1 and grand-mean centered Level-2 predictors. *SE* = standard error of *b*. $CI_{95\%}$ = 95% confidence intervals for *b*. β = standardized coefficient from multilevel regression with z-standardized predictors (Level-1 predictors were additionally person-mean centered after z-standardizing). The coefficient in bold is statistically significant at a level of $p < .005$. [1] Level-1 predictors, [2] Level-2 predictors ^ One-tailed p-values (see preregistered hypotheses and footnote 3).

Figure 4.3

Interaction Effect of Mood Valence and Arousal on Chosen Musical Energy



Note. N_{L1} = 1,631 observations from N_{L2} = 110 participants. Predictors were person-mean centered. For mood valence, higher values indicate more positive mood states. Please note that only the main effect of mood arousal on musical energy was significant, while the effect of mood valence and the interaction effect were not (see Table 4.3).

coefficient implies that a one-point increase in self-reported mood arousal (ranging from 1 to 6) goes along with an average increase of .01 in the played music's energy level (ranging from 0 to 1) if all other predictors are kept constant. Mood arousal also exhibited the largest effect size (i.e., standardized coefficient) with β = .08, hinting at the superiority of this Level-1 variable compared to other state- and trait-level predictors. Because no other predictor reached statistical significance, all other hypotheses regarding the musical energy criterion must be rejected. In particular, none of the Big Five personality domains (see H1.1b–1.5b) exhibited significant relations with chosen musical energy, and for Openness (see H1.1b) and Extraversion (see H1.3b), the beta coefficients even contradicted our expected directionality. Similarly, personality traits did not interact significantly with mood states (see H3.1b-3.5b), falsifying our third hypothesis.

Beyond these hypothesized associations, the Level-2 predictor *mean mood valence* (b = 0.02, β = .08, p = .302) and the interaction term between the Level-1 predictors *mood arousal* and *mood valence* (b = 0.01, β = .07, p = .009) exhibited the highest – albeit non-significant – effect sizes, indicating that participants who, on average, experience more positive mood states and participants in a simultaneously more positive and more aroused mood state (e.g., joy) may listen to more energetic music (see Figure 4.3). Despite the significant Level-1 predictor, personality traits and mood states only explained a very small fraction of the variance in chosen musical energy as implied by a marginal $R^2_{(m)}$ of .02. Thus, while personality and mood were more informative about the musical energy than about the musical valence of played songs, both aspects of music choice were not well explained by our predictors.

4.5 Discussion

The present study employed a longitudinal multimethod design to examine the music people select on a moment-to-moment basis on their smartphones. We extracted listeners' momentary music preferences in terms of musical valence and energy from mobile-sensed music-listening records and predicted them from self-reported personality traits and experience-sampled mood states. Based on the theoretical reasoning and past empirical findings, we expected to replicate trait-congruent associations with the Big Five personality domains (H1) and mood-congruent associations with affective valence and arousal (H2).

Furthermore, we assumed that personality traits would moderate the association between mood states and music choice (H3).

However, our multilevel regression models showed that personality traits and mood states accounted only for a small proportion of variance in music choice. For musical valence, none of the personality and mood predictors or their interactions reached statistical significance, leading us to reject all hypotheses for this outcome. For musical energy, only one predictor, namely mood arousal, exhibited a significant albeit still weak effect, indicating that people in more activated mood states prefer more energetic music, which was consistent with our second hypothesis on mood congruence (H2b). In contrast, we had to reject the remaining hypotheses for the musical energy model. In the following sections, we discuss potential reasons for this lack of effects and provide an outlook on other factors that may play a role in momentary music preferences.

4.5.1 Modeling Momentary Music Preferences

We used multilevel regression models to investigate whether momentary music choices are related to stable personality traits and fluctuating mood states in a personality- and mood-congruent manner. However, these models explained only a small proportion of inter- and intraindividual variance in chosen musical valence and energy and revealed only one significant association. While we discuss several explanations for these effects (or the lack thereof), they should be interpreted with great caution because the statistical power provided by our small sample at Level-2 ($N = 110$) was not sufficient to reliably detect true positive effects (Maas & Hox, 2005; Scherbaum & Ferreter, 2009).

5.5.1.1 The Role of Personality

We found the Big Five personality domains to be largely unrelated to music choice (see Tables 4.2 & 4.3), which contradicts past studies that repeatedly reported trait-congruent preferences for technical audio characteristics similar to those in our study (Anderson et al., 2021; Dobrota & Reić Ercegovac, 2015; Flannery & Woolhouse, 2021; Ladinig & Schellenberg, 2012; Sust et al., 2023) and broader musical style dimensions (Bonneville-Roussy et al., 2013; Greenberg et al., 2022; Nave et al., 2018; Rentfrow & Gosling, 2003).

Because the majority of these studies assessed overall music preferences based on self-report questionnaires (Bonneville-Roussy et al., 2013; Rentfrow & Gosling, 2003) or ratings of musical excerpts (Dobrota & Reić Ercegovac, 2015; Flannery & Woolhouse, 2021; Nave et al., 2018; Ladinig & Schellenberg, 2012; Greenberg et al., 2022), they were low in ecological validity and, thus, their findings may not have generalized to natural music-listening behavior (Greenberg & Rentfrow, 2017). In particular, our mobile listening context may have attenuated the role of personality if smartphone-induced "auditory bubbles" (Bull, 2005, p. 344) isolated listeners from their surroundings, potentially decreasing the relevance of music for self-expression compared to self-reports or laboratory settings (e.g., DeNora, 1999; Laiho, 2004; Lonsdale & North, 2011; Schäfer et al., 2013). However, these differences in study design cannot fully explain our absence of findings as two past studies successfully predicted personality traits from music choices in natural listening records (Anderson et al., 2021; Sust et al., 2023).

Another explanation is provided in a meta-analysis by Schäfer and Mehlhorn (2017) who had previously reported only weak associations between personality traits and music preferences across studies, so our models may have lacked sufficient power to detect such small effects. In support of this reasoning, our descriptive analyses demonstrated that the music selected in everyday life varied more strongly within than between individuals. With only 20-30% of the variance in music choice attributable to the grouping of music-listening events within persons, it is reasonable that stable traits like personality only account for the small interindividual variance proportion in music choice. Past studies investigating self-reported music choice on smartphones reported ICCs in a similar range and, accordingly, found that only momentary variables like listeners' current mood or situation, but not their personality traits, exhibited significant effects (Greb et al., 2019; Greb, Steffens & Schlotz, 2018; Randall & Rickard, 2017). Hence, the contextual factors of everyday listening situations may impose musical affordances that inhibit personality congruence in music-listening behavior in the sense of the situational strength concept (Cooper & Withey, 2009; Mischel, 1977; Snyder & Ickes, 1985).

In this context, it should be noted that our study design did not consider the causal direction between personality and music choice. While it is tempting to assume an effect (however small) of the relatively stable construct of personality on music preferences (Buss, 1987; Swann, 1987), causality could also point in the other direction as people may adjust their auditory environments to their personalities or vice versa (Bleidorn et al., 2020; Rauthmann, 2021).

5.5.1.2 The Role of Mood

Our findings show that music choice was largely unrelated to mood states, except for chosen musical energy, which exhibited a positive association with listeners' concurrent arousal (see Table 4.3). Such a congruence effect was previously reported for the arousal dimension of mood (Greb et al., 2019; Randall & Rickard, 2017; Thoma et al., 2012; Yang & Liu, 2013) and may indicate that music is used to enact engagement rather than disengagement strategies of mood regulation, that is, to maintain certain arousal states instead of up- or downregulating them (i.e., to energize or relax). However, this effect was small, with a standardized beta coefficient below .10, and we could not replicate the mood congruence effects for the valence dimension repeatedly found in the past (Chen et al., 2007; DeMarco et al., 2015; Ferwerda et al., 2015; Friedman et al., 2012; Greb et al., 2019; Kinghorn, 2021; Lee et al., 2013; Randall & Rickard, 2017; Taruffi & Koelsch, 2014; Thoma et al., 2012; Xue et al., 2018; Yang & Liu, 2013). Several differences in the study designs could possibly explain the deviation in our results.

First, previous studies often focused on mood congruence regarding the affect category of sadness (e.g., Chen et al., 2007; DeMarco et al., 2015; Friedman et al., 2012; Taruffi & Koelsch, 2014; Xue et al., 2018), while we assessed music choice and mood states in terms of valence and arousal, the two dimensions of the circumplex model of affect (Russell, 1980). As we predicted each dimension of music choice separately in one model, we could not differentiate between distinct affect categories like sadness vs. anger. If valence congruence exists only for sadness in particular, but not for negative mood in general, this operationalization may have attenuated effects. Furthermore, we modeled positive and negative valence as inversely related, whereas some researchers consider them independent (e.g., Tellegen et al., 1999). Because mood-congruent music choice was sometimes investigated only for negative but not positive valence (e.g., Chen et al., 2007; DeMarco et al., 2015; Lee et al., 2013; Taruffi & Koelsch, 2014), this aspect of our mood conceptualization may also have obscured congruency effects. However, this reasoning cannot fully account for our lack of findings because several researchers previously reported mood-congruent music choices for non-specified negative and positive valence states when operationalizing mood via the dimensional circumplex model (Greb et al., 2019; Randall & Rickard, 2017; Thoma et al., 2012; Yang & Liu, 2013). In addition, musical valence and energy correlated highly positively in our study, indicating that the dimensional approach may have indirectly represented happy (i.e., positive and energetic) vs. sad (i.e., negative and calm) songs.

A second methodological reason for our deviating results could be that we assessed natural music-listening behavior in everyday life, while most past studies investigated music choices made in listening experiments with mood induction (Chen et al., 2007; DeMarco et al., 2015; Ferwerda et al., 2015; Friedman et al., 2012; Lee et al., 2013; Thoma et al., 2012; Xue et al., 2018). In our ecologically more valid setting, participants, on average, listened to songs with neutral valence and slightly increased energy (examples of such songs are Coldplay's "Viva La Vida" or "Too Lost in You" by the Sugarbabes), which aligns with the distribution of music preferences on smartphones found in a larger sample by Sust et al. (2023). Thus, in contrast to the song samples used in listening experiments (Chen et al., 2007; Ferwerda et al., 2015; Thoma et al., 2012), the naturally chosen songs were not prototypically positive vs. negative or calm vs. energetic, possibly reducing congruency effects. Similarly, the mood states assessed in our study occurred in mundane routine contexts and were, correspondingly, rather neutral in valence and arousal, whereas those elicited through autobiographical memory, film clips, or other induction tools may have been more intense (e.g., Chen et al., 2007; DeMarco & Friedman, 2018; Ferwerda et al., 2015; Friedman et al., 2012; Thoma et al., 2012; Xue et al., 2018). Hence, the findings for mood-congruent music choices in laboratory settings may not have generalized to natural music-listening behavior in our study (see also Greenberg & Rentfrow, 2017). However, contrary to this reasoning, a few studies still found the effects of mood-congruency in self-reported momentary music preferences on smartphones (Greb et al., 2019; Kinghorn, 2021; Randall & Rickard, 2017).

While these studies had participants rate the emotionality of their played songs themselves, we used Music Information Retrieval to automatically represent mobile-sensed song choice in terms of two technical audio characteristics. This methodological deviation could be another explanation for the lack of findings in our study because our objective audio characteristics did not capture how participants subjectively perceived songs' emotionality, which, in turn, is related to their personality traits and mood states (Hunter et al., 2011; Vuoskoski & Eerola, 2011). Furthermore, technical features cannot represent the personal meaning or memories associated with certain songs, which play a role in the emotional effects of music (Juslin et al., 2014; Taruffi & Koelsch, 2014; van Goethem & Sloboda, 2011). If mood congruence depends on subjective musical perception, our study was unable to replicate such effects.

Finally, our study design exhibited an important limitation in the event-triggered experience-sampling scheme that could also explain the discrepancies between our results and the literature. Past studies used experimental setups (e.g., Chen et al., 2007; DeMarco et al., 2015; Ferwerda et al., 2015; Friedman et al., 2012; Lee et al., 2013; Thoma et al., 2012; Xue et al., 2018) or assessed mood states at the onset of natural music-listening episodes (e.g., Greb et al., 2019; Randall & Rickard, 2017), so their findings may be interpreted in the sense of mood-based music preferences. In contrast, we sampled mood states whenever participants opened a music app on their smartphones, which occurred unsystematically right before, during, or after music listening (see footnote 4), and aggregated music choices over a window surrounding these experience-sampling instances. Thus, our analyses confounded the effects of mood on music choice with those of music listening on mood states, preventing any causal inferences. In particular, as music can elicit various affective states (e.g., Eerola & Vuoskoski, 2012; Juslin & Laukka, 2004; Lundqvist et al., 2009) and was previously found to return mood to neutral states (Randall & Rickard, 2017), mood states sampled after music listening may capture the effects of music instead of mood-dependent music choices. Hence, ES with different timing may have captured different effects that we could not discern and that potentially canceled each other out.

5.5.1.3 Moderation Effects

With the main effects of personality and mood being insignificant or very small, it is not surprising that we found no interaction effects, indicating that none of the Big Five domains moderated the mood-congruency effects in our data. Based on empirical findings from mood regulation and coping literature (Agbaria & Mokh, 2022; Baranczuk, 2019; Carver & Connor-Smith, 2010; Connor-Smith & Flachsbart, 2007; Watson & Hubbard, 1996), we had assumed those high in Openness, Conscientiousness, Extraversion, and Agreeableness would exhibit a stronger preference for mood-congruent music as these traits are positively related to the use of engagement-style mood regulation strategies, which, in turn, may require congruent music choices. Furthermore, we had expected that individuals high in Neuroticism would have a weaker preference for mood-congruent music since this trait is positively related to the use of disengagement-style mood regulation strategies, requiring the choice of incongruent songs. In contrast to these hypotheses and our findings, two past studies reported interaction effects inverse to our assumed directionality, namely that those higher in Openness, Extraversion, and Agreeableness tend to choose more incongruent music in negative mood states, while those higher in Neuroticism tend to select more mood-congruent songs (Ferwerda et al., 2015; Taruffi & Koelsch, 2014). As these studies investigated self-reports (Taruffi & Koelsch, 2014) and listening experiments (Ferwerda et al., 2015) instead of natural listening behavior, their findings may not have generalized to music choices made on smartphones. Alternatively, this discrepancy could also indicate that music does not serve to enact strategies like (dis-) engagement or that congruent vs. incongruent song choice cannot be automatically matched to distinct mood regulation strategies. Furthermore, individuals may apply music-based mood regulation strategies not in a dispositional way but more flexibly depending on the context in which a mood state occurs, so contextual factors instead of stable traits may moderate the association between mood and music choices. For example, listeners in negative mood states may sometimes choose congruent music to help them resolve the negative situation and, other times, incongruent music to divert their attention from the (unresolvable) negative situation.

4.5.2 Limitations

Our study faced several limitations that should be considered when interpreting our findings. First, as mentioned above, our initially large sample (N = 476) was reduced to a small size (N = 110) due to a lack of music-listening and experience-sampling data, so our multilevel analyses were severely underpowered, possibly obscuring effects (see Maas & Hox, 2005; Scherbaum & Ferreter, 2009). To obtain a larger sample size, future studies should employ an elaborate pre-screening strategy to include only participants who regularly listen to music on their smartphones (see Greb et al., 2019). Furthermore, studies should improve the experience-sampling scheduling and present mood questionnaires whenever participants play music and not only when they open a music app to not miss music-listening instances controlled via the banner on the lock screen (see footnote 4). Such a design will also help disentangle the confounding effect of music listening on mood states discussed above if mood states are consistently sampled prior to music listening and if auto-regressions are considered in the modeling process.

Second, our sample composition may have restricted the generalizability of our results. As commonly the case in university recruitment contexts, our sample consisted of mostly young and female participants drawn from a WEIRD population (Henrich et al., 2010). Since music preferences vary by age and gender (e.g., Bonneville-Roussy et al., 2013; Greenberg et al., 2022) as well as between countries (e.g., Bello & Garcia, 2021; Park et al., 2019), follow-up studies should transfer our study design to more heterogeneous samples, which, however, have to be drawn from populations with sufficient smartphone penetration, possibly excluding, for example, certain age groups or countries. While our study's scope was also limited to owners of Android smartphones, users of different operating systems seem to not differ systematically, according to the literature (Götz et al., 2017; Keusch et al., 2020).

Third, our data collection may have been impacted by the COVID-19 pandemic. Even though our study did not take place during a lockdown, other legislative measures were still in place, restricting everyday life and, potentially, participants' music-listening behavior (Mathieu et al., 2020). In particular, the containment measures limited mobility and socializing behavior (Steinmetz et al., 2022), which, in turn, may have affected when and to what music participants listened to. For example, with home office or online classes reducing mobility, participants had fewer opportunities for mobile music listening on their smartphones, possibly opting for stationary devices (e.g., notebooks) at home instead, which could explain the lack of music-listening events in our data. Furthermore, the absence of social events like parties may have systematically altered music choices, reducing the relevance of cheerful (i.e., positive and energetic) music. Future studies may want to reassess momentary music preferences in natural music-listening behavior in times without COVID-19 containment measures.

Fourth, we cannot confirm whether participants actively chose (and liked) the music they listened to on their smartphones because music apps offer various editorial, algorithmic, or user-created playlists and allow listeners to select music via the shuffle mode, which they do especially while on the go (Heye & Lamont, 2010). In particular, our participants often started playing music without opening their music apps, indicating that they did not search for a specific song but simply played what was on before. In those instances, listeners may not have been invested in their music choice, potentially obscuring personality- and mood-congruency effects. Furthermore, automated music recommendations pose a risk of listeners getting stuck in "filter bubbles" (Petridis, 2022), that is, overly personalized areas in the recommender space that may limit the intraindividual variance in their momentary music preferences. Hence, researchers should try to discern active choices from automatic recommendations when sensing music-listening behavior, for example, by tracking keystrokes or by explicitly asking participants about their selection mode in ES.

4.5.3 Outlook on Music-Listening Research

The present study focused on personality traits and mood states, that is, person variables, to explain inter- and intraindividual variance in everyday music choices. However, natural music-listening behavior usually takes place in some situational context, defined, for example, by the current location, time of day, social company, and concurrent activities (Juslin et al., 2008; North et al., 2004; North & Hargreaves, 1996; Sloboda et al., 2001; Sloboda & O'Neill, 2001). Hence, to understand music choices, we should not only consider the attributes of the person but also those of their situations, as suggested by theories on person-environment transactions like the personality triad (e.g., Funder, 2006, 2009; Rauthmann, 2021). Indeed, both the objective cues (e.g., listeners' current location or activity) and the subjective experience of situations (e.g., how dutiful or sociable listeners perceive a situation, see Rauthmann et al., 2014) were previously shown to predict momentary music choices in a way that music preferences augmented the qualities of the listening situation (Behbehani & Steffens, 2021; Greb et al., 2019; Greb, Steffens & Schlotz, 2018; North & Hargreaves, 1996; Randall & Rickard, 2017; Yang & Teng, 2015). As an underlying mechanism, different listening situations may present specific affordances regarding the uses of music (Greb, Schlotz & Steffens, 2018; North et al., 2004; Randall & Rickard, 2017; Volokhin & Agichtein, 2018). More specifically, music can serve various uses beyond self-expression or mood regulation, like pure aesthetic enjoyment, as background noise, to support physical activities such as dancing, or to create a social connection with others (Chamorro-Premuzic & Furnham, 2007; Chin & Rickard, 2012; Lonsdale & North, 2011; Schäfer et al., 2013), which, are related to situational context and require different types of music (Chamorro-Premuzic et al., 2010; Getz et al., 2015; Greb, Steffens & Schlotz, 2018; Vella & Mills, 2017). Thus, it may depend on the current situation and corresponding use of music, what music people choose, and whether personality and mood play a role in momentary music preferences. In support of this reasoning, an initial study by Greb et al. (2019) found that the uses of music listening mediate the association of music choices with mood states and situational variables.

To further explore the complex dynamics between listeners' states, situations, music uses and their music choices on a moment-to-moment basis, future studies may want to extend our approach of integrating active and passive ambulatory assessment into smartphones.

Smartphone sensing not only allows researchers to objectively assess natural music-listening behavior through digital listening records but also to log various parameters of participants' listening situations from phone sensors, such as their current location from GPS, their activity from accelerometers, or their company from ambient noise (for overviews, see Harari et al., 2020; Schoedel et al., 2023). In addition, ES can obtain in situ self-reports about concurrent states, subjective situation perceptions, and uses of music in timely contingency (i.e., before, during, or after) with music listening on the smartphone. In sum, smartphones provide ample possibilities for investigating the dynamic interplay of manifold variables, which will broaden our understanding of music-listening behavior.

4.6 Conclusion

The present study employed a smartphone-based longitudinal sampling design to investigate momentary music choices in relation to enduring personality traits and concurrent mood states. Based on theoretical reasoning and past empirical findings, we expected that the musical valence and energy of chosen songs would be congruent with listeners' Big Five personality domains and their mood valence and arousal, with personality moderating mood-congruency. However, our multilevel regression models explained only a small fraction of variance in music choices, revealing only one significant albeit still weak mood-congruency effect indicating that listeners in more activated states chose more energetic songs. Beyond that, our models failed to replicate trait- and state-congruent momentary music preferences or interaction effects, which could be due to our limited statistical power. Nevertheless, the present study introduced an ecologically valid ambulatory assessment approach to studying inter- and intraindividual differences in natural music-listening behavior, which may be extended in numerous ways in future studies.

4.7 Acknowledgements

We thank the entire PhoneStudy team at LMU Munich for their continuous and diligent work on the PhoneStudy app, enabling our research. Special thanks go to Markus Bühner for providing the infrastructure for our research and Florian Bemmann for giving technical support during this study. We further thank Florian Pargent for his helpful advice on our multilevel modeling. Finally, thank all student researchers for supporting us with recruiting participants.

4.8 Appendix

Table 4.A1

Overview of Preregistered and Applied Data Exclusion Criteria

Order	Level	Preregistered Exclusions	Applied Exclusions	# Exclusions
476 participants with mobile-sensing data				
1	Participants	We will exclude participants without (ES) data on their mood states.	We excluded participants without any ES instances.	66 participants
2	Participants	We will consider only participants with music-listening events that have preceding mood self-reports.	We excluded participants without any music-listening data in their sensing data.	79 participants
3	Participants		We excluded participants without music-listening data in the 30-minute window[1] around any ES.	115 participants
→ 216 participants with windows of sensed music-listening data in timely proximity to ES instances				
4	Windows	We will exclude windows where less than 30% of the logged songs have available song-level information from Spotify.com.	We excluded windows containing less than one song[2] with available Spotify information.	3352 windows (24 participants)
5	Windows	We will exclude windows containing only songs listened to for less than 20 seconds.	There were no instances meeting this criterion.	/
6	Windows	We will exclude windows with a listening duration shorter than 3 minutes.	We excluded windows with less than 1 minute[2] of music listening.	30 windows (2 participants)
7	Windows	We will exclude participants without (ES) data on their mood states (see 1).	We excluded windows with NAs in either the arousal or valence item of the ES.	2 events (0 participants)
→ 190 participants with at least one valid music-listening event[3]				
8	Participants	We will exclude participants with fewer than 7 valid music-listening events.	We exclude participants with less than 4 valid music-listening events[2].	77 participants

Table 4.A1 (continued)

Order	Level	Preregistered Exclusions	Applied Exclusions	# Exclusions
9	Participants	We will exclude participants with zero variance in either of the two mood variables.	We excluded participants with zero variance in valence or arousal across music-listening events.	2 participants
→ 111 participants with a sufficient number of valid music-listening events				
10	Participants	We will exclude participants who listened to music on fewer than 5 study days.	We lowered this threshold to 3 study days[2]. There were no participants meeting this criterion.	/
11	Participants	We will exclude participants who listened to fewer than 5 different songs.	There were no participants meeting this criterion.	/
12	Participants	We will exclude participants for whom technical errors lead to strongly distorted music-listening data.	There were no participants meeting this criterion.	/
13	Participants	We will exclude participants without full data in the BFI-2-S.	We excluded participants without complete BFI-2-S survey data.	1 participant
→ 110 participants with all data required for testing our hypotheses				
14	Participants	We will exclude participants younger than 18 years.	There were no participants meeting this criterion.	/
→ 110 participants meeting all our inclusion criteria				

Note. The rationale behind these exclusions was to ensure a sufficient amount of high-quality data for the outcome (sensed music-listening) and predictor variables on Level-1 (experience-sampled mood states) and Level-2 (personality surveys) in our multilevel regression models. The applied preprocessing pipeline deviates from the preregistered order of exclusions to accommodate for the data structure which we were aware of only after data inspection. Applied exclusions highlighted in gray deviate from the preregistered criteria.

[1] We adapted the variable extraction strategy by changing the extraction window from the preregistered "30 minutes after an ES" to "15 minutes before AND after an ES" to obtain more music-listening events (see also footnote 3).

[2] We loosened the stringency in our exclusion thresholds to preserve a reasonably large dataset without risking a loss of data quality.

[3] Valid music-listening events are defined as completed ES instances (i.e., where both mood items were answered) surrounded by a 30-minute window where a) music was played for at least one minute and b) at least one played song had available song-level information from Spotify.com.

4.9 References

Agbaria, Q., & Mokh, A. A. (2022). Coping with stress during the Coronavirus outbreak: The contribution of big five personality traits and social support. *International Journal of Mental Health and Addiction*, *20*(3), 1854-1872. https://doi.org/10.1007/s11469-021-00486-2

Allport, G. W. (1927). Concepts of trait and personality. *Psychological Bulletin*, *24*(5), 284-293. https://doi.org/10.1037/h0073629

Ammar, A., Mueller, P., Trabelsi, K., Chtourou, H., Boukhris, O., Masmoudi, L., Bouaziz, B., Brach, M., Schmicker, M., Bentlage, E., How, D., Ahmed, M., Aloui, A., Hammouda, O., Paineiras-Domingos, L. L., Braakman-Jansen, A., Wrede, C., Bastoni, S., Pernambuco, C. S., … Hoeckelmann, A. (2020). Psychological consequences of COVID-19 home confinement: The ECLB-COVID19 multicenter study. *PloS one*, *15*(11): e0240204. https://doi.org/10.1371/journal.pone.0240204

Anderson, I., Gil, S., Gibson, C., Wolf, S., Shapiro, W., Semerci, O., & Greenberg, D. M. (2021). Just the way you are: Linking music listening on Spotify and personality. *Social Psychological and Personality Science, 12*(4), 561-572. https://doi.org/10.1177/1948550620923228

Barańczuk, U. (2019). The five factor model of personality and emotion regulation: A meta-analysis. *Personality and Individual Differences*, *139*, 217-227. https://doi.org/10.1016/j.paid.2018.11.025

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015a). *Parsimonious mixed models*. arXiv. https://doi.org/10.48550/arXiv.1506.04967

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015b). Fitting linear mixed effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science*, *2*(4), 396-403. https://doi.org/10.1111/j.1745-6916.2007.00051.x

Behbehani, S., & Steffens, J. (2021). Musical DIAMONDS: The influence of situational classes and characteristics on music listening behavior. *Psychology of Music, 49*(6), 1532-1545. https://doi.org/10.1177/0305735620968910

Bello, P., & Garcia, D. (2021). Cultural Divergence in popular music: the increasing diversity of music consumption on Spotify across countries. *Humanities and Social Sciences Communications, 8*(1), 1-8. https://doi.org/10.1057/s41599-021-00855-1

Bleidorn, W., Hopwood, C. J., Back, M. D., Denissen, J. J., Hennecke, M., Jokela, M., Kandler, C., Lucas, R. E., Luhmann, M., Orth, U., Roberts, B. W., Wagner, J., Wrzus, C., & Zimmermann, J. (2020). Longitudinal experience–wide association studies - A framework for studying personality change. *European Journal of Personality, 34*(3), 285-300. https://doi.org/10.1002/per.2247

Bonneville-Roussy, A., Rentfrow, P. J., Xu, M. K., & Potter, J. (2013). Music through the ages: Trends in musical engagement and preferences from adolescence through middle adulthood. *Journal of Personality and Social Psychology, 105*(4), 703. https://doi.org/10.1037/a0033770

Bull, M. (2005). No dead air! The iPod and the culture of mobile listening. *Leisure studies*, *24*(4), 343-355. https://doi.org/10.1080/0261436052000330447

Buss, D. M. (1987). Selection, evocation, and manipulation. *Journal of Personality and Social Psychology, 53*(6), 1214-1221. https://doi.org/10.1037/0022-3514.53.6.1214

Carver, C. S., & Connor-Smith, J. (2010). Personality and coping. *Annual Review of Psychology*, *61*, 679-704. https://doi.org/10.1146/annurev.psych.093008.100352

Chamorro-Premuzic, T., Fagan, P., & Furnham, A. (2010). Personality and uses of music as predictors of preferences for music consensually classified as happy, sad, complex, and social. *Psychology of Aesthetics, Creativity, and the Arts, 4*(4), 205-213. https://doi.org/10.1037/a0019210

Chamorro-Premuzic, T., & Furnham, A. (2007). Personality and music: Can traits explain how people use music in everyday life?. *British Journal of Psychology*, *98*(2), 175-185. https://doi.org/10.1348/000712606X111177

Charles, N. E., Strong, S. J., Burns, L. C., Bullerjahn, M. R., & Serafine, K. M. (2021). Increased mood disorder symptoms, perceived stress, and alcohol use among college students during the COVID-19 pandemic. *Psychiatry Research, 296*, 113706. https://doi.org/10.1016/j.psychres.2021.113706

Chen, L., Zhou, S., & Bryant, J. (2007). Temporal changes in mood repair through music consumption: Effects of mood, mood salience, and individual differences. *Media Psychology*, *9*(3), 695-713. https://doi.org/10.1080/15213260701283293

Chin, T., & Rickard, N. S. (2012). The music USE (MUSE) questionnaire: An instrument to measure engagement in music. *Music Perception*, *29*(4), 429-446. https://doi.org/10.1525/mp.2012.29.4.429

Conner, T. S., & Mehl, M. R. (2015). Ambulatory assessment: Methods for studying everyday life. In R. A. Scott & S. M. Kosslyn (Eds.), *Emerging Trends in the Social and Behavioral Sciences* (1st ed., pp. 1-15). Wiley. https://doi.org/10.1002/9781118900772.etrds0010

Connor-Smith, J. K., & Flachsbart, C. (2007). Relations between personality and coping: a meta-analysis. *Journal of Personality and Social Psychology*, *93*(6), 1080-1107. https://doi.org/10.1037/0022-3514.93.6.1080

Cooper, W. H., & Withey, M. J. (2009). The strong situation hypothesis. Personality and *Social Psychology Review, 13*(1), 62-72. https://doi.org/10.1177/1088868308329378

Datta, H., Knox, G., & Bronnenberg, B. J. (2018). Changing their tune: How consumers' adoption of online streaming affects music consumption and discovery. *Marketing Science*, *37*(1), 5-21. https://doi.org/10.1287/mksc.2017.1051

De Raad, B. (2000). *The Big Five Personality Factors: The Pycholexical Approach to Personality*. Hogrefe & Huber Publishers.

DeMarco, T. C., Taylor, C. L., & Friedman, R. S. (2015). Reinvestigating the effect of interpersonal sadness on mood-congruency in music preference. *Psychology of Aesthetics, Creativity, and the Arts*, *9*(1), 81-90. https://doi.org/10.1037/a0038691

DeNora, T. (1999). Music as a technology of the self. *Poetics*, *27*(1), 31-56. https://doi.org/10.1016/S0304-422X(99)00017-0

Dobrota, S., & Reić Ercegovac, I. (2015). The relationship between music preferences of different mode and tempo and personality traits – Implications for music pedagogy. *Music Education Research, 17*(2), 234-247. https://doi.org/10.1080/14613808.2014.933790

Downie, J. S. (2003). Music information retrieval. *Annual Review of Information Science and Technology*, *37*(1), 295-340. https://doi.org/10.1002/aris.1440370108

Eerola, T., Lartillot, O., & Toiviainen, P. (2009). Prediction of multidimensional emotional ratings in music from audio using multivariate regression models. In K. Hirata, G. Tzanetakis & K. Yoshii (Eds.), *Proceedings of the 10th International Society for Music Information Retrieval Conference* (pp. 621-626). International Society for Music Information Retrieval.

Eerola, T., & Vuoskoski, J. K. (2012). A review of music and emotion studies: Approaches, emotion models, and stimuli. *Music Perception: An Interdisciplinary Journal*, *30*(3), 307-340. https://doi.org/10.1525/mp.2012.30.3.307

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. *Psychological Methods, 12*(2), 121. https://doi.org/10.1037/1082-989X.12.2.121

Ferwerda, B., Schedl, M., & Tkalcic, M. (2015). Personality & emotional states: Understanding users' music listening needs. In A. I. Cristea, J. Masthoff, A. Said, & N. Tintarev (Eds.), *Posters, Demos, Late-breaking Results and Workshop Proceedings of the 23rd Conference on User Modeling, Adaptation, and Personalization: Vol. 1388,* CEUR-WS.org.

Flannery, M. B., & Woolhouse, M. H. (2021). Musical Preference: Role of personality and music-related acoustic features. *Music & Science*, *4*, 1-14. https://doi.org/10.1177/20592043211014014

Fournier, D. A., Skaug, H. J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M., Nielsen, A., & Sibert, J. (2012). AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software, 27*(2), 233-249. https://doi.org/10.1080/10556788.2011.597854

Friedman, R. S., Gordis, E., & Förster, J. (2012). Re-exploring the influence of sad mood on music preference. *Media Psychology*, *15*(3), 249-266. https://doi.org/10.1080/15213269.2012.693812

Funder, D. C. (2006). Towards a resolution of the personality triad: Persons, situations, and behaviors. *Journal of Research in Personality, 40*(1), 21-34. https://doi.org/10.1016/j.jrp.2005.08.003

Funder, D. C. (2009). Persons, behaviors and situations: An agenda for personality psychology in the postwar era. *Journal of Research in Personality, 43*(2), 120-126. https://doi.org/10.1016/j.jrp.2008.12.041

Furr, R. M. (2009). Personality psychology as a truly behavioural science. *European Journal of Personality*, *23*(5), 369-401. https://doi.org/10.1002/per.724

Gantz, W., Gartenberg, H. M., Pearson, M. L., & Schiller, S. O. (1978). Gratifications and expectations associated with pop music among adolescents. *Popular Music & Society*, *6*(1), 81-89. https://doi.org/10.1080/03007767808591113

Getz, L. M., Chamorro-Premuzic, T., Roy, M. M., & Devroop, K. (2012). The relationship between affect, uses of music, and music preferences in a sample of South African adolescents. *Psychology of Music, 40*(2), 164-178. https://doi.org/10.1177/0305735610381818

Ghosh, D., & Vogt, A. (2012). Outliers: An evaluation of methodologies. In American Statistical Association (Ed.), *Joint Statistical Meetings*: Vol. 12 (pp. 3455-3460). American Statistical Association.

Goldberg, L. R. (1990). An alternative" description of personality": The big-five factor structure. *Journal of Personality and Social Psychology, 59*(6), 1216-1229. https://doi.org/10.1037/0022-3514.59.6.1216

Götz, F. M., Stieger, S., & Reips, U. D. (2017). Users of the main smartphone operating systems (iOS, Android) differ only little in personality. *PloS ONE, 12*(5), e0176921. https://doi.org/10.1371/journal.pone.0176921

Greb, F., Schlotz, W., & Steffens, J. (2018). Personal and situational influences on the functions of music listening. *Psychology of Music, 46*(6), 763-794. https://doi.org/10.1177/0305735617724883

Greb, F., Steffens, J., & Schlotz, W. (2018). Understanding music-selection behavior via statistical learning: using the percentile-Lasso to identify the most important factors. *Music & Science*, *1*. https://doi.org/10.1177/2059204318755950

Greb, F., Steffens, J., & Schlotz, W. (2019). Modeling music-selection behavior in everyday life: a multilevel statistical learning approach and mediation analysis of experience sampling data. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.00390

Greenberg, D. M., & Rentfrow, P. J. (2017). Music and big data: A new frontier. *Current Opinion in Behavioral Sciences, 18*, 50-56. https://doi.org/10.1016/j.cobeha.2017.07.007

Greenberg, D. M., Wride, S. J., Snowden, D. A., Spathis, D., Potter, J., & Rentfrow, P. J. (2022). Universals and variations in musical preferences: A study of preferential reactions to Western music in 53 countries. *Journal of Personality and Social Psychology, 122*(2), 286-309. https://doi.org/10.1037/pspp0000397

Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology*, *2*(3), 271-299. https://doi.org/10.1037/1089-2680.2.3.271

Hale, T., Angrist, N., Goldszmidt, R., Kira, B., Petherick, A., Phillips, T., Webster, S., Cameron Blake, E., Hallas, L., Majumdar, S., & Tatlow, H. (2021). A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nature Human Behaviour*, *5*(4), 529–538. https://doi.org/10.1038/s41562-021-01079-8

Harari, G. M., Lane, N. D., Wang, R., Crosier, B. S., Campbell, A. T., & Gosling, S. D. (2016). Using smartphones to collect behavioral data in psychological science: Opportunities, practical considerations, and challenges. *Perspectives on Psychological Science*, *11*(6), 838-854. https://doi.org/10.1177/1745691616650285

Harari, G. M., Müller, S. R., Aung, M. S., & Rentfrow, P. J. (2017). Smartphone sensing methods for studying behavior in everyday life. *Current Opinion in Behavioral Sciences*, *18*, 83-90. https://doi.org/10.1016/j.cobeha.2017.07.018

Harari, G. M., Müller, S. R., & Gosling, S. D. (2020). Naturalistic assessment of situations using mobile sensing methods. In J. F. Rauthmann, R. A. Sherman, & D. C. Funder (Eds.), *The Oxford Handbook of Psychological Situations* (pp. 298–311). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190263348.013.14

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, *466*(7302), 29-29. https://doi.org/10.1038/466029a

Heye, A., & Lamont, A. (2010). Mobile listening situations in everyday life: The use of MP3 players while travelling. *Musicae Scientiae*, *14*(1), 95-120. https://doi.org/10.1177/102986491001400104

Hunter, P. G., Schellenberg, E. G., & Griffith, A. T. (2011). Misery loves company: mood-congruent emotional responding to music. *Emotion, 11*(5), 1068-1072. https://doi.org/10.1037/a0023749

Hunter, P. G., Schellenberg, E. G., & Schimmack, U. (2010). Feelings and perceptions of happiness and sadness induced by music: Similarities, differences, and mixed emotions. *Psychology of Aesthetics, Creativity, and the Arts*, *4*(1), 47-56. https://doi.org/10.1037/a0016873

IFPI. (2019). Global Music Listening Report 2019. (n.p.). https://www.ifpi.org/resources/

John, O. P., & Srivastava, S. (1999). The Big-Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of Personality: Theory and Research* (2nd ed., pp.102-138). Guilford.

Juslin, P. N., Harmat, L., & Eerola, T. (2014). What makes music emotionally significant? Exploring the underlying mechanisms. *Psychology of Music, 42*(4), 599-623. https://doi.org/10.1177/0305735613484548

Juslin, P. N., & Laukka, P. (2004). Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of New Music Research*, *33*(3), 217-238. https://doi.org/10.1080/0929821042000317813

Juslin, P. N., Liljeström, S., Västfjäll, D., Barradas, G., & Silva, A. (2008). An experience sampling study of emotional reactions to music: listener, music, and situation. *Emotion*, *8*(5), 668-683. https://doi.org/10.1037/a0013505

Katz, E., Blumler, J. G., & Gurevitch, M. (1973). Uses and gratifications research. *The Public Opinion Quarterly*, *37*(4), 509-523. http://www.jstor.org/stable/2747854

Kelley, K. (2018). *MBESS: The MBESS R package*. (Version 4.0.2) [R package]. https://CRAN.R-project.org/package=MBESS

Keusch, F., Bähr, S., Haas, G. C., Kreuter, F., & Trappmann, M. (2020). Coverage error in data collection combining mobile surveys with passive measurement using apps: Data from a German national survey. *Sociological Methods & Research*, 52(2), 841-878. https://doi.org/10.1177/0049124120914924

Kinghorn, E. E. (2021). *Mood, Music Choices, and the Emotional Outcomes of Music Listening: An Examination of the Moderating Role of Rumination using Experience-Sampling Methodology* [Doctoral dissertation, The University of Western Ohio]. Western Libraries.

Knobloch, S., & Zillmann, D. (2002). Mood management via the digital jukebox. *Journal of Communication*, *52*(2), 351-366. https://doi.org/10.1111/j.1460-2466.2002.tb02549.x

Krause, A. E., & North, A. C. (2016). Music listening in everyday life: Devices, selection methods, and digital technology. *Psychology of Music*, *44*(1), 129-147. https://doi.org/10.1177/0305735614559065

Kuch, M., & Wöllner, C. (2021). On the move: Principal components of the functions and experiences of mobile music listening. *Music & Science*, *4*. https://doi.org/10.1177/20592043211032852

Kushlev, K., & Heintzelman, S. J. (2018). Put the Phone Down: Testing a Complement-Interfere Model of Computer-Mediated Communication in the Context of Face-to-Face Interactions. *Social Psychological and Personality Science*, *9*(6), 702–710. https://doi.org/10.1177/1948550617722199

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82*(13), 1-26. https://doi.org/10.18637/jss.v082.i13

Ladinig, O., & Schellenberg, E. G. (2012). Liking unfamiliar music: Effects of felt emotion and individual differences. *Psychology of Aesthetics, Creativity, and the Arts*, *6*(2), 146-154. https://doi.org/10.1037/a0024671

Laiho, S. (2004). The psychological functions of music in adolescence. *Nordic Journal of Music Therapy*, *13*(1), 47-63. https://doi.org/10.1080/08098130409478097

Larsen, R. J. (2000). Toward a science of mood regulation. *Psychological Inquiry*, *11*(3), 129-141. https://doi.org/10.1207/S15327965PLI1103_01

Larson, R., & Csikszentmihalyi, M. (2014). The experience sampling method. In M. Csikszentmihalyi (Ed.), *Flow and the Foundations of Positive Psychology* (pp. 21–34). Springer Dordrecht.

Laurier, C., Lartillot, O., Eerola, T., & Toiviainen, P. (2009). Exploring relationships between audio features and emotion in music. In P.-S. Eerola, T. Eerola, T. Himberg, J. Louhivuori, P. Pirjo-Leena, & S. Saarikallio (Eds.), *7th Triennial Conference of European Society for the Cognitive Sciences of Music*. Jyväskylän yliopisto Msiikin laitos.

Lazarus, R. S. (1966). *Psychological Stress and the Coping Process*. McGraw-Hill

Lee, C. J., Andrade, E. B., & Palmer, S. E. (2013). Interpersonal relationships and preferences for mood-congruency in aesthetic experiences. *Journal of Consumer Research*, *40*(2), 382-391. https://doi.org/10.1086/670609

Lonsdale, A. J., & North, A. C. (2011). Why do we listen to music? A uses and gratifications analysis. *British journal of psychology*, *102*(1), 108-134. https://doi.org/10.1348/000712610X506831

Lorah, J. (2018). Effect size measures for multilevel models: definition, interpretation, and TIMSS example. *Large-Scale Assessments in Education, 6*(1). https://doi.org/10.1186/s40536-018-0061-2

Lundqvist, L. O., Carlsson, F., Hilmersson, P., & Juslin, P. N. (2009). Emotional responses to music: Experience, expression, and physiology. *Psychology of Music*, *37*(1), 61-90. https://doi.org/10.1177/0305735607086048

Maas, C. J., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, 1(3), 86-92. https://doi.org/10.1027/1614-2241.1.3.86

Mathieu, E., Ritchie, H., Rodés-Guirao, L., Appel, C., Giattino, C., Hasell, J., Macdonald, B., Dattani, S., Beltekian, D., Ortiz-Ospina, E., & Roser, M. (2020). *Coronavirus Pandemic (COVID-19)*. OurWorldInData.Org. https://ourworldindata.org/coronavirus

McCrae, R., R., & John, O., P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality, 60*(2), 175–215. https://doi.org/10.1111/j.1467-6494.1992.tb00970.x

Miranda, D., & Claes, M. (2009). Music listening, coping, peer affiliation and depression in adolescence. *Psychology of Music*, *37*(2), 215-233. https://doi.org/10.1177/0305735608097245

Mischel, W. (1977). The interaction of person and situation. In D. Magnusson & N. S. Endler (Eds.), *Personality at the Crossroads: Current Issues in Interactional Psychology* (pp. 333-352). Erlbaum.

Naragon-Gainey, K., McMahon, T. P., & Chacko, T. P. (2017). The structure of common emotion regulation strategies: A meta-analytic examination. *Psychological Bulletin*, *143*(4), 384-427.https://doi.org/10.1037/bul0000093

Nave, G., Minxha, J., Greenberg, D. M., Kosinski, M., Stillwell, D., & Rentfrow, J. (2018). Musical preferences predict personality: Evidence from active listening and Facebook likes. *Psychological Science, 29*(7), 1145-1158. https://doi.org/10.1177/0956797618761659

North, A. C., & Hargreaves, D. J. (1996). Situational influences on reported musical preference. *Psychomusicology: A Journal of Research in Music Cognition, 15*(1-2), 30-45. https://doi.org/10.1037/h0094081

North, A. C., Hargreaves, D. J., & Hargreaves, J. J. (2004). Uses of music in everyday life. *Music Perception*, *22*(1), 41-77. https://doi.org/10.1525/mp.2004.22.1.41

Park, M., Thom, J., Mennicken, S., Cramer, H., & Macy, M. (2019). Global music streaming data reveal diurnal and seasonal patterns of affective preference. *Nature Human Behaviour*, *3*(3), 230-236. https://doi.org/10.1038/s41562-018-0508-z

Petridis, S., Daskalova, N., Mennicken, S., Way, S. F., Lamere, P., & Thom, J. (2022). Tastepaths: Enabling deeper exploration and understanding of personal preferences in recommender systems. In G. Jacucci, S. Kaski, C. Conati, S. Stumpf, T. Ruotsalo, & K. Gajos (Eds.), *27th International Conference on Intelligent User Interfaces* (pp. 120-133). Association for Computing Machinery. https://doi.org/10.1145/3490099.3511156

Python Software Foundation. (2021). *Python: A dynamic, open source programming language* [Computer Software]. Python Software Foundation. https://www.python.org

Qiu, L., Chen, J., Ramsay, J., & Lu, J. (2019). Personality predicts words in favorite songs. *Journal of Research in Personality*, *78*, 25-35. https://doi.org/10.1016/j.jrp.2018.11.004

R Core Team (2022). *R: A language and environment for statistical computing* [Computer Software]. R Foundation for Statistical Computing. https://www.r-project.org

Rammstedt, B., Danner, D., Soto, C. J., & John, O. P. (2020). Validation of the short and extra-short forms of the Big Five Inventory-2 (BFI-2) and their German adaptations. *European Journal of Psychological Assessment, 36*(1), 149-161. https://doi.org/10.1027/1015-5759/a000481

Randall, W. M., & Rickard, N. S. (2017). Personal music listening: A model of emotional outcomes developed through mobile experience sampling. *Music Perception: An Interdisciplinary Journal*, *34*(5), 501-514. https://doi.org/10.1525/mp.2017.34.5.501

Rauthmann, J., F. (2021). Capturing interactions, correlations, fits, and transactions: A person-environment relations model. In J. F. Rauthmann (Ed.), *The Handbook of Personality Dynamics and Processes* (pp., 427-522). Elsevier. https://doi.org/10.1016/B978-0-12-813995-0.00018-2

Rauthmann, J. F., Gallardo-Pujol, D., Guillaume, E. M., Todd, E., Nave, C. S., Sherman, R. A., Ziegler, M., Jones, A. B., & Funder, D. C. (2014). The Situational Eight DIAMONDS: A taxonomy of major dimensions of situation characteristics. Journal of *Personality and Social Psychology, 107*(4), 677–718. https://doi.org/10.1037/a0037250

Rentfrow, P. J., & Gosling, S. D. (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), 1236-1256. https://doi.org/10.1037/0022-3514.84.6.1236

Rights, J. D., & Sterba, S. K. (2019). Quantifying explained variance in multilevel models: An integrative framework for defining R-squared measures. *Psychological Methods, 24*(3), 309-338. https://doi.org/10.1037/met0000184

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178. https://doi.org/10.1037/h007771

Saarikallio, S., & Erkkilä, J. (2007). The role of music in adolescents' mood regulation. *Psychology of Music*, *35*(1), 88-109. https://doi.org/10.1177/0305735607068889

Sakka, L. S., & Juslin, P. N. (2018). Emotion regulation with music in depressed and non-depressed individuals: Goals, strategies, and mechanisms. *Music & Science*, *1*. https://doi.org/10.1177/2059204318755023

Schäfer, T., & Mehlhorn, C. (2017). Can personality traits predict musical style preferences? A meta-analysis. *Personality and Individual Differences, 116*, 265-273. https://doi.org/10.1016/j.paid.2017.04.061

Schäfer, T., Sedlmeier, P., Städtler, C., & Huron, D. (2013). The psychological functions of music listening. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00511

Scherbaum, C. A., & Ferreter, J. M. (2009). Estimating statistical power and required sample sizes for organizational research using multilevel modeling. *Organizational Research Methods, 12*(2), 347-367. https://doi.org/10.1177/1094428107308906

Schoedel, R., Kunz, F., Bergmann, M., Bemmann, F., Bühner, M., & Sust, L. (2023). Snapshots of Daily Life: Situations Investigated Through the Lens of Smartphone *Sensing. Journal of Personality and Social Psychology.* Advance online publication. https://doi.org/10.1037/pspp0000469

Shaw, M., Rights, J. D., Sterba, S. K., & Flake, J. K. (2020). *r2mlm: R-Suqared Measures for Multilevel Models*. PsyArXiv. https://doi.org/10.31234/osf.io/xc4sv.

Sloboda, J. A., & O'Neill, S. A. (2001). Emotions in everyday listening to music. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and Emotion: Theory and Research (pp. 415-429)*. Oxford University Press.

Sloboda, J. A., O'Neill, S. A., & Ivaldi, A. (2001). Functions of music in everyday life: An exploratory study using the Experience Sampling Method. *Musicae Scientiae*, *5*(1), 9-32. https://doi.org/10.1177/102986491140117 4

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An Introduction to Basic and Advanced Multilevel Modeling.* Sage Publishing.

Snyder, M., & Ickes, W. (1985). Personality and social behavior. In G. Lindzey & E. Aronson (Eds.), *Handbook of Social Psychology* (Vol. 3, pp. 883–947).

SAA: Society for Ambulatory Assessment. (2018). *Society for Ambulatory Assessment Bylaws* (n.p.). https://ambulatory-assessment.org/bylaws/

Soto, C. J., & John, O. P. (2017). Short and extra-short forms of the Big Five Inventory–2: The BFI-2-S and BFI-2-XS. *Journal of Research in Personality, 68*, 69-81. https://doi.org/10.1016/j.jrp.2017.02.004

Spotify. (2023). Get Track's Audio Features, *Spotify for Developers*. https://developer.spotify.com/documentation/web-api/reference/#/operations/get-several-audio-features

Stachl, C., Au, Q., Schoedel, R., Gosling, S. D., Harari, G. M., Buschek, D., Völkel, S. T., Schuwerk, T., Oldemeier, M., Ullmann, T., Hussmann, H., Bischl, B., & Bühner, M. (2020). Predicting personality from patterns of behavior collected with smartphones. *Proceedings of the National Academy of Sciences (PNAS)*, *117*(30), 17680-17687. https://doi.org/10.1073/pnas.1920484117

Stein, N. L., Ornstein, P. A., Tversky, B., & Brainerd, C. (2013). *Memory for Everyday and Emotional Events*. Psychology Press.

Steinmetz, H., Batzdorfer, V., Scherhag, J., & Bosnjak, M. (2022). *The ZPID Lockdown Measures Dataset for Germany* [Data set]. PsychArchives. https://doi.org/10.23668/PSYCHARCHIVES.6676

Sust, L., Stachl, C., Kudchadker, G., Bühner, M., & Schoedel, R. (2023). Personality Computing with Naturalistic Music Listening Behavior: Comparing Audio and Lyrics Preferences. *Collabra: Psychology*, *9*(1). https://doi.org/10.1525/collabra.75214

Swann, W. B. (1987). Identity negotiation: Where two roads meet. *Journal of Personality and Social Psychology, 53*(6), 1038-1051. https://doi.org/10.1037/0022-3514.53.6.1038

Tahlier, M., Miron, A. M., & Rauscher, F. H. (2013). Music choice as a sadness regulation strategy for resolved versus unresolved sad events. *Psychology of Music*, *41*(6), 729-748. https://doi.org/10.1177/0305735612446537

Taruffi, L., & Koelsch, S. (2014). The paradox of music-evoked sadness: An online survey. *PLoS One*, *9*(10): e110490. https://doi.org/10.1371/journal.pone.0110490

Tellegen, A., Watson, D., & Clark, L. A. (1999). On the dimensional and hierarchical structure of affect. *Psychological science, 10*(4), 297-303. https://doi.org/10.1111/1467-9280.00157

ter Bogt, T., Canale, N., Lenzi, M., Vieno, A., & van den Eijnden, R. (2021). Sad music depresses sad adolescents: A listener's profile. *Psychology of Music*, *49*(2), 257-272. https://doi.org/10.1177/0305735619849622

Thoma, M. V., Ryf, S., Mohiyeddini, C., Ehlert, U., & Nater, U. M. (2012). Emotion regulation through listening to music in everyday situations. *Cognition & Emotion*, *26*(3), 550-560. https://doi.org/10.1080/02699931.2011.595390

Ushey, K. (2022). *Renv: Project Environemnts* (Version 0.15.5) [R package]. https://CRAN.R-project.org/package=renv

Van Berkel, N., Ferreira, D., & Kostakos, V. (2017). The experience sampling method on mobile devices. *ACM Computing Surveys, 50*(6), 1-40. https://doi.org/10.1145/3123988

Van den Tol, A. J., & Edwards, J. (2015). Listening to sad music in adverse situations: How music selection strategies relate to self-regulatory goals, listening effects, and mood enhancement. *Psychology of Music*, *43*(4), 473-494. https://doi.org/10.1177/0305735613517410

Van Goethem, A., & Sloboda, J. (2011). The functions of music for affect regulation. *Musicae scientiae, 15*(2), 208-228. https://doi.org/10.1177/1029864911401174

Vella, E. J., & Mills, G. (2017). Personality, uses of music, and music preference: The influence of openness to experience and extraversion. *Psychology of Music*, *45*(3), 338-354. https://doi.org/10.1177/0305735616658957

Volokhin, S., & Agichtein, E. (2018). Understanding music listening intents during daily activities with implications for contextual music recommendation. In C. Shah, N. J. Belkin, K. Byström, J. Huang, & F. Scholer (Eds.), *Proceedings of the 2018 Conference on Human Information Interaction and Retrieval* (pp. 313-316). Association for Computing Machinery. https://doi.org/10.1145/3176349.3176885

Vuoskoski, J. K., & Eerola, T. (2011). The role of mood and personality in the perception of emotions represented by music. *Cortex, 47*(9), 1099-1106. https://doi.org/10.1016/j.cortex.2011.04.011

Watson, D., & Hubbard, B. (1996). Adaptational style and dispositional structure: Coping in the context of the Five-Factor model. *Journal of Personality*, *64*(4), 737-774. https://doi.org/10.1111/j.1467-6494.1996.tb00943.x

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological bulletin*, *98*(2), 219-235. https://doi.org/10.1037/0033-2909.98.2.219

Wickham, H., & Girlich, M., (2022). *tidyr: Tidy Messy Data* (Version 1.2.1) [R package]. https://tidyr.tidyverse.org

Wickham, H., Girlich, M., Ruiz, E. (2023). *dbplyr: A 'dplyr' Back End for Databases* (Version 2.2.1) [R package]. https://dbplyr.tidyverse.org/

Wrzus, C., & Mehl, M. R. (2015). Lab and/or field? Measuring personality processes and their social consequences. *European Journal of Personality*, *29*(2), 250-271. https://doi.org/10.1002/per.1986

Xue, C., Li, T., Yin, S., Zhu, X., & Tan, Y. (2018). The influence of induced mood on music preference. *Cognitive Processing*, *19*, 517-525. https://doi.org/10.1007/s10339-018-0872-7

Yang, Y. H., & Liu, J. Y. (2013). Quantitative study of music listening behavior in a social and affective context. *IEEE Transactions on Multimedia, 15*(6), 1304-1315. https://doi.org/10.1109/TMM.2013.2265078

Yang, Y. H., & Teng, Y. C. (2015). Quantitative study of music listening behavior in a smartphone context. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *5*(3), 1-30. https://doi.org/10.1145/2738220

Zacher, H., & Rudolph, C. W. (2021). Individual differences and changes in subjective wellbeing during the early stages of the COVID-19 pandemic. *American Psychologist*, *76*(1), 50-62. https://doi.org/10.1037/amp0000702

Zillmann, D. (2015). Mood management: Using entertainment to full advantage. In L. Donohew, H. E. Sypher, & E. T. Higgins (Eds.), *Communication, Social Cognition, and Affect* (pp. 147-171). Lawrence Erlbaum Associates, Inc.

5 General Discussion

The present dissertation employed smartphone sensing to investigate individual differences in natural music-listening behavior in relation to psychological constructs.

The first study focused on overall music preferences quantified via various audio and lyrics characteristics and habitual listening behaviors to predict the Big Five dimensions of personality in a machine-learning framework. Out-of-sample prediction performances showed that only the domain of Openness was successfully predicted, while Conscientiousness, along with several personality facets, exhibited small to moderate non-significant prediction performances. Thereby, audio- and lyrics-based music preferences contributed distinctly to personality predictions, and the most predictive single music preference variables displayed predominantly trait-congruent associations.

The second study assessed momentary music preferences and modeled them from enduring personality traits, concurring mood states, and their respective interactions in a multilevel regression framework. Based on the literature, it was expected that the musical valence and energy preferred in a given moment should be congruent with listeners' Big Five personality domains and their mood valence and arousal, with personality domains moderating mood-congruency. However, the models revealed that personality and mood explained only a small proportion of variance in music preferences, with only one significant but weak mood-congruency effect for the arousal dimension. Beyond that, the models failed to replicate the expected trait- and state-congruent associations, which may be related to the limited statistical power of the study.

While these findings were discussed in detail in the respective articles above, it should be noted that both studies of this dissertation obtained fewer and weaker effects than past research relying on traditional assessment approaches such as questionnaires or listening experiments (e.g., Bonneville-Roussy et al., 2013; Chen et al., 2007; Greenberg et al., 2016, 2022; Rentfrow & Gosling, 2003; Taruffi & Koelsch, 2014). This discrepancy in results may indicate that natural music-listening behavior is more complex than self-reported music preferences and related to many more factors not included in this dissertation. However, it should also be mentioned that recent steps of digitalization have not only changed the way researchers can assess music listening but also the way listeners consume music. For example, portable listening devices in combination with Internet-based streaming services have increased the freedom of choice of where and when to play what music (e.g., Bull, 2005; Krause & North, 2016; North et al., 2004), resulting in higher quantity and diversity in music consumption (Datta et al., 2018). In turn, the role of person variables such as personality and mood may have changed with these new listening occasions.

5.1 Methodological Contributions and Future Directions

The present dissertation employed state-of-the-art methodological approaches from computational and statistical sciences to efficiently collect, numerically represent, and adequately model natural music-listening behavior.

First and foremost, the dissertation demonstrated that smartphones are a promising tool for efficiently assessing music-listening behavior in an ecologically valid and longitudinal manner. While collecting behavioral music-listening data in the field has long been practically infeasible (see Baumeister et al., 2007; Funder, 2001), the rise of smartphones as popular music devices has paved the way for the ambulatory assessment of listening behaviors (see Conner & Mehl, 2015; Wrzus & Mehl, 2015). More specifically, passive smartphone sensing (Harari et al., 2016; Miller, 2012) now allows for unobtrusively collecting music-listening records, alongside other usage behaviors and sensor data, from participants' private smartphones, as evident in both empirical studies of this dissertation. In addition, smartphones can administer event-triggered ES (Conner et al., 2007; Van Berkel et al., 2017), actively asking participants to answer short questionnaires whenever they listen to music, as shown in the second study. This combination of passive and active ambulatory assessments enables the study of natural music-listening behavior in relation to the listener and their listening context, covering all three components of the personality triad, namely the behavior, the person, and the situation (Funder, 2006; Rauthmann, 2021). However, while this dissertation considered the behavior (i.e., overall and momentary music preferences) and the person (i.e., personality traits and mood states) on stable and momentary levels, it neglected the situation component (see Rauthmann, 2021). Hence, to exploit the full potential of smartphone-based ambulatory assessments, future studies should also collect situational information, which may, for example, shed light on the functionality of music listening on smartphones (e.g., as background noise while commuting, see Chamorro-Premuzic & Furnham, 2007; Schäfer et al., 2013). For this purpose, studies may collect objective situation cues via smartphone sensing (e.g., listening location from GPS sensors) and subjective situation perceptions via ES (e.g., perceived Sociality of a listening situation; see Rauthmann et al., 2014) as previously suggested by Schoedel et al. (2023).

Second, this dissertation proposed an automated preprocessing pipeline for extracting psychologically meaningful variables from smartphone-sensed digital listening records. In particular, both articles focused on music preferences and quantified them via the musical characteristics of the songs played by participants. In contrast to commonly used genre categorizations (e.g., Anderson et al., 2021; Rentfrow & Gosling, 2003), which represent inconsistent meta-information about musical pieces like their geographical origin (e.g., Latin) or stereotypes associated with the artist (e.g., Punk, Aucouturier & Pachet, 2003; Rentfrow & Gosling, 2007), the musical characteristics proposed here are intrinsic to the songs' melodies and lyrics. They were obtained via music information retrieval by Spotify.com (Spotify, 2022) and NLP conducted by the authors (see the first study), highlighting the role of machine learning in data preprocessing. Unlike human labeling used in the past (e.g., Greenberg et al., 2016; Knobloch & Zillman, 2002), this automated annotation allowed for an unbiased and efficient representation of music preferences from large samples of natural listening data (cf. Juslin et al., 2014; Vuoskoski & Eerola, 2011). When aggregating music preferences across played songs in the sensed listening records, the first study presented above considered the entire study duration, building average music preferences, while the second study created momentary music preferences over 30-minute time windows. As participants in the second study exhibited considerable intraindividual fluctuation in music choices (see also Greb et al., 2019), it seems worthwhile to keep studying music listening at more granular levels. In doing so, future researchers may also choose different short-scale time frames and, for example, consider music preferences on a daily or song-level basis. Beyond music preferences, digital listening records from smartphones can also be assessed for habitual listening behaviors (Greenberg & Rentfrow, 2017). While the first study considered some habitual behaviors such as listening durations, future studies may want to investigate more sophisticated variables like daily variations in listening or skipping behaviors (see Anderson et al., 2021).

Third, this dissertation mimicked the cycle of empirical research with its two empirical studies integrating two complementary modeling approaches (Mahmoodi et al., 2017; Yarkoni & Westfall, 2017). The first study was purely exploratory, using flexible machine-learning algorithms to make out-of-sample predictions from a large predictor space (see Breiman, 2001). In contrast, the second study tested preregistered hypotheses with inferential multilevel models, taking into account the data structure of music-listening events nested within persons (see Bates et al., 2015). While both of these approaches aimed at doing justice to the complex structure of natural music-listening data, future studies should integrate them to better understand the dynamics at play. Multilevel machine learning could help to capture non-linear patterns among a multitude of music-listening variables while also disentangling within- and between-person effects.

In sum, the methodological approaches presented in this dissertation may be extended in numerous ways in the future to gain new insights into natural music-listening behavior. Such insights are not only of theoretical value for personality science but may also have practical implications for the improvement of music recommender systems. While automated recommendations have gained importance in the face of millions of songs to choose from, recommender systems usually neglect contextual aspects of the listener and the situation and are, hence, often unsatisfactory (Schedl et al., 2018). Accordingly, researchers in human-computer interaction have started to develop concepts for user-centered and context-aware recommender systems (e.g., Gillhofer & Schedl, 2015; Lozano Murciego et al., 2021) that may benefit from empirical findings on natural music-listening behavior in personality science.

5.2 Limitations of the Smartphone-Sensing Approach

Beyond the study-specific limitations reviewed in the respective articles above, the approach of administering passive ambulatory assessments via smartphones presents some challenges, both for investigating music-listening behavior, in particular, and for psychological research, in general.

5.2.1 Assessing Music Listening on Smartphones

The collection of music-listening data via smartphone sensing is subject to two important limitations. First, investigating only the music-listening events occurring on smartphones can introduce a sampling bias if participants regularly use other music devices as well. That is because music-listening behavior may systematically differ between devices for several reasons. For instance, different types of devices vary in their song selection, ranging from limited choice among pre-purchased physical records on analog devices to seemingly endless choice on web-enabled digital devices. Accordingly, music devices also differ in the way music is typically chosen, whereby streaming music on smartphones is particularly suitable for discovering new music (Krause & Brown, 2021). Furthermore, music devices may vary with regard to the situations in which they are most commonly used. As an example, smartphones were previously found to mainly serve music-listening purposes while commuting (Kuch & Wöllner, 2021), while stationary devices like notebooks may serve for playing music during work. As initial support for this notion, past researchers showed that the time of day relates to the use of different music-listening devices (Krause et al., 2015). In extension to this limitation, not everybody uses smartphones for music listening, introducing a complementary sampling bias on the person level. More specifically, demographic variables like age and personality traits may determine who plays music on the smartphone (Krause et al., 2015; Krause & Brown, 2021).

A second limitation of sensing music-listening data in the field is the lack of control over whether the played music was actually chosen by participants themselves. On the one hand, persons uninvolved in the study, such as friends at a party or family members during a car ride, may select songs by directly accessing participants' smartphones or through requests. On the other hand, music apps suggest songs via editorial or algorithmic playlists and provide a shuffle mode, allowing listeners to play songs without consciously selecting them. Such listening instances where music is not actively chosen by participants may introduce noise to the data and obscure psychological effects.

5.2.2 Conducting Smartphone-Sensing Studies

A more general line of limitations relates to the practical application of smartphone sensing in psychological studies, regardless of their study scope. First, developing and maintaining smartphone-sensing apps requires intensive interdisciplinary collaboration with researchers from informatics departments (Lazer et al., 2009; Miller, 2012). In particular, the rapid technological turnover of the smartphone industry calls for constant adaption of sensing functionalities and intensive technical study support (Schoedel & Mehl, in press; Wrzus & Mehl, 2015). While there are several open (e.g., AWARE, Ferreira et al., 2015) and commercial sensing app solutions (e.g., EARS, Lind et al., 2018), these do not offer the same capabilities as custom-designed apps and, for example, usually cannot log music-listening records (Sust et al., 2023). Second, researchers face multiple hurdles from an ethics and data protection perspective when planning smartphone-sensing studies (Harari, 2020; Harari et al., 2016, Miller, 2012). On the one hand, participants must be carefully informed about the data collected and should ideally be granted control over what data types they share (Beierle et al., 2019; Bemmann et al., 2022; Harari, 2020). On the other hand, the collected data must be handled with great caution as they contain highly privacy-sensitive information (e.g., GPS data), which can hardly be fully anonymized (Gasson et al., 2011; Lazer et al., 2009; Miller, 2012). Hence, researchers must take extensive measures to maintain data privacy and security, both in the development of sensing apps (e.g., by saving only aggregates instead of raw data through on-device preprocessing) and in data storage concepts (e.g., by using local servers, Beierle et al., 2018, 2019; Bemmann & Buschek, 2020). Third, the privacy-invasive nature of smartphone sensing makes it difficult to recruit study participants (Keusch et al., 2019; Miller, 2012). As a consequence, sample sizes usually remain small, and selectivity biases preclude people with strong privacy concerns from participating (Keusch et al., 2019; Wenz et al., 2019). Fourth, another source of selectivity biases originates from the technical requirements of sensing apps for participants' smartphones because most apps only work on phones running the Android operating system, precluding people owning iOS devices (Schoedel & Mehl, in press; Sust et al., 2023). Finally, smartphone sensing produces high volumes of unstructured digital data (e.g., time-stamped usage events), which require extensive preprocessing efforts to obtain meaningful variables as evident in both studies of this dissertation. For this task, psychology researchers must acquire considerable technical and statistical know-how (Miller, 2012; Schoedel & Mehl, in press; Yarkoni, 2012). Furthermore, extensive preprocessing pipelines introduce many researcher degrees of freedom as they require a large number of analytical decisions (see Schoedel et al., 2020). To conclude, smartphone-sensing studies are complex and time-consuming to conduct due to administrative and technical challenges.

For the same reasons, smartphone-sensing studies can be challenging to align with the principles of open science (Wrzus & Schoedel, 2023). For example, preregistrations may be difficult to implement because concrete data preprocessing steps are often not foreseeable due to sensing bugs. Here, it can be helpful to draft preregistrations at a lower level of detail and to include forks for different analysis paths, as was done in the second study of this dissertation. For full transparency, processing decisions that were not preregistered in detail should then be documented thoroughly in the methods section or, if that is not possible due to paper scope, in the supplemental material (Wrzus & Schoedel, 2023). Finally, providing open code is essential for ensuring transparency amongst the unlimited possibilities of aggregating smartphone-sensing data (see Schoedel et al., 2020). Another challenge to openness in smartphone-sensing studies is that usually not all data can be made fully public (Wrzus & Schoedel, 2023). More specifically, raw smartphone-sensing logs contain privacy-sensitive information that cannot be shared because of the legal boundaries of data privacy. In addition, raw data are often enriched with third-party external data (e.g., song lyrics in the first study), which cannot be published due to copyright issues. Nevertheless, sensing studies should at least provide the aggregated data for formal analysis as done in the two articles of this dissertation.

5.3 Conclusion

The present dissertation employed smartphones to investigate inter- and intraindividual differences in natural music-listening behavior. It comprises two empirical studies relating overall and momentary music preferences to enduring personality traits and momentary mood states. These studies applied state-of-the-art methodological approaches to efficiently collect, numerically represent, and jointly model music-listening data.

First, the dissertation demonstrated that smartphone sensing has become a useful tool for investigating music listening in an ecologically valid manner across time, replacing traditional approaches like self-report questionnaires and listening experiments. Furthermore, it showed how integrating passive and active ambulatory assessment in event-triggered ES can enrich sensed music-listening records with self-reported contextual data. Second, the dissertation illustrated how digital music-listening records can be aggregated to preference variables, using the intrinsic musical characteristics of played songs' melodies and lyrics in an automated manner without relying on biased manual annotation. Third, the studies of this dissertation exemplified the empirical research cycle, starting with exploratory machine-learning predictions and following up with confirmatory inferential multilevel regressions as complementary modeling approaches.

However, both empirical studies also demonstrated the limitations of smartphone-sensing studies for investigating music-listening behavior and for psychological research in general, pointing out administrative and technical challenges. Nevertheless, the methodological framework presented in this dissertation provided new insights and corroborated past findings on music-listening behavior. It may be extended in multiple ways to advance theory building in psychology, but also to derive practical implications, for example, for improving automated music recommendations.

5.4 References

Anderson, I., Gil, S., Gibson, C., Wolf, S., Shapiro, W., Semerci, O., & Greenberg, D. M. (2021). Just the way you are: Linking music listening on Spotify and personality. *Social Psychological and Personality Science, 12*(4), 561-572. https://doi.org/10.1177/1948550620923228

Aucouturier, J. J., & Pachet, F. (2003). Representing musical genre: A state of the art. *Journal of New Music Research, 32*(1), 83-93. https://doi.org/10.1076/jnmr.32.1.83.16801

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015b). Fitting linear mixed effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science*, *2*(4), 396-403. https://doi.org/10.1111/j.1745-6916.2007.00051.x

Beierle, F., Tran, V. T., Allemand, M., Neff, P., Schlee, W., Probst, T., Pryss, R., & Zimmermann, J. (2018). Context data categories and privacy model for mobile data collection apps. *Procedia Computer Science, 134*, 18-25. https://doi.org/10.1016/j.procs.2018.07.139

Beierle, F., Tran, V. T., Allemand, M., Neff, P., Schlee, W., Probst, T., Zimmermann, J., & Pryss, R. (2019). What data are smartphone users willing to share with researchers?. *Journal of Ambient Intelligence and Humanized Computing, 11*, 2277-2289. https://10.1007/s12652-019-01355-6

Bemmann, F., & Buschek, D. (2020). LanguageLogger: A Mobile Keyboard Application for Studying Language Use in Everyday Text Communication in the Wild. In J. Nichols (Ed.), *Proceedings of the ACM on Human-Computer Interaction*, *4* (pp. 1–24). Association for Computing Machinery. https://doi.org/10.1145/3397872

Bemmann, F., Windl, M., Erbe, J., Mayer, S., & Hussmann, H. (2022). The influence of transparency and control on the willingness of data sharing in adaptive mobile apps. In J. Nichols (Ed.), *Proceedings of the ACM on Human-Computer Interaction, 6* (pp. 1-26). Association for Computing Machinery. https://doi.org/10.1145/3546724

Bonneville-Roussy, A., Rentfrow, P. J., Xu, M. K., & Potter, J. (2013). Music through the ages: Trends in musical engagement and preferences from adolescence through middle adulthood. *Journal of Personality and Social Psychology, 105*(4), 703. https://doi.org/10.1037/a0033770

Bull, M. (2005). No dead air! The iPod and the culture of mobile listening. *Leisure Studies*, *24*(4), 343-355. https://doi.org/10.1080/0261436052000330447

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5-32. https://doi.org/10.1023/A:1010933404324

Chamorro-Premuzic, T., & Furnham, A. (2007). Personality and music: Can traits explain how people use music in everyday life?. *British Journal of Psychology, 98*(2), 175-185. https://doi.org/10.1348/000712606X111177

Chen, L., Zhou, S., & Bryant, J. (2007). Temporal changes in mood repair through music consumption: Effects of mood, mood salience, and individual differences. *Media Psychology*, *9*(3), 695-713. https://doi.org/10.1080/15213260701283293

Conner, T. S., Barrett, L. F., Tugade, M. M., & Tennen, H. (2007). Idiographic personality: The theory and practice of experience sampling. In R. W. Robbins, R. C. Fraley, & R. F. Krieger (Eds.), *Handbook of Research Methods in Personality Psychology* (pp. 79-96). Guilford Press.

Conner, T. S., & Mehl, M. R. (2015). Ambulatory assessment: Methods for studying everyday life. In R. A. Scott & S. M. Kosslyn (Eds.), *Emerging Trends in the Social and Behavioral Sciences* (1st ed., pp. 1-15). Wiley. https://doi.org/10.1002/9781118900772.etrds0010

Datta, H., Knox, G., & Bronnenberg, B. J. (2018). Changing their tune: How consumers' adoption of online streaming affects music consumption and discovery. *Marketing Science*, *37*(1), 5-21. https://doi.org/10.1287/mksc.2017.1051

Ferreira, D., Kostakos, V., & Dey, A. K. (2015). AWARE: Mobile context instrumentation framework. *Frontiers in ICT, 2*. https://doi.org/10.3389/fict.2015.00006

Funder, D. C. (2001). Personality. *Annual Review of Psychology, 52*, 197-221. https://doi.org/10.1146/annurev.psych.52.1.197

Funder, D. C. (2006). Towards a resolution of the personality triad: Persons, situations, and behaviors. *Journal of Research in Personality, 40*(1), 21-34. https://doi.org/10.1016/j.jrp.2005.08.003

Gasson, M. N., Kosta, E., Royer, D., Meints, M., & Warwick, K. (2011). Normality mining: Privacy implications of behavioral profiles drawn from GPS enabled mobile phones. *IEEE Transactions on Systems, Man, and Cybernetics—Part C: Applications and Reviews, 41*, 251–261. https://doi.org/10.1109/TSMCC.2010.2071381.

Gillhofer, M., & Schedl, M. (2015). Iron maiden while jogging, debussy for dinner? An analysis of music listening behavior in context. In X. He, S. Luo, D. Tao, C. Xu, J. Yang, M. A. Hasan (Eds.), *Proceedings of the 21st International Conference on MultiMedia Modeling* (pp. 380-391). Springer International Publishing.

Greb, F., Steffens, J., & Schlotz, W. (2019). Modeling music-selection behavior in everyday life: a multilevel statistical learning approach and mediation analysis of experience sampling data. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.00390

Greenberg, D. M., Kosinski, M., Stillwell, D. J., Monteiro, B. L., Levitin, D. J., & Rentfrow, P. J. (2016). The song is you: Preferences for musical attribute dimensions reflect personality. *Social Psychological and Personality Science, 7*(6), 597-605. https://doi.org/10.1177/1948550616641473

Greenberg, D. M., & Rentfrow, P. J. (2017). Music and big data: A new frontier. *Current Opinion in Behavioral Sciences, 18*, 50-56. https://doi.org/10.1016/j.cobeha.2017.07.007

Greenberg, D. M., Wride, S. J., Snowden, D. A., Spathis, D., Potter, J., & Rentfrow, P. J. (2022). Universals and variations in musical preferences: A study of preferential reactions to Western music in 53 countries. *Journal of Personality and Social Psychology, 122*(2), 286-309. https://doi.org/10.1037/pspp0000397

Harari, G. M. (2020). A process-oriented approach to respecting privacy in the context of mobile phone tracking. *Current Opinion in Psychology, 31*, 141-147. https://doi.org/10.1016/j.copsyc.2019.09.007

Harari, G. M., Lane, N. D., Wang, R., Crosier, B. S., Campbell, A. T., & Gosling, S. D. (2016). Using smartphones to collect behavioral data in psychological science: Opportunities, practical considerations, and challenges. *Perspectives on Psychological Science*, *11*(6), 838-854. https://doi.org/10.1177/1745691616650285

Juslin, P. N., Harmat, L., & Eerola, T. (2014). What makes music emotionally significant? Exploring the underlying mechanisms. *Psychology of Music, 42*(4), 599-623. https://doi.org/10.1177/0305735613484548

Keusch, F., Struminskaya, B., Antoun, C., Couper, M. P., & Kreuter, F. (2019). Willingness to participate in passive mobile data collection. *Public Opinion Quarterly, 83*(S1), 210-235. https://doi.org/10.1093/poq/nfz007

Knobloch, S., & Zillmann, D. (2002). Mood management via the digital jukebox. *Journal of Communication*, *52*(2), 351-366. https://doi.org/10.1111/j.1460-2466.2002.tb02549.x

Krause, A. E., & Brown, S. C. (2021). A uses and gratifications approach to considering the music formats that people use most often. *Psychology of Music, 49*(3), 547-566. https://doi.org/10.1177/0305735619880608

Krause, A. E., & North, A. C. (2016). Music listening in everyday life: Devices, selection methods, and digital technology. *Psychology of Music*, *44*(1), 129-147. https://doi.org/10.1177/0305735614559065

Krause, A. E., North, A. C., & Hewitt, L. Y. (2015). Music-listening in everyday life: Devices and choice. *Psychology of Music, 43*(2), 155-170. https://doi.org/10.1177/0305735613496860

Kuch, M., & Wöllner, C. (2021). On the Move: Principal Components of the Functions and Experiences of Mobile Music Listening. *Music & Science, 4*, 1-14. https://doi.org/10.1177/20592043211032852

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., & Alstyne, M. V. (2009). Computational social science. *Science, 323* (5915), 721–723. https://doi.org/10.1126/science.1167742

Lind, M. N., Byrne, M. L., Wicks, G., Smidt, A. M., & Allen, N. B. (2018). The effortless assessment of risk states (EARS) tool: An interpersonal approach to mobile sensing. *JMIR Mental Health*, *5*(3), e10334. https://doi.org/10.2196/10334

Mahmoodi, J., Leckelt, M., van Zalk, M., Geukes, K., & Back, M. (2017). Big data Approaches in social and behavioral science: Four key trade-offs and a call for integration. *Current Opinion in Behavioral Sciences, 18*, 57–62. https://doi.org/10.1016/j.cobeha.2017.07.001

Miller, G. (2012). The smartphone psychology manifesto. *Perspectives on Psychological Science, 7* (3), 221–237. https://doi.org/10.1177/1745691612441215

Lozano Murciego, Á., Jiménez-Bravo, D. M., Valera Román, A., De Paz Santana, J. F., & Moreno-García, M. N. (2021). Context-aware recommender systems in the music domain: A systematic literature review. *Electronics, 10*(13), 1555. https://doi.org/10.3390/electronics10131555

North, A. C., Hargreaves, D. J., & Hargreaves, J. J. (2004). Uses of music in everyday life. *Music Perception*, *22*(1), 41-77. https://doi.org/10.1525/mp.2004.22.1.41

Rauthmann, J., F. (2021). Capturing interactions, correlations, fits, and transactions: A person-environment relations model. In J. F. Rauthmann (Ed.), *The handbook of personality dynamics and processes* (pp., 427-522). Elsevier.

Rauthmann, J. F., Gallardo-Pujol, D., Guillaume, E. M., Todd, E., Nave, C. S., Sherman, R. A., Ziegler, M., Jones, A. B., & Funder, D. C. (2014). The Situational Eight DIAMONDS: A taxonomy of major dimensions of situation characteristics. *Journal of Personality and Social Psychology*, *107*(4), 677–718. https://doi.org/10.1037/a0037250

Rentfrow, P. J., & Gosling, S. D. (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), 1236-1256. https://doi.org/10.1037/0022-3514.84.6.1236

Rentfrow, P. J., & Gosling, S. D. (2007). The content and validity of music-genre stereotypes among college students. *Psychology of Music, 35*(2), 306-326. https://doi.org/10.1177/0305735607070382

Schäfer, T., Sedlmeier, P., Städtler, C., & Huron, D. (2013). The psychological functions of music listening. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00511

Schedl, M., Zamani, H., Chen, C. W., Deldjoo, Y., & Elahi, M. (2018). Current challenges and visions in music recommender systems research. *International Journal of Multimedia Information Retrieval, 7*, 95-116. https://doi.org/10.1007/s13735-018-0154-2

Schoedel, R., Kunz, F., Bergmann, M., Bemmann, F., Bühner, M., & Sust, L. (2023). Snapshots of Daily Life: Situations Investigated Through the Lens of Smartphone *Sensing. Journal of Personality and Social Psychology.* Advance online publication. https://doi.org/10.1037/pspp0000469

Schoedel, R., & Mehl, M. R. (in press). Mobile sensing methods. In H. T. Reis, T. West, & C. M. Judd (Eds.), *Handbook of Research Methods in Social and Personality Psychology* (3rd ed.). Cambridge University Press.

Schoedel, R., Pargent, F., Au, Q., Völkel, S. T., Schuwerk, T., Bühner, M., & Stachl, C. (2020). To challenge the morning lark and the night owl: Using smartphone sensing data to investigate day–night behaviour patterns. *European Journal of Personality, 34*(5), 733-752. https://doi.org/10.1002/per.2258

Spotify. (2022). Get Track's Audio Features, *Spotify for Developers*. https://developer.spotify.com/documentation/web-api/reference/#/operations/get-several-audio-features

Sust, L., Talaifar, S., & Stachl, C. (2023). Mobile Application Usage in Psychological Research. In Mehl, M. R., Eid, M., Wrzus, C., Harari, G. M., & Ebner-Priemer, U. W. (Eds.), *Mobile Sensing in Psychology: Methods and Applications* (pp. 184-214). Guilford Press.

Taruffi, L., & Koelsch, S. (2014). The paradox of music-evoked sadness: An online survey. *PLoS One*, *9*(10): e110490. https://doi.org/10.1371/journal.pone.0110490

Van Berkel, N., Ferreira, D., & Kostakos, V. (2017). The experience sampling method on mobile devices. *ACM Computing Surveys, 50*(6), 1-40. https://doi.org/10.1145/3123988

Vuoskoski, J. K., & Eerola, T. (2011). The role of mood and personality in the perception of emotions represented by music. *Cortex, 47*(9), 1099-1106. https://doi.org/10.1016/j.cortex.2011.04.011

Wenz, A., Jäckle, A., & Couper, M. P. (2019). Willingness to use mobile technologies for data collection in a probability household panel. *Survey Research Methods, 13*(1), 1-22. https://doi.org/10.18148/srm/2019.v1i1.7298

Wrzus, C., & Mehl, M. R. (2015). Lab and/or field? Measuring personality processes and their social consequences. *European Journal of Personality*, *29*(2), 250-271. https://doi.org/10.1002/per.1986

Wrzus, C., & Schoedel, R. (2023). Transparency and Replicability in Mobile Sensing Research. In Mehl, M. R., Eid, M., Wrzus, C., Harari, G. M., & Ebner-Priemer, U. W. (Eds.), *Mobile Sensing in Psychology: Methods and Applications* (pp. 53-81). Guilford Press.

Yarkoni, T. (2012). Psychoinformatics. *Current Directions in Psychological Science, 21*(6), 391–397. https://doi.org/10.1177/0963721412457362

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science, 12*(6), 1100-1122. https://doi.org/10.1177/1745691617693393