Original Manuscript

Social Psychological and
Personality Science
1-12

DOI: 10.1177/1948550620923228
journals.sagepub.com/home/spp
SAGE

"Just the Way You Are": Linking Music Listening on Spotify and Personality

Ian Anderson[1], Santiago Gil[1], Clay Gibson[1], Scott Wolf[1], Will Shapiro[1], Oguz Semerci[1], and David M. Greenberg[2,3,4]

Abstract

Advances in digital technology have put music libraries at people's fingertips, giving them immediate access to more music than ever before. Here we overcome limitations of prior research by leveraging ecologically valid streaming data: 17.6 million songs and over 662,000 hr of music listened to by 5,808 Spotify users spanning a 3-month period. Building on interactionist theories, we investigated the link between personality traits and music listening behavior, described by an extensive set of 211 mood, genre, demographic, and behavioral metrics. Findings from machine learning showed that the Big Five personality traits are predicted by musical preferences and habitual listening behaviors with moderate to high accuracy. Importantly, our work contrasts a recent self-report-based meta-analysis, which suggested that personality traits play only a small role in musical preferences; rather, we show with big data and advanced machine learning methods that personality is indeed important and warrants continued rigorous investigation.

Keywords

musical preferences, personality, computational social science, Big Five personality model

Today, online services facilitate many of our daily activities, from social interaction (Facebook, Twitter, Snapchat) to information retrieval (Google, Bing, Yahoo) to content consumption (YouTube, Netflix, Spotify). The digital nature of these interactions allows data to be logged precisely and at a scale not possible in previous decades. Companies collect and analyze these data to fuel the services they provide: Google uses user-level search history data to optimize search rankings (Covington et al., 2016; Das et al., 2007; Horling & Kulick, 2009), Netflix leverages user feedback to personalize content recommendations (Koren et al., 2009), and Twitter can use feedback to provide personalized news recommendations (Abel et al., 2011). This rich individual-level information provides social scientists with unprecedented opportunities to advance our understanding of human behavior (Greenberg & Rentfrow, 2017). In this article, we leverage data from Spotify to provide the first ecologically valid behavioral evidence that musical preferences and habitual listening behavior are linked to personality traits.

Recent social psychological, personality, and computational research has developed a basis for successfully predicting human characteristics from digital records (i.e., "digital footprints"). For example, personality traits are accurately predicted by Facebook likes, and computer-generated models have more predictive accuracy than humans (Kosinski et al., 2013; Youyou et al., 2015). Personality is also predicted from language use in Facebook status updates (Schwartz et al., 2013), Twitter posts (Quercia et al., 2011; Skowron et al., 2016), and more subtle behaviors such as patterns of keyboard and mouse use (Khan et al., 2008; Marcus et al., 2006). In this article, we use a similar approach using digital records to address theory and research into musical preferences and habitual listening behavior.

Contemporary research on musical preferences has applied interactionist theories to music (Buss, 1987; Swann et al., 2002), positing that people select musical environments that reflect their psychological traits and needs. Prior research has found converging support for this theory when applied to musical preferences, finding that across multiple methods, samples, and geographic regions, personality is correlated with preferences for features, genres, and styles (Brown, 2012; Dunn et al., 2011; Fricke et al., 2018; Greenberg et al., 2016; Langmeyer et al., 2012; Nave et al., 2018; Rentfrow & Gosling, 2003).

[1] Spotify, New York, NY, USA
[2] Interdisciplinary Department of Social Sciences, Bar-Ilan University, Ramat Gan, Israel
[3] Department of Music, Bar-Ilan University, Ramat Gan, Israel
[4] Department of Psychiatry, University of Cambridge, United Kingdom

Corresponding Author:
Ian Anderson, Spotify, 150 Greenwich Street, 62nd Floor, New York, NY 10007, USA.
Email: iananderson@spotify.com

However, a recent meta-analysis on the personality correlates of musical preferences concluded that the effects of personality are small and "...barely account for individual differences in musical preferences" (Schäfer & Mehlhorn, 2017). The studies in their meta-analysis though had predominantly used self-report methods and basic statistical analyses which are both profound limitations. To help move beyond these limitations, Nave et al. (2018) examined the link between musical preferences and personality using a stimuli-based measure and an artist-specific measure derived from Facebook likes of musical artists. They then applied out-of-sample techniques and least absolute shrinkage and selection operator (lasso) regression to find that preferences accurately predicted personality. However, this study too had methodological limitations. The stimuli-based approach that measured affective reactions to musical excerpts and measuring Facebook likes are artificial in the sense that they do not measure *actual* listening behavior. Although they serve as robust behavioral methods, they do not capture the music that a person listens to day-in and day-out. Therefore, the literature is in need of a study that observes musical preferences "in the wild."

Furthermore, stimuli-based methods and digital behaviors do not capture listening behaviors beyond musical preferences, including habitual listening behaviors. For example, the extent to which someone repeatedly plays a song, listens to new music over old music, and temporal aspects of music listening. Indeed, prior research has explored aspects of musical use (e.g., background listening vs. a primary activity) and its links to personality (Chamorro-Premuzic & Furnham, 2007), but these studies too have been limited by self-report methods and small samples. Therefore, needed in the literature are habitual aspects of music listening captured in the real world. Such observations, which require large amounts of data over time with many people, can now be measured using music streaming data.

Music streaming data pose at least six advantages over other types of digital records from human behavior. First, listening to music involves a significant amount of individual choice (including not just what type of music but when, where, and how to listen) and therefore has the potential to capture subtle information about personal preferences and attitudes. Second, people listen to music across a wide range of situations—socializing, exercising, sleeping—which captures a more complete picture of a person's daily activities and routines. Third, music induces and communicates emotions, evokes autobiographical memories, affects people's moods, and activates brain regions linked to emotion and creativity (Juslin, & Laukka, 2003; Juslin & Västfjäll, 2008; Levitin & Grafton, 2016; Limb & Braun, 2008; McPherson et al., 2016; Rickard, 2004; Salimpoor et al., 2011; Taruffi & Koelsch, 2014; Zentner et al., 2008), providing a large window into a person's emotional life. Fourth, compared to data sources that involve circumstantial and occasional activity, listening to music spans much longer timescales, sometimes extending over an individual's entire daily activity (North et al., 2004; Sloboda et al., 2001), potentially capturing more stable behaviors and traits. Fifth, prior big data studies on personality have often relied on data that may suffer from social desirability biases (e.g., liked Facebook pages and Twitter activity). Since a user's full listening history is never shared publicly and Spotify includes an "incognito mode" which prevents others from knowing what a user is currently playing, music streaming data do not suffer from the same biases. Sixth, prior studies used behavioral data generated by participants in an intermittent and inactive way. For example, most Twitter users are active less than once per day (Greenwood et al., 2016) and visit it for the explicit intent of engaging with social media. Whereas music crosses contextual boundaries including passive engagement and spans longer time scales, so while the signal may be less clear, it should represent a broader picture of a listener's personality. For these reasons, music streaming data provide a unique lens in which to observe and understand the complexity of human individuality.

Music streaming services can provide moment-to-moment data on the music people listen to, enabling researchers, for the first time, to define musical taste through the accumulation of everyday listening. In this article, we describe the nuances of listening behavior using an extensive set of 211 mood, genre, demographic, and behavioral variables. This more expansive and ecological representation of taste builds upon prior theory and research on music feature preferences (Fricke et al., 2018, 2019; Greenberg et al., 2016) and extends it to habitual listening behaviors. Based on interactionist theories, we hypothesize that real-life musical choices and patterns of interaction with the Spotify music streaming service will be linked to personality traits. Since this study is descriptive, we do not have an explicit blueprint of hypotheses; however, we do have general expectations of the patterns between personality traits and music listening behaviors. For example, consistent with previous research, we hypothesize that those who discover more new music or listen to more varied music will score higher on Openness (Greenberg et al., 2016), those who listen to more aggressive music will score lower on Agreeableness (Greenberg et al., 2015).

Method

Participants and Procedure

All participants were registered Spotify users in the United States. To qualify for the survey, participants had to be 18–75 years old and active on the app in the 30 days preceding data collection. Each participant provided informed consent prior to participating in the study, including the option to remove their data from the study prior to publication. Users with no variance in their survey responses or with no streaming in the 30 days prior to data collection were removed, leaving a final sample composed of 5,808 participants (54% male; median age = 26, min age = 18, max age = 75). The selected participants had listened to a combined count of over 17.6 million streams over a 3-month time period, where a stream is defined as a song that was played on Spotify for at least 30 s. This study was given ethical approval by the internal review board (IRB) at Spotify.

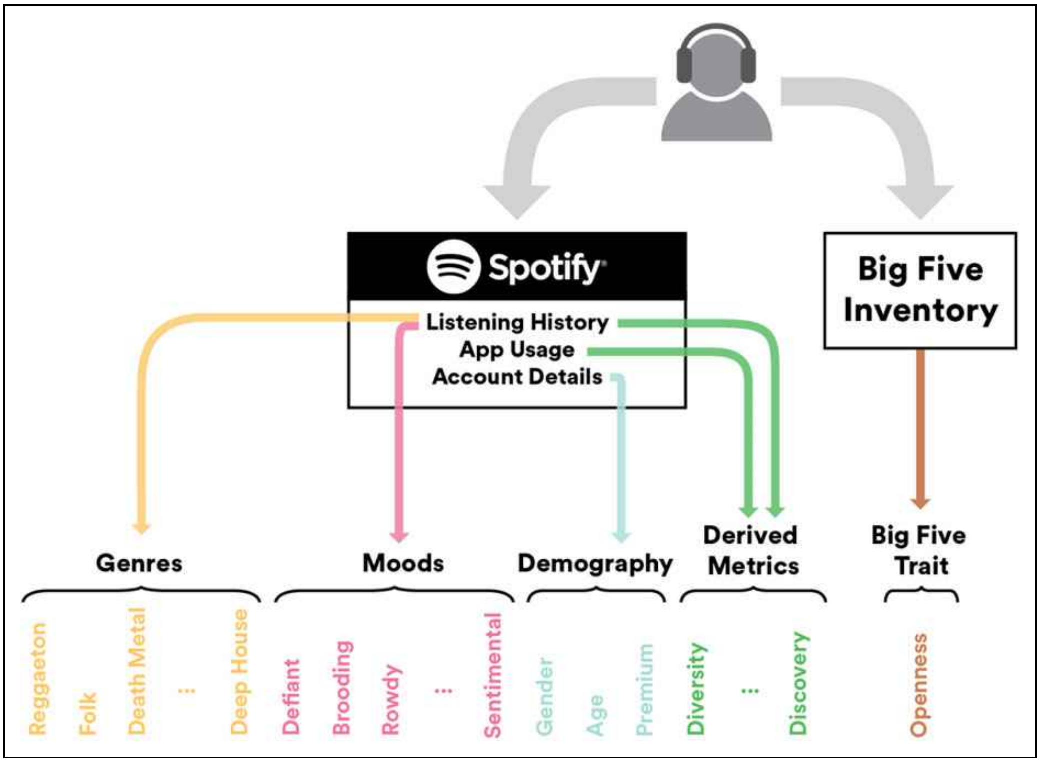

Figure 1. User features and traits independently derived from interactions with the music streaming service and from responses to the Big Five Inventory survey, respectively.

Features

As input to our model, we gathered a variety of data related to users' demography (age and gender), musical taste, and in-app tendencies over a 3-month period from November 2017 to February 2018.

Genre and mood vectors. In prior literature examining the connection between music and personality, a person's musical preferences were derived from self-reported information based on a small set of genres, styles, or stimuli (Greenberg et al., 2015, 2016; North, 2010; Rentfrow & Gosling, 2003). For this analysis, we took users' listening data and mapped it to genre and mood vectors (see Figure 1). Genres are produced internally by Spotify through data curators labeling clusters of artists via a machine-assisted approach that takes into account aggregate listening, acoustic properties, and cultural knowledge (Johnston, 2018). Gracenote, a third-party music metadata service, infers a song's mood by processing the audio signal into features (e.g., rhythm and harmony), then passing these features into a supervised classifier trained to predict the mood from a predefined taxonomy (Summers, 2016). For each user, their streams were mapped to 66 Spotify genres and 25 Gracenote moods then aggregated and normalized to get the percentage of listening from each genre or mood.

Derived metrics. We computed a suite of derived metrics—features constructed using Spotify's data—which quantify aspects of individual behavior on the platform. These included simple aggregations of the user's data, such as the types of platforms used (e.g., Mac, speaker, game console), the total playtime, and the number of playlists created. We also included more complex computations: measures of the diversity of one's musical preferences, their preference for discovering new music, regularity of listening habits, and tendency to listen to music representative of their formative years. Additionally, the audio of the songs listened to over the 3-month period were mapped to a suite of features determined from a supervised learning approach, which served to identify the acoustic profile of each user's listening history. In total, 123 of these features were computed and used in the model. Calculations and definitions of these metrics including discovery, diversity, contextual listening, and audio attributes are detailed in the Supplemental Material (including Table S1).

Product flag. Spotify offers two product versions: Free and Premium. The Free product is available without charge but comes with some feature restrictions. Free users are served advertisements between songs and cannot play tracks on demand when using the mobile app (as of March 2018). In addition, Free product users have a limited number of skips per hour and are unable to download music for off-line listening. The Premium product requires a subscription and removes these limitations, giving users unrestricted access. These differences between the products have a significant impact on how users engage with the application. For this reason, we added an additional feature to indicate which product the user was on during the 3-month period. In our sample, there were 2,974 users of the Free product and 2,834 users of the Premium product.

Measuring Personality

Participants completed the Big Five Inventory (BFI: John & Srivastava, 1999), a 44-item questionnaire measuring the Big Five personality traits: Openness, Conscientiousness, Extraversion, Agreeableness, and Emotional Stability.[1] Each trait is computed by averaging the responses to 8–10 Likert-style questions on a 5-point scale. The BFI has been well tested, and we return Cronbach's αs in excess of .73 for each trait (Openness = .73, Conscientiousness = .80, Extraversion = .85, Agreeableness = .77, Emotional Stability = .84).

Model Selection

Given that predictors in our model have varying distributions, we transformed the numerical values of each predictor to achieve a more standardized distribution. The standardization technique was chosen for each feature based on its distribution. For example, while a log transformation was used for the number of plays in the last 3 months as it spans orders of magnitude, the percentage of streams of Jazz music was transformed using a logit transformation.

Overfitting was a concern when choosing a model, given our sample size and number of predictors (211, including all genre, mood, demographic, and derived metrics). A common approach to avoid overfitting is to use regularized methods such as lasso and ridge regression (Hoerl & Kennard, 1970; Tibshirani, 1996). Lasso regression is known for its interpretability and bias toward lower dimensional models (Tibshirani, 1996) and has been used in previous studies on personality and music (Nave et al., 2018). Ridge regression (Hoerl & Kennard, 1970) uses an alternative that shows better performance when there are high correlations between predictors (Tibshirani, 1996) but does not allow predictors to drop out of the model. Neither model outperforms the other in all situations (Fu, 1998; Tibshirani, 1996). Instead, we chose elastic net regularization regression, which linearly combines[2] both techniques to address concerns of data sets with overly broad feature sets while outperforming either technique individually (Zou & Hastie, 2005). To account for possible nonlinearities between predictors—if personality is expressed differently between Free and Premium, for example—we also used random forest regression for completeness.[3] Both models' hyperparameters were decided based on the minimization of mean of the root-mean-square error (RMSE) using grid search. Although the nonlinear models occasionally outperformed the linear models, the overall performance was similar, so we report the best performance among the two models.

We constructed all regression models using the same set of metrics (see Figure 2) to predict the numerical values of each of the Big Five personality traits independently. For each model, we performed 10-fold cross-validation to test the out-of-sample accuracy of the model; all variable selection and parameter tuning happens within each training set independently. Any model improvement observed when variables are selected prior to cross-validation would represent an unrealistic estimate of true out-of-sample prediction accuracy (Hastie et al., 2009). Following similar papers that predicted personality (Kosinski et al., 2013), we measured prediction accuracy by computing the Pearson correlation between the predicted values and the measured personality traits for the test group of each fold. We report the average correlations across the 10 folds.

Results

Prediction

Mean of the RMSE from 10-fold cross-validation showed moderate to high prediction for each of the Big Five personality traits: .811 for Extraversion, .777 for Emotional Stability, .621 for Agreeableness, .618 for Conscientiousness, and .530 for Openness. Independent regressions were then performed for each trait. Table 1 summarizes our prediction results (rs range from .262 to .374). These results are greater in magnitude than those found in previous research by Nave et al. (2018) that use stimuli-based methods and Facebook likes to assess musical preferences. That our results yielded higher correlations is not surprising since we included metrics that assessed not only musical preferences but also habitual listening behaviors.

Our regression results do outperform univariate correlates, but not substantially. This is due to the cross-validation technique we use in our study. It is important to stress that these prediction accuracies represent a reliable estimation of our ability to predict the personality traits of subjects outside of the current study. Cross-validation is widely used to make regression results more resilient against out-of-sample effects and possibilities of overfitting. Naturally, these will be smaller than correlation values that could be achievable within the sample.

Of the five personality traits, Emotional Stability and Conscientiousness were the two most predictable from our data (rs = .374 and .363, respectively). Since these two traits are significantly related to age and gender, it is not surprising that including these variables with our behavioral metrics yielded significant improvement for our prediction ability (Soto et al., 2011).

Mood and Genres

To gain insight into our models, we have presented the significant correlations for each of the traits in Figure 3. A general observation is that—aside from well-known correlations between demographic characteristics and personality traits like gender and Emotional Stability, and age and Conscientiousness—the largest correlations were for mood and genre information (i.e., features relating to the sonic and emotive aspects of the music; all Pearson correlations between personality and the derived metrics, genres, and moods, are presented in Tables S2–S4 and Figures S1–S3 in the Supplemental Material).

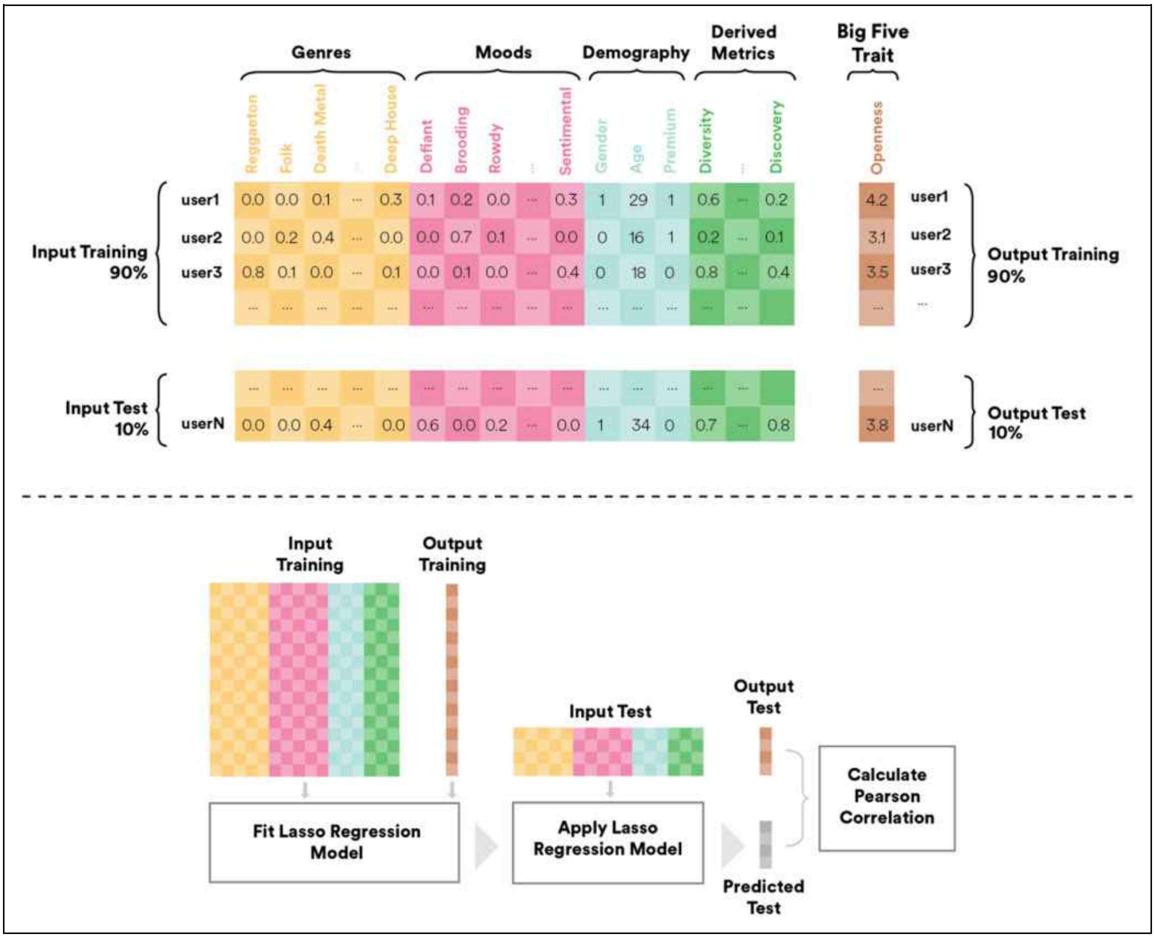

Figure 2. Regression, cross-validation procedure, and output of a correlation between predicted and actual values of each personality trait.

Table 1. Pearson Product–Moment Correlations (Averaged Across 10-Folds) Between Predicted Values From Regression and Actual Values for Each Trait and Product Pair.

Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability
.309 95% CI [.285, .332]	.363 95% CI [.340, .385]	.294 95% CI [.270, .317]	.262 95% CI [.238, .286]	.374 95% CI [.351, .396]

Note. CI = confidence interval.

The specific moods and genre metrics that correlate with each trait show meaningful patterns. Openness was positively correlated with listening to Atmospheric (r = .139, 95% confidence interval (CI) [.114, .165]), Folk (r = .154, 95% CI [.129, .179]), Reggae (r = .085, 95% CI [.059, .110]), or Afropop (r = .112, 95% CI [.086, .137]). The pattern captures a general preference for less popular genres compared to the U.S. population. For mood metrics, Openness was positively correlated with listening to "Sentimental" (e.g., "Freddie Freeloader" by Miles Davis and "April Come She Will" by Simon & Garfunkel; r = .093, 95% CI [.068, .119]) and "Melancholy" music (e.g., "Dust In The Wind" by Kansas and Frank Ocean's "Moon River"; r = .130, 95% CI [.104, .155]).

Emotional Stability correlated positively with Blues (r = .086, 95% CI [.061, 0112]), Old Country (r = .086, 95% CI [.060, .111]), Soul (r = .076, 95% CI [.050, 0101]), and music with "Lively" moods (e.g., "Down On The Corner" by Creedence Clearwater Revival and "Let The Good Times Roll" by Ray Charles; r = .054, 95% CI [.029, .080]), and negatively with "Brooding" (e.g., "Take Care" by Drake and "Karma Police" by Radiohead; r = −.088, 95% CI [−.114, −.063]) or "Defiant" moods (e.g., "Mask Off" by Future and "3005" by Childish Gambino; r = −.082, 95% CI [−.107, −.056]) and music from Indie (r = −.099, 95% CI [−.125, −0.074]), Emo (r = −.155, 95% CI [−.180, −.130]), and Regional Music from Korea (r = −.079, 95% CI [−.105, −.054]). Agreeableness correlated negatively with Punk (r = −.103, 95% CI [−.129,

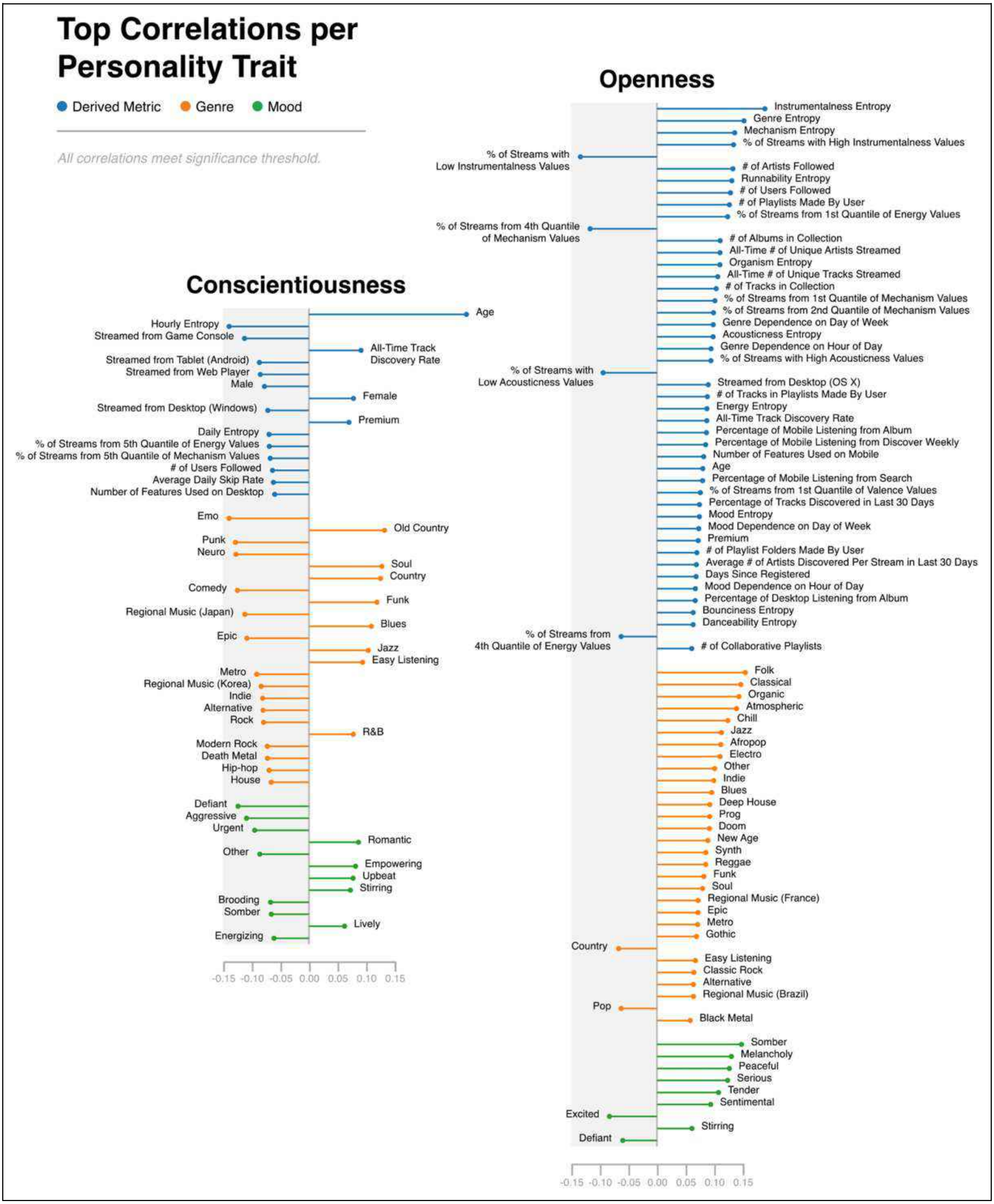

Figure 3. A selection of significant correlations ($ps < 4.7 \times 10^{-5}$) for each personality trait, organized by variable category (blue = derived metric; orange = genre; green = mood). The complete correlation tables and figures are available in Supplemental Material.

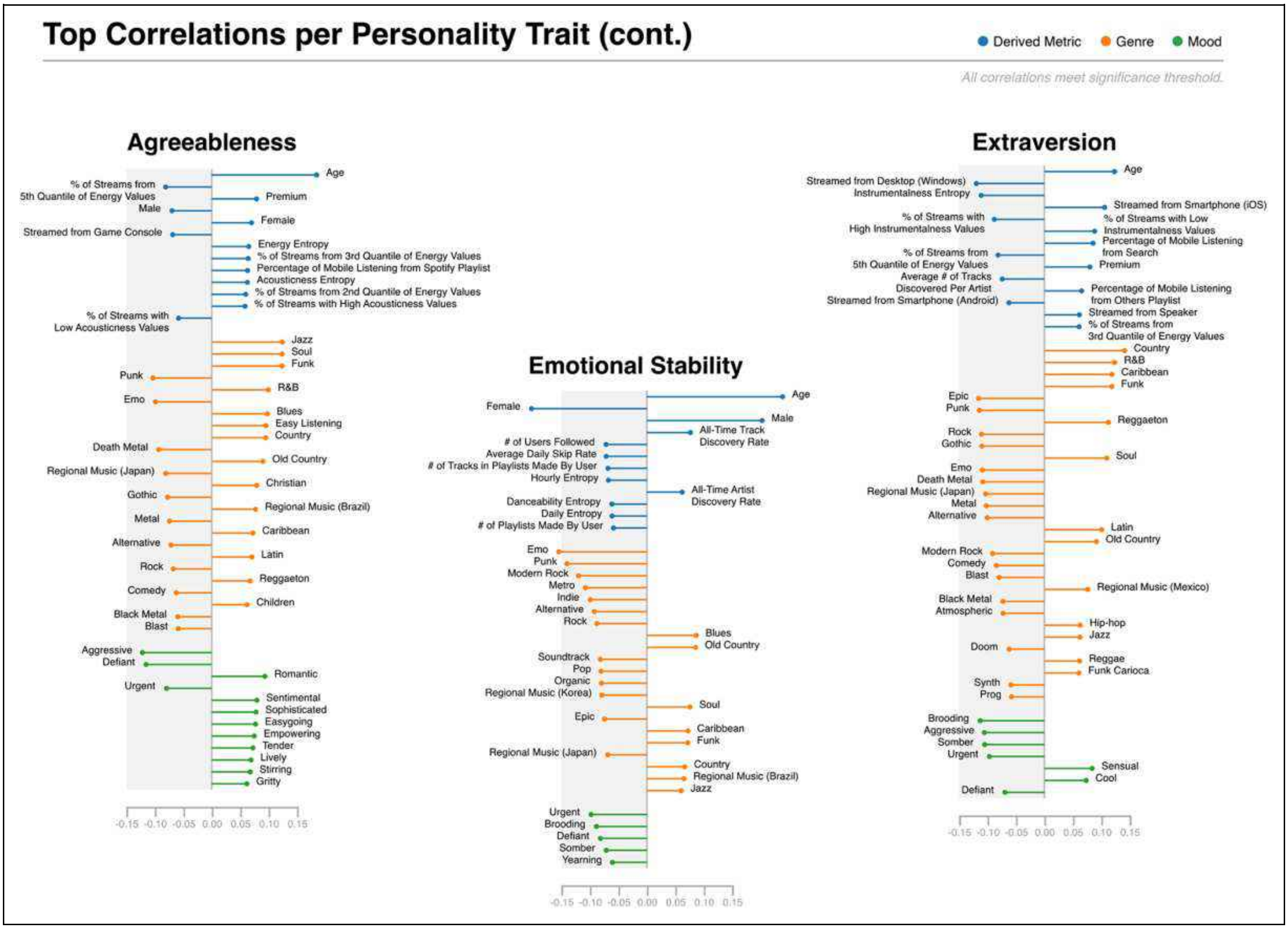

Figure 3. (continued)

−.078]), Death Metal ($r = -.093$, 95% CI [−.119, −.068]), or other "aggressive" music (e.g., "Boss" by Lil Pump and "Last Resort" by Papa Roach; $r = -.122$, 95% CI [−.147, −.097]) and correlated positively with Jazz ($r = .124$, 95% CI [.099, 0.149]), Soul ($r = .124$, 95% CI [.098, .149]), and "Sophisticated" music (e.g., "Fly Me To The Moon" by Frank Sinatra; $r = .078$, 95% CI [.052, .103]).

Conscientiousness was negatively correlated with Rock ($r = -.079$, 95% CI [−.105, −.054]), Comedy ($r = -.125$, 95% CI [−.150, −.100]), and Alternative ($r = -.080$, 95% CI [−.106, −.055]) genres, and "Energizing" (e.g., "Happy" by Pharrell Williams; $r = -.061$, 95% CI [−.087, −.036]) and "Excited" moods (e.g., "I Wanna Dance With Somebody" by Whitney Houston and "California Gurls" by Katy Perry; $r = -.055$, 95% CI [−.081, −.030]). Conscientiousness was positively correlated with Funk ($r = .119$, 95% CI [.094, .144]), Easy Listening ($r = .055$, 95% CI [.029, .081]), or "Romantic" music (e.g., "All of Me" by Billie Holiday and "La vie en rose" by Édith Piaf; $r = .087$, 95% CI [.061, .112]). Extraversion was negatively correlated with listening to Rock ($r = -.111$, 95% CI [−.136, −.085]), Metal ($r = -.102$, 95% CI [−.134, −.083]), and "Urgent" music (e.g., "Locked Out of Heaven" by Bruno Mars, "Cheap Thrills" by Sia; $r = -.097$, 95% CI [−.122, −.071]), and positively correlated with Funk ($r = .118$, 95% CI [.093, .144]), Reggaeton ($r = .112$, 95% CI [.087, .138]), or "Sensual" music (e.g., "LOVE. FEAT. ZACARI." by Kendrick Lamar and "Same Old Love" by Selena Gomez; $r = .084$, 95% CI [.058, .110]).

Derived Metrics

In addition to mood and genre information, we found transparent and theoretically consistent patterns of correlations between our derived metrics and personality traits. Emotional Stability correlated negatively with average skip rate ($r = -.071$, 95% CI [−.097, −.046]), indicating that participants who were more neurotic tended to be more selective of what they listened to at any given moment. Openness correlated positively with all-time track discovery rate ($r = .087$, 95% CI [.061, .112]) and genre entropy ($r = .152$, 95% CI [.127, .177]; see definitions in Supplemental Material), suggesting that users scoring high on Openness are more receptive to exploring different types of music.

We also observed significant trends between personality and how and when users play music. People who scored high on Conscientiousness tended to concentrate their listening to a narrow window of time of day across multiple weeks ($r = -.034$, 95% CI [−.060, −.009]), indicating that they structure their day more rigidly than those who score lower on Conscientiousness. Those who scored higher on Extraversion tended to listen more from others' playlists ($r = .066$, 95% CI [.040, .091]). This leads to several possible explanations: Extroverts may have (1) a greater reliance on their social network's suggestions, (2) a preference for discovering music based on their peer group, or (3) musical preferences driven by group identity. On the other hand, those who listened more to music suggested by Spotify (rather than others) tended to score higher on Agreeableness ($r = .062$, 95% CI [.037, .088]) or Conscientiousness ($r = .056$, 95% CI [.031, .082]), implying a greater likelihood of considering recommendations.

Discussion

Here we investigated the links between human personality and musical listening behavior on the Spotify streaming service. We used metrics from Spotify that characterize the music people listen to in their daily lives, the context in which they do so, and their habitual listening behaviors. Our results showed several main findings. First, the results showed that the metrics from Spotify behavior were moderately to highly predictive of personality. There is no unified standard for benchmarking predictive performance across big data personality studies, but our regression and correlations perform comparably or outperform results from prior studies of this nature (Golbeck et al., 2011; Kosinski et al., 2013; Mairesse et al., 2007; De Montjoye et al., 2013). This suggests that music streaming data are on par or better than services like Facebook and Twitter in its predictive validity of human personality. Second, when compared to prior big data music studies (Nave et al., 2018), our regression results suggest not surprisingly that combining both musical preferences and habitual music behavior (e.g., listening contexts) is more predictive of human personality than relying only on musical preferences. Third, and very importantly, our results provide a stark contrast to the recent meta-analysis that concluded that personality plays little role in musical preferences (Schäfer & Mehlhorn, 2017). Our results using big data and advanced machine learning techniques show the opposite: There is a great deal of information about personality that is communicated through musical preferences. This is an important distinction of how different conclusions can be drawn using big data.

More specifically, our regression results show that Emotional Stability and Conscientiousness are most predictable from music listening behaviors compared to the other Big Five traits. This result may be driven by the way users engage with music streaming platforms. Regression accuracy may be less for Agreeableness and Extraversion because Spotify's applications are limited in opportunities for social interactions and therefore provide less opportunity for agreeable people and extroverts to engage in a way that aligns with their traits. In contrast, people who score low on Emotional Stability (i.e., high on Neuroticism) may select music to regulate their emotions (e.g., searching for music with matching emotional content), and users who score high on Conscientiousness may choose music based on goal-oriented behavior (e.g., study music, workout music).

Our findings are based on cross-sectional and correlational data. We decided to build a model that predicts personality from musical behavior because of the statistical difficulties that would arise from applying data reduction to the more than 200 music usage variables and accounting for variance and individual differences in the millions of streams and hundreds of thousands of listening hours by users, all which were able to be accounted for by making these variables the predictors. Furthermore, our approach is consistent with contemporary research designs in big data music and personality studies, which also used personality traits as outcome variables (Nave et al., 2018). However, it is worth mentioning that there are at least two possible underlying mechanisms at work in the interaction between music and personality. On the one hand, people may seek out music that reflects their personalities as interactionist theories would suggest (Buss, 1987; Rentfrow et al., 2011). Alternatively, people's personalities may be shaped by the music they are exposed and listening to. This is not unprecedented considering longitudinal evidence shows musical training impacts brain maturation (Habibi et al., 2017). Given the cross-sectional nature of our data, we are unable to conclude which mechanism is driving out findings. However, we speculate that a combination of these two mechanisms (and perhaps others) is at play. Future research needs to undertake rigorous longitudinal investigations to develop an understanding of the development of musical preferences and personality throughout the life span and how they impact each other. This is particularly important for understanding childhood and adolescent development when there are increased social pressures and a focus on identity formation (Bonneville-Roussy et al., 2013).

Our study had several limitations. First, the sample was exclusive to U.S. users of Spotify. Therefore, we do not know the extent to which our findings generalize across geographic regions in other Westernized cultures. Furthermore, while recent empirical evidence suggests that music is universal in form and function (Mehr et al., 2019), since our results are based on streaming services that require users to have internet-enabled devices, we are unable to generalize our findings to non-Westernized cultures. Testing how the links between musical preferences and habitual listening behaviors manifest across cultures (in Westernized and non-Westernized around the world) is a ripe area for future research. Second, our understanding of human personality was reliant on a self-report assessment of the Big Five model of personality. Future research should extend our work by investigating other constructs including psychological values (Schwartz, 1992), cognitive profiles (Greenberg et al., 2015), and narrative identities (McAdams, 2008). Third, while the present study takes into account the streaming behavior and self-reported personality of 5,808 listeners, the size of Spotify's audience allows for a much larger study. A larger user sample would allow for more flexibility on model choice (e.g., deep learning, models trained on demographic segments), ultimately yielding more accurate predictions. There is also the potential to assimilate other inputs (e.g., longer listening windows, additional derived metrics) into the models, further improving predictive ability. Fourth, we relied on self-report and behavioral data from Spotify and therefore were unable to draw insights about the role of human biology on musical preferences and habits. Given the vast volume of research on the cognitive neuroscience of music and the emerging literature (Peretz & Zatorre, 2012) on the social neuroscience of music (e.g., the role of oxytocin) (Keeler et al., 2015), future research could begin to link streaming behavior with brain scanning, genetic, and physiological data.

However, such future research and applications must be conducted within the strict boundaries of ethical data usage, collection, and storage policies. A user's digital history is extraordinarily personal and sensitive and should be treated with proper consideration of the conceivable misuses and unintended externalities. We affirm that our methodology, surveying, and data governance were all done under such a framework, including an IRB at Spotify that scrutinized our design and methods before granting formal ethical approval. We disavow any future research and applications which violate ethical standards of data usage and are not transparent about privacy to its users.

In conclusion, we used data from a naturalistic environment to show how personality is intertwined with music listening behavior. The observations were drawn from a substantial data set that included 5,808 users, 211 mood, genre, and behavioral variables, 17.6 million streams, and over 662,000 hr of listening collected over a span of 3 months. The observed links between personality and listening behavior are robust and ecologically valid. The present work is a model for how psychological methods can be fused with cutting-edge technology and big data for scientific inquiry. Finally, the results show that personality does in indeed play an important role in musical preferences and warrants continued rigorous investigation.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Spotify.

ORCID iD

Ian Anderson https://orcid.org/0000-0002-8308-497X

Supplemental Material

The supplemental material is available in the online version of the article.

Notes

1. "Neuroticism" is traditionally used, but we chose to use the term "Emotional Stability" as it more acutely defines the trait.
2. The elastic net procedure is to first conduct ridge regression with a fixed regularization parameter then do lasso shrinking of those coefficients. The double shrinkage is corrected by an intermediate rescaling.
3. Random forest techniques create an ensemble of decision trees on random subspaces of the predictor space then averaging their predictions (Breiman, 2001; Ho, 1998). Although individual trees overfit, in aggregate variance is reduced. We explored other regression methods—stepwise regression, support vector machines (Cortes & Vapnik, 1995), and so on—but none showed significantly better performance, so we restricted the final results to elastic net and random forest.

References

Abel, F., Gao, Q., Houben, G. J., & Tao, K. (2011). Analyzing user modeling on Twitter for personalized news recommendations. *Lecture Notes in Computer Science* (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 1–12. https://doi.org/10.1007/978-3-642-22362-4_1

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), 703–717. https://doi.org/10.1037/a0033770

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

Brown, R. A. (2012). Music preferences and personality among Japanese university students. *International Journal of Psychology*, *47*(4), 259–268. https://doi.org/10.1080/00207594.2011.631544

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

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

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*, 273–297. https://doi.org/10.1007/BF00994018

Covington, P., Adams, J., & Sargin, E. (2016). Deep neural networks for YouTube recommendations. *Proceedings of the 10th ACM Conference on Recommender Systems*. https://doi.org/10.1145/2959100.2959190

Das, A. S., Datar, M., Garg, A., & Rajaram, S. (2007). Google news personalization. *Proceedings of the 16th International Conference on World Wide Web*. https://doi.org/10.1145/1242572.1242610

De Montjoye, Y. A., Quoidbach, J., Robic, F., & Pentland, A. (2013). Predicting personality using novel mobile phone-based metrics.

Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). https://doi.org/10.1007/978-3-642-37210-0_6

Dunn, P. G., Ruyter, B. D., & Bouwhuis, D. G. (2011). Toward a better understanding of the relation between music preference, listening behavior, and personality. *Psychology of Music*, *40*(4), 411–428. https://doi.org/10.1177/0305735610388897

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*, 98–102. https://doi.org/10.1016/j.jrp.2018.06.004

Fricke, K. R., Greenberg, D. M., Rentfrow, P. J., & Herzberg, P. Y. (2019). Assessing music preference from actual listening behavior: Data from over 1 million users and over 200,000 songs. *Psychology of Music*. https://doi.org/10.1177/0305735619868280

Fu, W. (1998). Penalized regression: The Bridge versus the Lasso. *Journal of Computational and Graphical Statistics*, *7*, 397–416.

Golbeck, J., Robles, C., Edmondson, M., & Turner, K. (2011). Predicting personality from Twitter. *Proceedings of 2011 IEEE International Conference on Privacy, Security, Risk and Trust and IEEE International Conference on Social Computing, PASSAT/socialcom 2011*. https://doi.org/10.1109/PASSAT/SocialCom.2011.33

Greenberg, D. M., Baron-Cohen, S., Stillwell, D. J., Kosinski, M., & Rentfrow, P. J. (2015). Musical preferences are linked to cognitive styles. *PLoS ONE*, *10*(7). https://doi.org/10.1371/journal.pone.013115

Greenberg, D. M., Kosinski, M., Stillwell, D. J., Monteiro, B. L., Levitin, D. J., & Rentfrow, P. J. (2016). The song is you. *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

Greenwood, S., Perrin, A., & Duggan, M. (2016). *Social media update 2016*. Pew Research Center.

Habibi, A., Damasio, A., Ilari, B., Veiga, R., Joshi, A. A., Leahy, R. M., Haldar, J. P., Varadarajan, D., Bhushan, C., & Damasio, H. (2017). Childhood music training induces change in micro and macroscopic brain structure: Results from a longitudinal study. *Cerebral Cortex*, *28*(12), 4336–4347. https://doi.org/10.1093/cercor/bhx286

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining*, *inference, and prediction* (2nd ed., pp. 241–248). Springer Publishing Company. https://doi.org/10.1007/978-0-387-84858-7

Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(8): 832–844. https://doi.org/10.1109/34.709601

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*. https://doi.org/10.1080/00401706.1970.10488634

Horling, B., & Kulick, M. (2009). Personalized Search for everyone. In *Google Blog*.

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* (pp. 102–138). Guilford Press.

Johnston, M. (2018). How Spotify discovers the genres of tomorrow. Spotify for Artists Blog. https://artists.spotify.com/blog/how-spotify-discovers-the-genres-of-tomorrow

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*(5), 770–814. https://doi.org/10.1037/0033-2909.129.5.770

Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences*, *31*(5), 559–575 https://doi.org/10.1017/s0140525x08005293

Keeler, J. R., Roth, E. A., Neuser, B. L., Spitsbergen, J. M., Waters, D. J. M., & Vianney, J. M. (2015). The neurochemistry and social flow of singing: Bonding and oxytocin. *Frontiers in Human Neuroscience*, *9*, 5180. https://doi.org/10.3389/fnhum.2015.00518

Khan, I. A., Brinkman, W., Fine, N., & Hierons, R. M. (2008). Measuring personality from keyboard and mouse use. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/1473018.1473066

Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, *42*(8), 30–37. https://doi.org/10.1109/MC.2009.263

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, *110*(15), 5802–5805. https://doi.org/10.1073/pnas.1218772110

Langmeyer, A., Guglhör-Rudan, A., & Tarnai, C. (2012). What do music preferences reveal about personality? *Journal of Individual Differences*, *33*(2), 119–130. doi:10.1027/1614-0001/a000082

Levitin, D. J., & Grafton, S. T. (2016). Measuring the representational space of music with fMRI: A case study with Sting. *Neurocase*, *22*(6), 548–557. https://doi.org/10.1080/13554794.2016.1216572

Limb, C. J., & Braun, A. R. (2008). Neural substrates of spontaneous musical performance: An fMRI study of jazz improvisation. *PLoS ONE*, *3*(2), e1679. https://doi.org/10.1371/journal.pone.0001679

Mairesse, F., Walker, M. A., Mehl, M. R., & Moore, R. K. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research*, *30*, 457–500. https://doi.org/10.1613/jair.2349

Marcus, B., Machilek, F., & Schütz, A. (2006). Personality in cyberspace: Personal web sites as media for personality expressions and impressions. *Journal of Personality and Social Psychology*, *90*(6), 1014–1031. https://doi.org/10.1037/0022-3514.90.6.1014

McAdams, D. P. (2008). Personal narratives and the life story. In O. John, R. W. Robins, & L. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed.). Guilford Press. https://doi.org/10.1136/ewjm.173.1.32

McPherson, M. J., Barrett, F. S., Lopez-Gonzalez, M., Jiradejvong, P., & Limb, C. J. (2016). Emotional intent modulates the neural substrates of creativity: An fMRI study of emotionally targeted improvisation in jazz musicians. *Scientific Reports*, *6*(1). https://doi.org/10.1038/srep18460

Mehr, S. A., Singh, M., Knox, D., Ketter, D. M., Pickens-Jones, D., Atwood, S., Lucas, C., Jacoby, N., Egner, A. A., Hopkins, E. J., Howard, R. M., Hartshorne, J. K., Jennings, M. V., Simson, J., Bainbridge, C. M., Pinker, S., O'Donnell, T. J., Krasnow, M. M., & Glowacki, L. (2019). Universality and diversity in human song. *Science*, *366*(6468), eaax0868. https://doi.org/10.1126/science.aax0868

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*. https://doi.org/10.1177/0956797618761659

North, A. C. (2010). Individual differences in musical taste. *American Journal of Psychology*. https://doi.org/10.5406/amerjpsyc.123.2.0199

North, A. C., Tarrant, M., & Hargreaves, D. J. (2004). The effects of music on helping behavior: A field study. *Environment and Behavior*. https://doi.org/10.1177/0013916503256263

Peretz, I., & Zatorre, R. J. (2012). *The cognitive neuroscience of music*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198525202.001.0001

Quercia, D., Kosinski, M., Stillwell, D., & Crowcroft, J. (2011). Our Twitter profiles, our selves: Predicting personality with Twitter. In *Proceedings of 2011 IEEE International Conference on Privacy, Security, Risk and Trust and IEEE International Conference on Social Computing, PASSAT/SocialCom 2011*. https://doi.org/10.1109/PASSAT/SocialCom.2011.26

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

Rickard, N. S. (2004). Intense emotional responses to music: A test of the physiological arousal hypothesis. *Psychology of Music*, *32*(4), 371–388. https://doi.org/10.1177/0305735604046096

Salimpoor, V. N., Benovoy, M., Larcher, K., Dagher, A., & Zatorre, R. J. (2011). Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nature Neuroscience*, *14*(2), 257–262. https://doi.org/10.1038/nn.2726

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

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

Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In M. P. Zanna (Ed.), *Advances in experimental social psychology (Vol. 25*, pp. 1–65). Academic Press. https://doi.org/10.1016/S0065-2601(08)60281-6

Skowron, M., Ferwerda, B., Tkalčič, M., & Schedl, M. (2016). Fusing social media cues : Personality prediction from Twitter and Instagram. *WWW’16 Companion*. https://doi.org/10.1145/2872518.2889368

Sloboda, J. A., Oneill, 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/10.1177/102986490100500102

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), 330–348. https://doi.org/10.1037/a0021717

Summers, C. (2016). Feel like listening? Computing musical mood at Gracenote. *Gracenote Tech Blog*. https://www.gracenote.com/computing-musical-mood-at-gracenote/

Swann, W. B., Jr., Rentfrow, P. J., & Guinn, J. S. (2002). Self-verification: The search for coherence. In M. Leary & J. Tagney (Eds.) *Handbook of self and identity* (pp. 367–383). 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

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, *112*(4), 1036–1040. https://doi.org/10.1073/pnas.1418680112

Zentner, M., Grandjean, D., & Scherer, K. R. (2008). Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion*, *8*(4), 494–521. https://doi.org/10.1037/1528-3542.8.4.494

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

Author Biographies

Ian Anderson is a machine learning engineer at Spotify, New York, USA. His primary research interests are machine learning, recommender systems, and their impact on society. He holds a Ph.D. in physics from Johns Hopkins University.

Santiago Gil is an independent scientist and entrepreneur.

Clay Gibson is a software engineer at Spotify, New York, USA. His main research interests are related to machine learning and recommender systems. He holds a B.S. in Psychology from Yale University, USA.

Scott Wolf is a Quantitative User Researcher at Google in New York City.

Will Shapiro is the co-founder and CEO of Topos and Adjunct Assistant Professor at The Cooper Union, where he teaches courses on data science, machine learning and cities.

Oguz Semerci is a software engineer at Spotify, New York, USA. His main research interests are related to machine learning and recommender systems. He holds a Ph.D. in Electrical Engineering from Tufts University, USA.

David M. Greenberg, PhD is a Zuckerman Postdoctoral Scholar at Bar-Ilan University in Israel and a visiting researcher at Cambridge University in England. His research investigates the impact of music on individuals and society using big data, psychology, and social neuroscience.

Handling Editor: Gregory Webster