Musical Preferences Predict Personality: Evidence from Active Listening and Facebook Likes

Journal:	*Psychological Science*
Manuscript ID	PSCI-16-1292.R2
Manuscript Type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Nave, Gideon; Caltech, Computation and Neural Systems Minxha, Juri; California Institute of Technology, Computation and Neural Systems Greenberg, David; University of Cambridge, Psychology Kosinski, Michal; Stanford University, Graduate School of Business Stillwell, David; University of Cambridge, Psychometrics Centre Rentfrow, Peter; University of Cambridge, Social & Developmental Psychology; Jason Rentfrow,
Keywords:	Music, Personality, Online behavior, Prediction, Machine learning

Musical Preferences Predict Personality: Evidence from Active Listening and Facebook Likes

Gideon Nave, Department of Marketing, The Wharton School of the University of Pennsylvania
Juri Minxha, Computation & Neural Systems, Caltech
David M Greenberg, Department of Psychology, University of Cambridge
Michal Kosinski, Graduate School of Business, Stanford University
David Stillwell, Judge Business School, University of Cambridge
Jason Rentfrow,* Department of Psychology, University of Cambridge

* Corresponding author, pjr39@cam.ac.uk

Abstract

Research over the past decade has shown that various personality traits are communicated through musical preferences. One limitation of that research is external validity, as most studies have assessed individual differences in musical preferences using self-reports of music-genre preferences. Are personality traits communicated through behavioral manifestations of musical preferences? We address this question in two large-scale online studies with demographically diverse populations. Study 1 (N=22,252) shows that reactions to unfamiliar musical excerpts predicted individual differences in personality - most notably openness and extraversion - above and beyond demographic characteristics. Moreover, these personality traits were differentially associated with particular music-preference dimensions. The results from Study 2 (N=21,929) replicated and extended these findings by showing that an active measure of naturally-occurring behavior, Facebook Likes for musical artists, also predicted individual differences in personality. In general, our findings establish the robustness and external validity of the links between musical preferences and personality.

Introduction

With the proliferation of Internet-based services for sharing and streaming music on demand, personalized music is becoming a more central and prominent fixture in many people's lives. This increase coincides with a growing interest in understanding the psychological basis of musical preferences. Over the past decade, several studies have investigated individual differences in musical preferences with the aim of identifying its structure and psychological correlates. In general, these investigations offer promising evidence that musical preferences can be reduced to and conceptualized by a few broad dimensions, and that various aspects of musical preferences are associated with individual differences in a range of psychological variables.

Informed by theory and research on person-environment interactions, a number of studies have examined associations between musical preferences and personality (e.g., Miranda and Claes 2008; Greenberg et al. 2015, 2016; Langmeyer, Guglhör-Rudan, and Tarnai 2012; Schäfer and Mehlhorn 2017; Rentfrow and Gosling 2003). The motivation for these studies has been to develop and test the hypothesis that individuals are drawn to musical styles that satisfy and reinforce their psychological needs. The results suggest, for example, that people who have a need for creative and intellectual stimulation prefer unconventional and complex musical styles, and that people who are sociable and enthusiastic prefer musical styles that are energetic and lively (Rentfrow and Gosling 2003).

Although the results from studies on the links between musical preferences and personality generally converge, past research suffers important limitations. One limitation concerns the way in which musical preferences are measured. The most common method for assessing musical preferences relies on self-reported preferences for musical genres (e.g., classical, rock, rap, etc.), treated as proxies for musical preferences. This is problematic for three reasons. First, there is no consensus about which genres to measure, with studies employing from a few to over 30 genres and subgenres (e.g., George et al. 2007; Yeoh and North 2010). Second, participants may differ in their definitions and interpretations of what type of music different genres represent, which in turn might add undesirable noise to the measurement of their musical tastes. Third, it is not clear to what extent findings from survey studies represent the actual preferences and behaviors of the participants in the real world.

Another limitation of past studies is their reliance on samples of college students (e.g., Brown 2012; Langmeyer, Guglhör-Rudan, and Tarnai 2012; George et al. 2007; Palmer and Griscom 2013; Vuoskoski and Eerola 2011; Rentfrow and Gosling 2003). As music is particularly important to young people (Bonneville-Roussy et al. 2013) and their peer-group relations (Delsing, et al., 2008), college students may report stronger preferences for musical genres that are popular among their peers, due to social desirability.

To overcome these limitations, we conducted two studies investigating whether the links between musical preferences and personality generalize across different assessment methods and across age-diversified samples. Our primary objective was to determine if individual differences in the Big Five personality domains can be predicted from musical preferences. In Study 1 we used a machine-learning "predictive" approach (Yarkoni and Westfall 2017) to examine whether participants' preference ratings following active listening to novel musical stimuli can be used for out-of-sample predictions of their personalities. Study 2 replicates and extends Study 1 using an ecologically valid behavioral measure of musical preferences: Facebook Likes of musical artists.

Study 1: Preferences for novel music following active listening predict personality

Methods

Participants. We used data from a sample of 22,252 *MyPersonality* users from 153 different countries.[1] Majority of the participants self-reported gender (N = 20,770; 62% female); about half reported their age (N = 10,414, median = 22, interquartile range= 7), 45% of which reported being over 22 years of age, the typical age of a college graduate in the US. Among 17,988 users who reported their geographical location, 25% (N = 4,517) lived in countries other than the US, UK, or Canada. All respondents received feedback about their musical preferences (according to

[1] The myPersonality Facebook app ran from 2007 to 2012. It presented the opportunity for Facebook users to take real scientific research questionnaires and get feedback on their results. Overall, more than 6 million users took at least one questionnaire. The raw data used in the current study is available for researchers on the project's website, http://mypersonality.org.

the MUSIC model, further details below) and their Big Five personality following the questionnaire. The study's sample included all MyPersonality users who (1) completed a Big Five personality questionnaire, and (2) completed at least one music-preference survey (further details below).

Personality. Respondents' personality profiles were estimated using the International Personality Item Pool (IPIP) questionnaire measuring Five Factor Model (FFM) of personality (20 to 100 item long; Goldberg et al. 2006/2)**.**

Musical preferences. Preferences for Western music can be reduced to a few dimensions (Colley, 2008; George et al., 2007; Rentfrow, Goldberg, & Levitin, 2011; Rentfrow, Goldberg, Stillwell, Kosinski, Gosling, & Levitin, 2012; Rentfrow & Gosling, 2003; Rentfrow & McDonald, 2009; Schäfer & Sedlmeier, 2009). Analyses of the psychological, social, and auditory characteristics of the dimensions suggests they can be defined as: Mellow, Unpretentious, Sophisticated, Intense, and Contemporary (MUSIC). The Mellow dimension represents music that is romantic, relaxing, and slow, and comprises soft rock, R&B, and adult contemporary musical pieces. The Unpretentious dimension represents music that is uncomplicated, relaxing, and acoustic, and comprises country, folk, and singer/songwriter pieces. The Sophisticated dimension represents music that is inspiring, complex, and dynamic, and comprises classical, operatic, world, and jazz pieces. The Intense dimension represents music that is distorted, loud, and aggressive, and comprises classic rock, alternative rock, punk, and heavy metal pieces. The Contemporary dimension represents music that is percussive, electric, and not sad, and comprises rap, electronic dance music, Latin, and Euro pop pieces. Recent work indicated that the MUSIC model accounts for 55% to 59% of the variance in preferences for Western music (Bonneville-Roussy et al. 2013; Rentfrow et al. 2012; Rentfrow et al., 2011).

We estimated musical preferences using surveys designed according to the five-factor MUSIC model (Rentfrow et al., 2011; Rentfrow et al. 2012).[2] Each survey comprised 25 different 15-second musical excerpts, with five excerpts representing each factor. Overall, there were 6 different musical surveys (Rentfrow et al., 2011; Rentfrow et al. 2012). Two surveys (Mix_A, $N = 17{,}904$; Mix_B, $N = 10{,}840$), consisted of excerpts from a multitude of genres and subgenres, the copyrights of which were purchased from Getty Images; thus, it was unlikely that participants had previous exposure to them. Four other surveys included commercially released music by known artists, of which two surveys consisted of only rock excerpts (Rock_A, $N = 2{,}758$; Rock_B, $N = 1{,}748$), and two surveys included only jazz excerpts (Jazz_A, $N = 1{,}590$; Jazz_B, $N = 8{,}887$). All of the excerpts were used as stimuli that represent the 5-factor MUSIC model in previous work.[3] Each participant was assigned to one of three conditions [Mix, Jazz, Rock] and took its corresponding survey "A". Then, participants were given the opportunity to take the second survey ("B"), always in the same condition as survey A. Surveys with missing responses were excluded from further analysis.

---

[2] The excerpts are available for download on the project's page in the open science framework (OSF): https://osf.io/nfqb9/?view_only=dff0271c8e0049bc88738e5a7b51ec2f

[3] The mixed genre excerpts were used in (Rentfrow, Goldberg, and Levitin 2011; Rentfrow et al. 2012; Greenberg et al. 2015, 2016). A sub-sample (about 5%) of the "mix" respondents in the current study were also included in (Rentfrow et al. 2012). The rock and jazz excerpts were used in (Rentfrow et al. 2012; Greenberg et al. 2015) and (Greenberg et al. 2016).

Prediction algorithm. For each of the Big Five personality traits, we conducted out-of-sample predictions based on 1) preference ratings for the 25 musical excerpts; 2) survey responses + gender and age; 3) gender and age alone. Predictions were carried out using the following nested cross-validation procedure[4]:

1. We randomly split the entire data set into ten groups of participants.[5]
2. For each of the 10-holdout groups, we trained a linear model to predict each of the Big Five personality traits by fitting a linear regression with a Least Absolute Shrinkage and Selection Operator (LASSO) penalty to the remaining 90% of the data (Tibshirani 1996; Camerer, Nave, and Smith 2017). The tuning parameter, λ was optimized via 10-fold cross validation (Stone 1974), performed within each training set.[6]
3. Using that trained model, we conducted out-of-sample predictions for the remaining 10% of the data (i.e., the holdout group). We estimated the predictive accuracy by calculating the Pearson's correlation between the actual and predicted personality-trait scores.[7]

Results

Preferences for novel music predict personality traits

Here, we report personality predictions based on the responses to the survey with the largest number of responses, Mix_A (*N = 17,904*), and discuss further replications in the section that follows. The results are summarized in Fig. 1A and Table 1. For all the personality traits, we found reliable correlations between the music-based personality predictions and the actual traits (all p-values <0.001). The highest correlation was observed for openness ($r(17{,}904)=0.25$, 95% confidence interval (CI)=[0.23 0.26]), followed by extraversion ($r(17{,}902)=0.16$, 95% CI=[0.14 0.17]), agreeableness ($r(17{,}903)=0.15$, 95% CI=[0.14 0.17]), neuroticism ($r(17{,}905)=0.12$, 95% CI=[0.10 0.13]), and conscientiousness ($r(17{,}174)=0.11$, 95% CI=[0.10 0.13]). The music-based predictors of openness, extraversion and agreeableness were significantly better than a baseline model that predicted personality solely using gender and age (we rejected the null hypothesis of equality in out-of-sample predictive accuracies at the $p<0.01$ level, Steiger's z-test (Steiger 1980)).[8] Adding musical preferences to the baseline model (gender and age) significantly increased the predictive accuracy for all of the Big Five traits (all p's<0.012, Steiger's z-test). To put these results in perspective, knowledge of one's musical preferences reveals nearly as much about their personality trait of openness as their behavior at work reveals to a work colleague; for the remaining traits, predictive accuracy ranged between 41% (conscientiousness) and 66% (neuroticism) of a colleague's accuracy (Youyou, Kosinski, and Stillwell 2015).

---

[4] Analyses scripts and pre-processed data are available on the project's OSF page: https://osf.io/nfqb9/?view_only=dff0271c8e0049bc88738e5a7b51ec2f

[5] The random partition of the data into ten hold-out groups was conducted independently for each of the models (i.e., there was a different partition for every combination of personality trait and dependent variables).

[6] Optimization was performed using the 'lassoglm' function implemented in 'Matlab', under its default setting. Thus, we first estimate λ_MAX - the largest value of the penalty parameter λ that gives a non-null model, and perform optimization by exploring a geometric sequence of 100 values between 0.0001λ_MAX and λ_MAX. The values of the tuning parameter λ for the main models (averaged across the 10 folds estimating personalities from the survey responses) are available on Table S3. As the LASSO procedure is sensitive to the scale of the inputs, all independent variables were standardized (z-scored) prior to model training.

[7] We estimated the correlations and confidence intervals on all of the data collapsed.

[8] The results hold when limiting the training and testing samples to participants who self-reported age and gender, ruling out the possibility that sample size differences underlie the improved predictive accuracy of the music-preferences based models.

These results indicate that preferences for short musical excerpts contain predictive information about personality traits. However, they do not allow us to tease apart whether this information arises from our participants' unique musical tastes (represented by the liking of individual excerpts) or from their tendencies to like music in general. To further investigate this issue, we constructed, for each of the Big Five traits, an additional "general baseline model", that included (i) participants' general evaluative tendencies (i.e., mean preference rating from all the musical pieces), (ii) a general music liking factor, calculated by fitting a bi-factor model (Reise, Moore, and Haviland 2010) to the liking ratings[9], and (iii) gender and age.

Contrasting the predictive accuracies of the model that includes responses to individual survey items, gender and age (Table 1, row 2) with the general baseline model (Table 1, row 4), allows us to disentangle the predictive accuracies arising from specific versus general musical preferences. We find that the additional predictive accuracies obtained by including the individual survey responses (above the general baseline model) was highest for openness ($\Delta r = 0.09$, 55% increase) and extraversion ($\Delta r = 0.08$, 79%). However, they were less pronounced for the three other traits ($\Delta r = 0.03$, 17% for neuroticism, $\Delta r = 0.02$, 15% for agreeableness, and only $\Delta r = 0.01$, 7% for conscientiousness).[10] Thus, both specific and general musical preferences underlie the capacity to predict personality from musical preferences, where the former play a substantial role for the cases of openness and extraversion, and the latter underlie the capacity to predict the other three traits.

Finally, we explored the generalizability of our findings to two sub-populations that are typically under-represented in laboratory studies conducted in college students. First, we found that all of the results hold when limiting the estimates of predictive accuracy to participants who self-reported residing outside the US, UK, or Canada (N=1,596, see Table S1): adding musical preferences significantly increased the predictive accuracy of the baseline model that included only age and gender in this sub-population (for neuroticism $p=0.039$, for all other traits p's < 0.01, Steiger's z-test). The results hold when restricting the analyses to participants that self-reported being over 30 years of age (N=1,528, see Table S1): adding the musical preferences survey increased the predictive accuracy of the baseline demographic model for all traits (openness, extraversion and agreeableness: p's < 0.01; for conscientiousness $p=0.047$, for neuroticism $p=0.060$, Steiger's z-test).

[9] The general music liking factor was estimated by fitting a bifactor model to Mix_A survey responses. The model included a general factor and five additional orthogonal dimensions (representing the MUSIC factors), and was fitted using the 'psych' library of the statistical software 'R'. The general factor significantly loaded on 18 of the 25 response items, but did not substantially load on the "intense" excerpts (see Table 3). The empirical question of whether a valid general music liking factor exists is beyond the scope of this paper.

[10] The predictive accuracy of the general baseline model was significantly smaller that the model that also included the ratings of individual items, for four of the traits (all p's<0.028, Steiger's z-test). For conscientiousness, predictive accuracy was greater in the model that included the ratings of individual items, though the difference in predictive accuracy did not reach statistical significance (p=0.29).

Table 1: Predictive accuracy of musical preference based personality predictors (out of sample), for all Big Five traits, test Mix_A (*N=17,904, N=8,100 with age and gender information*).

Mix A	Openness					Conscientiousness					Extraversion					Agreeablness					Neuroticism				
	N	R			p	N	R	95% CI		p	N	R	95% CI		p	N	R			p	N	R	95% CI		p
MUSIC	17,904	.25	.23	.26	<.001	17,904	.11	.10	.13	<.001	17,902	.16	.14	.17	<.001	17,903	.15	.14	.17	<.001	17,174	.12	.10	.13	<.001
Music+Gender+Age	8,100	.25	.23	.27	<.001	8,100	.16	.14	.18	<.001	8,099	.17	.15	.19	<.001	8,100	.18	.16	.20	<.001	7,933	.19	.16	.21	<.001
Gender+Age	8,100	.03	.01	.05	.013	8,100	.13	.11	.15	<.001	8,099	-.01	-.03	.01	.432	8,100	.06	.04	.08	<.001	7,933	.15	.13	.17	<.001
M+g+Gender+Age	8,100	.16	.14	.18	<.001	8,100	.15	.13	.17	<.001	8,099	.10	.07	.12	<.001	8,100	.15	.13	.18	<.001	7,933	.16	.14	.18	<.001
MUSIC (Gender+Age sample)	8,100	.24	.22	.26	<.001	8,100	.10	.08	.12	<.001	8,099	.17	.15	.20	<.001	8,100	.17	.15	.19	<.001	7,933	.12	.10	.15	<.001

Figure 1A: Correlations between music preference-based big five personality predictors (out of sample) and actual personalities, test A (*N=17,904*). Error bars denote 95% CIs.

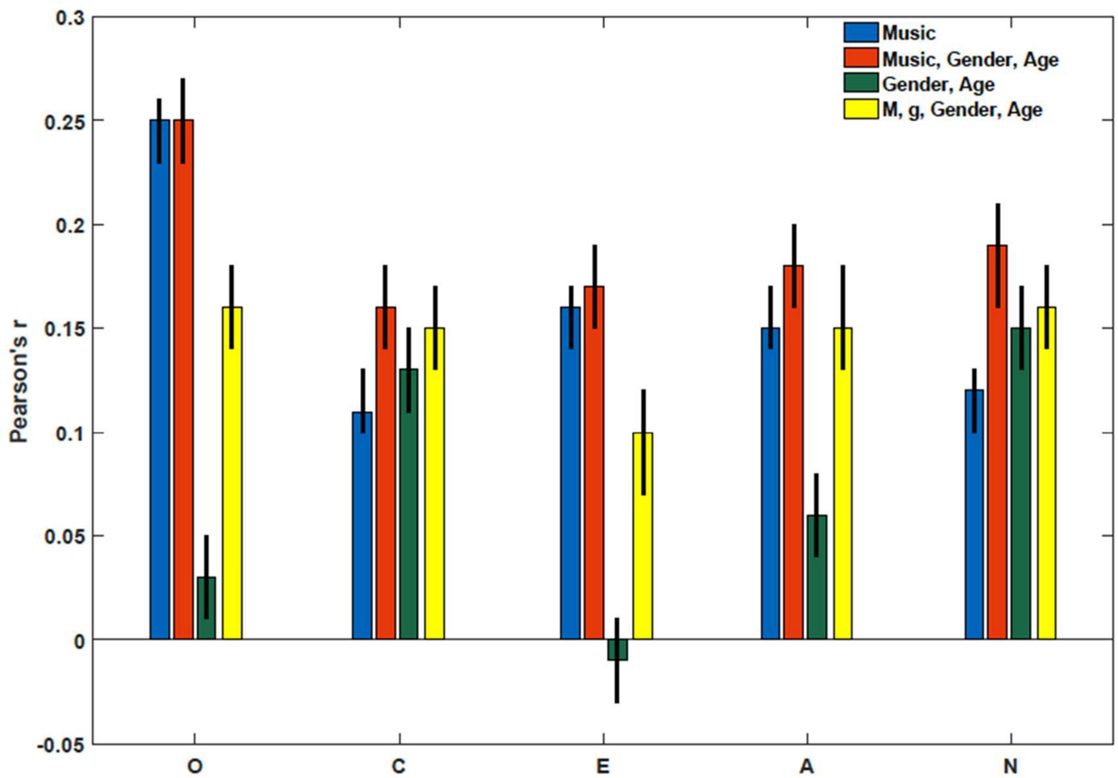
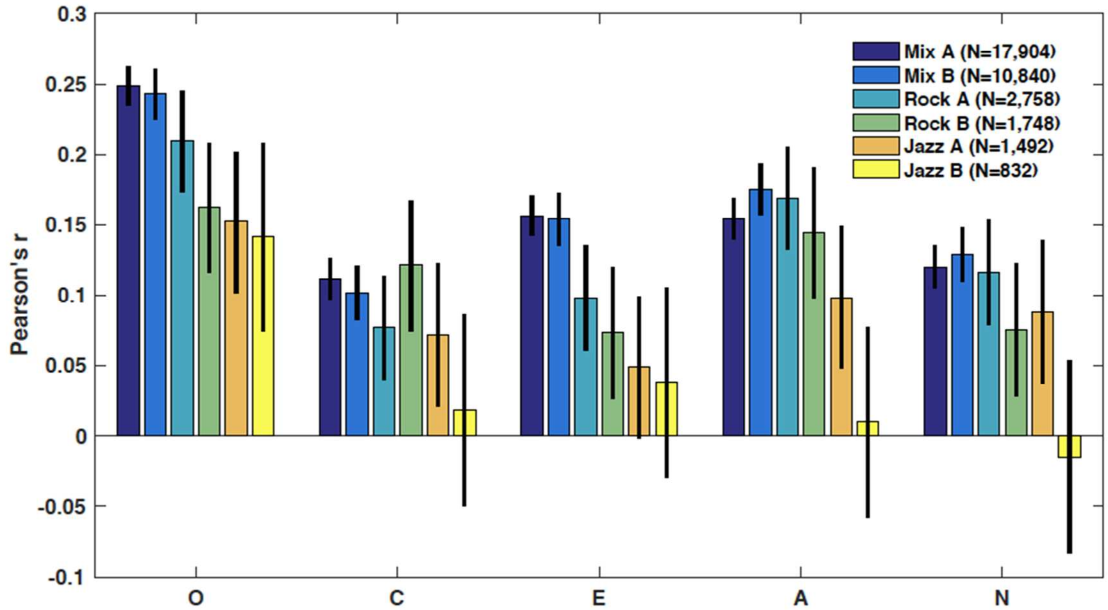


Figure 1B: Correlations between music-based big five personality predictors (out of sample) and actual personalities across tests and genres. Error bars denote 95% CIs.



Replication across tests and genres

To evaluate the robustness of the predictive accuracy results, we carried out the same analyses again for the other five musical preferences surveys. It is important to bear in mind that the sample sizes for these surveys were significantly smaller (between 5% and 45% of Mix_A's sample size), and therefore (1) predictive accuracy was expected to decline, as the models' parameter estimates were less stable, and (2) the capacity to detect effects decreased due to reduced statistical power, especially for the traits for which the associations between preferences and personality were expected to be smaller (conscientiousness and neuroticism).

The results are summarized in Fig. 1B and SOM Table S2. The most similar replication used survey Mix_B, that was taken by a subpopulation (about 45%) of Mix_A respondents, and like Mix_A, consisted of excerpts from multiple genres. The predictive accuracies of the models trained using Mix_B were significantly greater than zero (all p's < 0.001), and their point estimates were greater than the lower bounds of the 95% CIs of the predictive accuracies obtained from Mix_A responses, for all of the Big Five traits. Furthermore, adding Mix_B survey responses to the baseline demographic model (constructed from age and gender) significantly improved the predictive accuracies for openness, extraversion and agreeableness (p's < 0.001, Steiger's z-test), providing a successful replication of survey Mix_A for these traits. For neuroticism the improvement was marginally significant ($p=0.11$), and for conscientiousness we did not detect a reliable improvement ($p=0.47$).

Next, we repeated the analyses for the rock and jazz surveys. These surveys were designed to capture the dimensions of the MUSIC model, while containing excerpts from exclusively one genre. For the two rock surveys (Rock_A, N = 2,758; Rock_B, N = 1,748) the predictive accuracies of all ten models (5 personality traits x 2 surveys) were reliably greater than zero (all p's<0.01), and adding the responses for these musical surveys to the baseline model (gender and age) increased the predictive accuracy of the models for all traits except neuroticism (openness, extraversion and agreeableness: p's<0.01; conscientiousness p<0.10).

The models using responses to Jazz_A (N=1,590) had statistically significant predictive accuracies (p's<0.01) for all traits except extraversion. Adding the responses for these musical surveys to the baseline model (gender and age) increased the predictive accuracy of all traits, though the improvement was not statistically significant, perhaps due to the small sample. For Jazz_B (the smallest survey, N=887) we detected a reliable predictive accuracy only when predicting openness (p<0.001), and marginally significant (p<0.10) predictive accuracies for agreeableness and neuroticism, likely because the small sample (about 20 times smaller than Mix_A) might have been insufficient for model training.

Robustness of the 5-factor MUSIC model in a large diverse sample

Apart from examining the capacity to predict personality from liking of music, our data provide a unique opportunity to estimate the robustness of the 5-factor MUSIC model (Rentfrow et al.,2011; Rentfrow et al. 2012), and the capacity of our musical surveys to capture it. To do this, we subjected the survey responses of the participants who answered both surveys A and B of the mix genre excerpts (i.e., Mix_A and Mix_B, *N* = 10,840) to Principal Component Analysis (PCA).[11] Investigating the projections of the different musical excerpts onto each of the principal components, revealed that each group of excerpts, that was a-priori selected to represent a MUSIC dimension, mapped into a unique principle component, for which the average projection was an order of magnitude greater than the projection onto the four other components (Table 2, see supplementary Fig. S1, A-E for projections of individual excerpt). Further, the first five principal components explained 59% of the variance in the data (Fig. S1 F). Similar results were obtained for the responses to the Jazz and Rock surveys, and are published elsewhere (Rentfrow et al. 2012).

As the musical surveys used by the current investigation were specifically designed to capture the MUSIC model, examining these results in isolation would not allow concluding that *all types of Western music* are captured by the five-factor framework. However, it is important to keep in mind that the MUSIC model was originally constructed based on exploratory research that used a wide variety of musical pieces that are different from the ones used in the present research (Rentfrow, Goldberg, and Levitin 2011). The current results corroborate that the MUSIC model is a robust framework for organizing individual differences in preferences for music, and demonstrate its generalizability to a large and diverse population.

[11] A similar analysis using the same participants who took the jazz and rock excerpts, and a small sub-sample (about 5%) of the mixed survey respondents was previously published in (Rentfrow et al. 2012).

Table 2: Average loadings of the excerpts' liking ratings on the first five principal components of the data. Each row represents the 10 excerpts from surveys Mix_A and Mix_B that a-priori represented each of the five MUSIC dimensions.

A-priori MUSIC dimension	F1	F2	F3	F4	F5
Mellow	0.024	0.005	0.049	0.020	**-0.241**
Unpretentious	0.024	0.012	**0.272**	-0.016	-0.017
Sophisticated	**0.284**	-0.004	-0.003	0.001	-0.008
Intense	-0.003	**0.308**	-0.002	0.001	-0.003
Contemporary	0.014	0.007	-0.004	**0.288**	-0.025

Table 3: Average loadings of the excerpts' liking ratings on the general factor and the first five principal components of the data, extracted using a bi-factor model. Each row represents the 5 excerpts from surveys Mix_A that a-priori represented each of the five MUSIC dimensions.

A-priori MUSIC factor	General	F1	F2	F3	F4	F5
Mellow	**0.625**	-0.029	-0.007	0.061	0.062	**0.312**
Unpretentious	**0.366**	0.082	-0.069	**0.472**	-0.121	-0.146
Sophisticated	**0.463**	-0.013	**0.579**	-0.041	-0.025	-0.073
Intense	0.075	**0.786**	-0.009	0.015	-0.008	-0.005
Contemporary	**0.439**	0.004	-0.004	-0.012	**0.588**	-0.008

Links between the Big Five and MUSIC dimensions

The results indicate that the MUSIC model can be recovered from preference ratings for novel musical stimuli, and that personality traits can be predicted from these ratings. We now turn to investigate whether, and to what extent, systematic associations between the Big Five and preferences for specific MUSIC dimensions exist.

In order to tease apart the different MUSIC components from general liking tendencies, we performed a bi-factor analysis on the individual responses to survey Mix_A. The analysis resulted in five factors that captured the (a-priori defined) MUSIC dimensions, as well as a general liking factor (see Table 3). We then calculated the partial-correlation between the Big Five traits and the projections of participants' preference on (i) the general liking factor, and (ii) the lower dimensions capturing the MUSIC dimensions. These partial correlations controlled for gender and age, and for the low-order MUSIC dimensions they also controlled for the general liking factor.

The results are summarized in Table 4, and show that two personality traits are associated with preferences for specific MUSIC dimensions, above demographics and the general liking tendency. In particular, openness is associated with greater liking of sophisticated music ($r(8{,}097) = 0.16$, 95% CI = [0.14 0.18], $p<0.001$), and less liking of mellow ($r(8{,}097) = -0.12$, 95% CI = [-0.10 -0.14], p<0.001) and contemporary music (r(8,098) = -0.11, 95% CI = [-0.09 -0.13], p<0.001), where extraversion is associated with preference for unpretentious music (r(8,096) = 0.13, 95% CI = [0.11 0.15], p<0.001). Openness and extraversion are also associated with general liking of music (openness: r(8,098) = 0.14, 95% CI = [0.12 0.16] p<0.001; extraversion: r(8,097) = 0.10, 95% CI = [0.08 0.12], p<0.001). For the remaining three traits, none of the specific correlations exceeded r = 0.06, and agreeableness was the only trait associated with general liking of music (r(8,098) = 0.14 95% CI = [0.12 0.16] p<0.001).

We further explored the links between personality and preferences for the individual excerpts representing the MUSIC dimensions in all of the six musical surveys, by estimating the univariate correlations between responses to the different survey questions (i.e., specific excerpts) and personality traits. In Fig. 2, each 6 x 5 framed square represents a different combination of a Big Five trait (row) and a MUSIC factor (column). For example, the top left square represents the correlations between openness and the different excerpts capturing the Mellow dimension. Each row within this square represents one of the six different surveys, and contains the five different excerpts that correspond to the mellow factor in the survey.

Several patterns emerge in the correlation map. Most notably, the correlations are typically small in size (none was greater than r = .21), and are positive for all of the traits except neuroticism. In line with the partial-correlations reported above (for survey Mix_A), openness most strongly correlated with liking the sophisticated excerpts, and extraversion was most strongly correlated with evaluating the unpretentious excerpts more positively.

Table 4: Partial correlations between the Big Five traits and the general music liking factor as well as the lower-order MUSIC dimensions, extracted by performing a bi-factor model on the responses to survey Mix_A. The correlations control for gender and age, and for the lower dimension they also control for the general factor.

	General	M	U	S	I	C
O	**0.14**	**-0.12**	-0.02	**0.16**	0.07	**-0.11**
C	0.06	0.05	0.02	-0.03	-0.02	0.00
E	**0.10**	-0.05	**0.13**	-0.06	0.00	-0.01
A	**0.14**	0.06	0.00	-0.06	0.00	0.02
N	-0.06	0.03	-0.06	0.02	0.04	0.00

Figure 2: correlations between big five personality traits and liking of different musical excerpts.

		M					U					S					I					C				
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
O	Jazz A	-0.01	0.02	0.14	0.12	0.03	0.11	0.09	0.03	0.08	0.10	0.15	0.08	0.11	0.09	0.13	0.10	0.07	0.06	0.10	0.10	0.09	0.10	0.07	0.14	0.10
	Jazz B	0.12	0.03	0.02	0.02	0.10	0.09	0.10	0.08	0.11	0.05	0.08	0.06	0.06	0.10	0.10	0.09	0.09	0.05	0.10	0.10	0.07	0.07	0.04	0.12	0.13
	Rock A	0.03	0.12	0.12	0.05	0.16	0.10	0.09	0.08	0.07	0.04	0.17	0.12	0.07	0.14	0.05	0.06	0.02	0.06	0.04	0.09	0.10	0.02	0.07	0.09	0.05
	Rock B	0.09	0.09	0.00	0.11	0.13	0.08	0.08	0.06	0.04	0.08	0.10	0.03	0.13	0.06	0.03	0.06	0.05	0.03	0.08	0.01	0.08	0.05	-0.02	0.03	0.10
	Mix A	0.07	0.02	0.05	0.00	0.04	0.06	0.08	0.03	0.05	-0.03	0.17	0.16	0.17	0.17	0.17	0.09	0.06	0.07	0.07	0.06	0.02	0.02	-0.04	0.01	0.06
	Mix B	0.15	0.03	0.09	0.07	0.03	0.10	0.05	0.03	0.01	-0.01	0.21	0.17	0.13	0.11	0.16	0.06	0.07	0.02	0.08	0.06	0.09	-0.02	0.00	0.05	0.11
C	Jazz A	0.08	0.05	0.02	0.03	0.05	0.04	0.01	0.00	0.06	0.00	-0.01	-0.02	-0.01	0.02	0.01	-0.01	-0.02	-0.01	-0.01	-0.09	-0.04	-0.04	-0.04	-0.03	-0.02
	Jazz B	0.08	0.07	0.08	0.06	0.12	0.08	0.07	0.06	0.10	0.07	-0.02	-0.02	-0.02	0.00	-0.04	-0.04	-0.02	-0.04	0.01	-0.05	-0.06	-0.01	-0.06	-0.01	-0.02
	Rock A	0.06	-0.01	0.01	0.04	0.00	0.04	-0.02	-0.05	-0.03	0.00	0.01	0.04	-0.04	-0.02	-0.01	-0.03	-0.07	-0.05	-0.07	-0.05	0.03	0.01	-0.01	0.05	0.01
	Rock B	0.01	-0.02	0.07	-0.04	-0.01	0.11	-0.07	0.04	-0.04	-0.03	-0.02	0.07	-0.01	0.00	-0.04	-0.09	0.00	-0.05	0.01	0.00	-0.03	0.00	0.01	0.07	-0.03
	Mix A	0.06	0.06	0.05	0.06	0.04	0.03	-0.03	0.04	0.02	0.05	-0.01	0.03	0.04	-0.02	0.01	-0.04	-0.02	-0.02	-0.05	-0.03	0.01	0.02	0.04	0.03	-0.01
	Mix B	0.00	0.03	0.04	0.03	0.04	0.02	0.01	0.02	0.06	0.02	-0.01	0.01	0.00	0.02	-0.03	-0.03	-0.02	-0.06	-0.04	-0.04	-0.02	0.06	0.02	0.00	0.01
E	Jazz A	0.03	0.02	0.03	0.06	0.03	0.07	0.09	0.02	0.07	0.08	0.06	0.03	0.06	0.01	0.03	0.03	0.06	0.07	0.06	0.07	0.01	0.00	0.03	0.05	0.07
	Jazz B	0.02	0.02	0.06	0.00	0.02	0.05	0.07	0.03	0.02	0.06	0.04	0.04	0.07	0.06	0.05	0.02	0.08	0.04	0.10	0.06	-0.04	0.03	0.03	0.05	0.04
	Rock A	0.03	0.07	0.05	0.06	0.03	0.06	0.07	0.05	0.05	0.11	0.06	0.03	0.03	0.06	0.06	0.03	0.00	0.02	0.02	0.07	0.06	0.06	0.06	0.05	0.04
	Rock B	0.07	0.05	0.06	0.02	0.05	0.08	0.07	0.03	0.11	0.05	0.05	0.08	0.08	0.07	0.06	0.02	0.03	0.06	0.07	0.05	0.05	0.07	0.05	0.06	0.05
	Mix A	0.03	0.03	0.06	0.04	0.08	0.05	0.04	0.12	0.11	0.13	0.00	0.01	0.01	0.03	0.05	0.02	-0.01	0.03	0.00	0.00	0.03	0.02	0.04	0.05	0.03
	Mix B	0.07	0.09	0.06	0.08	0.03	0.07	0.04	0.09	0.11	0.13	0.05	0.05	0.07	0.06	0.05	0.01	0.03	-0.01	0.04	0.03	0.05	0.05	0.04	0.04	0.07
A	Jazz A	0.07	0.06	0.06	0.04	0.05	0.08	0.12	0.05	0.05	0.03	0.05	0.06	0.03	0.05	0.02	0.04	0.01	0.06	0.01	0.03	0.04	0.03	0.06	0.03	0.08
	Jazz B	0.11	0.06	0.07	0.06	0.09	0.08	0.05	0.05	0.00	0.06	-0.02	-0.01	-0.01	-0.01	0.00	0.02	0.02	0.03	0.06	0.03	0.03	0.02	0.06	0.03	0.07
	Rock A	0.08	0.11	0.07	0.10	0.07	0.07	0.13	-0.02	0.01	0.04	0.07	0.08	0.06	0.14	0.12	0.02	-0.08	-0.02	0.00	0.05	0.08	0.08	0.10	0.12	0.10
	Rock B	0.11	0.11	0.14	0.04	0.08	0.07	0.01	0.03	0.09	-0.01	0.06	0.07	0.10	0.07	-0.01	0.00	0.02	-0.04	0.05	0.02	0.07	0.06	0.06	0.09	0.05
	Mix A	0.11	0.10	0.08	0.11	0.11	0.06	0.00	0.04	0.03	0.02	0.01	0.05	0.03	0.01	0.04	0.01	-0.02	0.03	0.00	0.00	0.06	0.08	0.08	0.09	0.02
	Mix B	0.07	0.06	0.06	0.14	0.10	0.05	0.03	0.03	0.08	0.02	0.03	0.00	0.06	0.08	0.03	0.02	0.01	-0.07	0.04	0.01	0.05	0.10	0.09	0.06	0.13
N	Jazz A	-0.02	-0.03	-0.05	-0.01	0.00	-0.07	-0.06	-0.06	-0.06	-0.11	-0.04	-0.01	-0.05	-0.03	-0.05	-0.04	-0.03	-0.04	-0.04	-0.02	-0.04	-0.02	-0.03	-0.02	-0.06
	Jazz B	-0.04	0.02	-0.07	-0.01	-0.08	-0.07	-0.06	-0.09	-0.05	-0.05	-0.04	-0.01	-0.04	-0.04	-0.03	-0.04	-0.05	-0.02	-0.05	0.00	0.01	-0.03	0.01	-0.02	0.00
	Rock A	-0.08	-0.04	-0.03	-0.09	0.01	-0.03	-0.05	-0.01	-0.03	-0.08	-0.08	-0.02	-0.05	-0.05	-0.08	0.03	0.06	0.03	0.03	0.03	-0.04	-0.05	-0.05	-0.03	-0.04
	Rock B	-0.02	0.01	-0.08	0.02	0.01	-0.06	-0.01	0.01	-0.06	0.01	-0.03	-0.08	-0.04	-0.10	0.01	0.04	0.01	0.03	0.01	-0.01	-0.02	-0.03	-0.02	-0.02	-0.02
	Mix A	0.00	-0.04	-0.04	-0.06	-0.02	-0.04	-0.01	-0.09	-0.06	-0.08	-0.01	-0.01	-0.02	-0.03	-0.06	0.00	0.02	0.01	0.03	0.02	-0.04	-0.03	-0.03	-0.05	-0.03
	Mix B	-0.01	-0.05	-0.06	-0.01	-0.01	-0.07	-0.02	-0.06	-0.07	-0.08	-0.01	0.01	-0.08	-0.06	-0.02	0.01	-0.02	0.04	0.02	0.00	-0.06	-0.05	-0.05	-0.05	-0.01

0.2
0.15
0.1
0.05
0
-0.05
-0.1

Study 2: Musical Facebook Likes predict personality traits

The results from Study 1 indicated that preferences for unfamiliar musical stimuli contain some valid information about personality. The aim of Study 2 was to replicate and extend these results to real-world behavior by investigating whether naturally-occurring statements of musical preferences, as represented by Facebook Likes of musical artists, also predict personality traits.

The Like feature is a mechanism used by Facebook users to publicly express their positive association with online content, by generating a digital record that is accessible to their friends, Facebook, software developers who provide services to users, as well as outside parties, including governments and industries. Facebook Likes represent a very generic class of digital records, similar to Web search queries or credit card purchases, and are used to signal positive associations with many different types of content, including photos, friends' status updates, and Facebook pages of products, sports, books, restaurants, popular websites, and musicians. There is evidence that Facebook Likes, in general, contain information about many personal attributes, from religiosity and political views to sexual orientation and personality (Kosinski, Stillwell, and Graepel 2013). However, that work examined Likes in general, irrespective of content, so it is not clear whether Likes for specific types of content are reliably associated with personality.[12] Thus, Study 2 not only evaluates the generalizability of Study 1 to behavioral indicators of musical preferences, it also examines whether, and to what extent, Likes of musical artists alone betray information about the personalities of Facebook users.

[12] Various Facebook likes were shown to contain personality-related information, but many of the findings are difficult to interpret. For a general list of facebook likes that are most predictive of personality traits, see http://www.pnas.org/content/suppl/2013/03/07/1218772110.DCSupplemental/st01.pdf

Methods

Participants. We used data from a sample of 21,929 MyPersonality users (65% females), with a median age of 21 (interquartile distance = 5).[13] The study included all of the participants in the MyPersonality database who (1) completed a Big Five personality questionnaire (2) had at least 20 "Likes" of musical artists that were used for personality prediction (further details below) and (3) shared information about their age and gender.

Personality. Study2 used the same IPIP measure described in Study 1 (Goldberg et al., 2006).

Musical Facebook Likes. In order to focus our analysis on the predictive power of musical artists Likes, we first filtered out all the Likes that were not categorized by Facebook as music-related. We then searched all of the remaining Likes in EchoNest (http://the.echonest.com/) - a major online musical database containing over 3 million artists, and excluded all Likes that did not appear in the database. Next, we excluded all users that had less than 20 Likes, and included only artists that had at least 20 Likes.[14] This resulted a large, sparse logical matrix L, in which each row r represented a participant and each column c represented an artist, such that L(r,c) equals one if participant r Likes artist c and zero otherwise. The matrix L has dimensions 21,929 (users) x 62,036 (artists).

Prediction algorithm. For each of the Big Five personality traits, we conducted three out-of-sample predictions based on 1) the musical Likes matrix; 2) the Like matrix, gender and age; and 3) gender and age alone.

Predictions were carried out using the following procedure:

1. We randomly split the participants into ten groups, in a similar fashion to Study 1.
2. For each of the 10-holdout groups, we reduced the dimensionality of the liking matrix L to *N users x 500*, by performing sparse singular value decomposition (SVD) on the remaining 90% of the data.
3. For each of the 10-holdout groups, we trained a linear model to predict each of the Big Five personality traits, by fitting a linear regression with a LASSO penalty to the remaining 90% of the data. The tuning parameter, λ was optimized via 10-fold cross validation, performed within each training set, in a similar fashion to Study 1. All of the independent variables were standardized priori model training, as penalized regression models (such as the LASSO) are sensitive to the scale of the inputs.
4. Using the trained model, we conducted out-of-sample prediction on the 10% of the data that comprised the holdout group.[15] We estimated the goodness of fit by calculating the Pearson's correlation between the actual and predicted personalities.

---

[13] A small proportion of the participants of study 2 (roughly 3.5%) were also in the subject pool of study 1.

[14] The study's sample was constructed in the following manner: (1) users with less than 20 musical artist likes were excluded (2) artists with less than 20 likes among the remaining users were excluded (3) the first two steps were repeated iteratively until convergence, to ensure that each user had at least 20 likes and that each artist was associated with at least 20 users.

[15] In order to generate predictions, the liking matrix of the hold-out group was first projected onto the first 500 dimensions of the training data, calculated in step 2.

Results

Musical Facebook Likes predict personality traits

The results are summarized in Fig. 3. and Table 5. We found reliable correlations between the music Likes-based personality predictors and all of the Big Five personality traits (all p-values <0.001). Like in Study 1, the highest predictive accuracy was for openness (r(21,929)=0.30, 95% CI=[0.29 0.31]), followed by extraversion (r(21,929)=0.21, 95% CI=[0.20 0.22]), conscientiousness (r(21,929)=0.19, 95% CI=[0.17 0.20]), neuroticism (r(21,929)=0.18, 95% CI=[0.17 0.20]), and agreeableness (r(21,929)=0.17, 95% CI=[0.15 0.18]). To put our results in perspective, the predictive accuracy of the music Likes-based model for openness and neuroticism was roughly the same as a personality prediction made by a co-worker. For the other traits, accuracy ranged between 55% (agreeableness) and 77% (conscientiousness) of the accuracy of a work colleague's prediction (Youyou, Kosinski, and Stillwell 2015).

For all the traits, the musical Likes-based predictors were substantially more accurate than the baseline demographic model (all p-values <0.001, Steiger's z-test). Adding Likes to the baseline model significantly improved the results for all traits but neuroticism (see Fig. 3, (all p's<0.001, Steiger's z-test).[16] For neuroticism the predictive accuracies of the Like-based model and the demographic model were similar, suggesting that the former model's predictive accuracy stems from information that is also captured by age and gender.

Personality inferences based on Facebook Likes were more accurate, on average, compared to inferences based on active listening. As Facebook Likes contain meta information about the performing artist that goes beyond the pure auditory content, this finding is somewhat unsurprising. However, this finding should be interpreted with caution, as we cannot rule out the possibility that other factors, that are not directly related to metadata (e.g., differences in measurement error between these two types of variables), underlie the differences in predictive accuracy between the models.[17]

The results indicate that Facebook Likes of musical artists carry personality-relevant information. However, they do not allow us to tease apart the different contributions of particular musical tastes (i.e., liking of specific artists) from more general tendencies, such as a general tendency to like musical Facebook pages (e.g., high openness individuals tend to like more artists), or an inclination to like popular pages (e.g., agreeable individuals tend to like artists that are liked by others). To investigate this issue, we constructed an additional "general baseline model", predicting each of the Big Five traits, using (i) *# of Likes*: a single scalar denoting the total amount of musical artists that each participant liked, and (ii) *popularity score*: a scalar

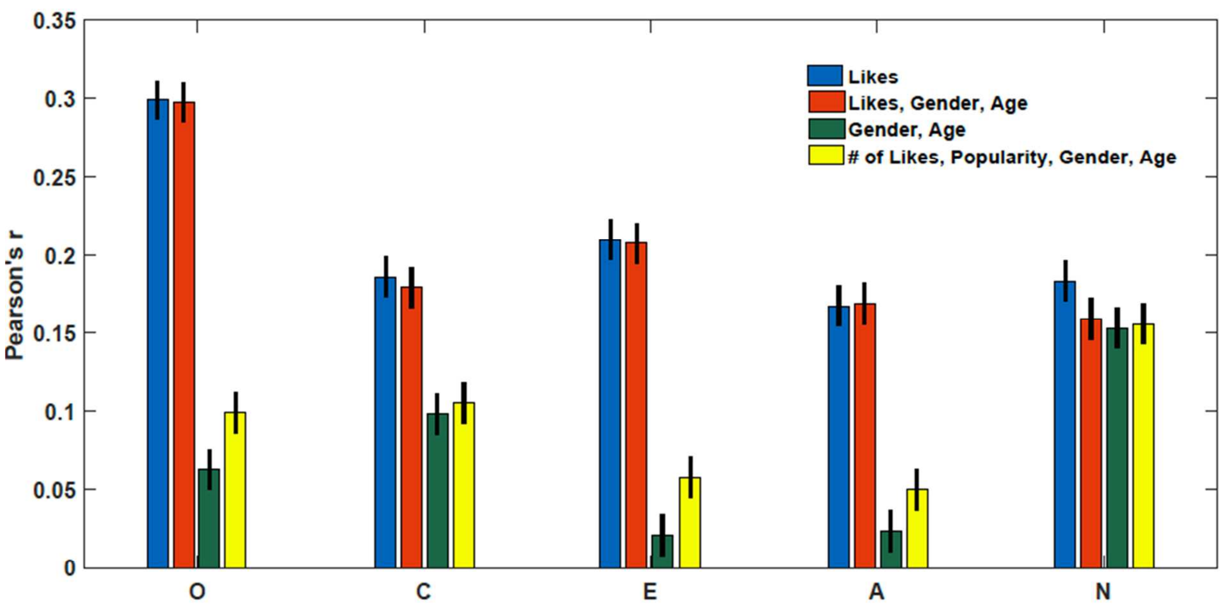
[16] Somewhat surprisingly, the predictive accuracy of the "likes only" model was slightly greater than the accuracy of the full model containing likes, gender and age, for some of the traits (Fig. 3). A possible account (apart from sampling error), is that age and gender are highly predictable from facebook likes (Kosinski, Stillwell, and Graepel 2013), which might have generated multicollinearity in the full model (Chong and Jun 2005). Note that the "likes only" model contains orthogonal features constructed using SVD.

[17] When repeating the analysis of study 2 using a subsample of 17,904 participants (the sample size of study 1 for the Mix_A survey), the predictive accuracy of the Likes-based models only slightly decreased, and was still greater than the excerpt-based models (see Table 5, bottom row). This suggests that the larger training set was not the main cause for the superior performance of the Like-based models.

denoting the average popularity across all artists liked by the user, such that popularity was defined as the logged number of Likes that an artist had across the study's participants, normalized by the total number of users, and (iii) gender and age.

Contrasting the predictive accuracies of the model that include individual Likes, gender and age (Table 5, row 2) with the general baseline model (Table 5, row 4), allows us to disentangle the predictive accuracies arising from specific, versus general, musical liking tendencies. We find that the additional predictive capacities obtained by including the individual Likes were substantial for four of the traits. For openness, we found $\Delta r = 0.20$ (200% increase), followed by extraversion ($\Delta r = 0.15$, 260%), agreeableness ($\Delta r = 0.12$, 240%), and conscientiousness ($\Delta r = 0.07$, 70%). The increase was not pronounced for neuroticism ($\Delta r < 0.01$, only 2%).[18] In conclusion, the majority of predictive personality information can be attributed to individual Likes rather than general tendencies, with the exception of neuroticism.

Figure 3: Correlations between Facebook music Likes-based big five personality predictors (out of sample) and actual personalities (*N*=21,929). Error bars denote 95% CIs.



[18] The general baseline model was inferior to a model that also included the ratings of individual items, for all of the traits except neuroticism (all p's<0.001, Steiger's z-test).

Table 5: Predictive accuracy of music artist Like based personality predictors (out of sample), for all big 5 traits (*N=21,929*).

Likes	Openness					Conscientiousness					Extraversion					Agreeablness					Neuroticism				
	N	R	95% CI		p	N	R	95% CI		p	N	R	95% CI		p	N	R			p	N	R	95% CI		p
Music likes	21,929	.30	.29	.31	<.001	21,929	.19	.17	.20	<.001	21,929	.21	.20	.22	<.001	21,929	.17	.15	.18	<.001	21,929	.18	.17	.20	<.001
Likes+Gender+Age	21,929	.30	.29	.31	<.001	21,929	.18	.17	.19	<.001	21,929	.21	.19	.22	<.001	21,929	.17	.16	.18	<.001	21,929	.16	.15	.17	<.001
Gender+Age	21,929	.04	.03	.06	<.001	21,929	.12	.11	.14	<.001	21,929	.00	-.02	.01	.924	21,929	.04	.02	.05	<.001	21,929	.17	.15	.18	<.001
# Likes + Popularity + Gender + Age	21,929	.10	.09	.11	<.001	21,929	.11	.09	.12	<.001	21,929	.06	.04	.07	<.001	21,929	.05	.04	.06	<.001	21,929	.16	.14	.17	<.001
Music likes, downsampled	17,904	.29	.28	.31	<.001	17,904	.18	.17	.20	<.001	17,904	.20	.19	.22	<.001	17,904	.16	.15	.18	<.001	17,904	.18	.16	.19	<.001

Discussion

Recent research has suggested that individual differences in musical preferences and personality traits are linked. Using a diverse sample composing tens of thousands of participants, we corroborated these findings and further extended them in four important ways.

First, our results show that affective reactions to 15-second excerpts of novel musical pieces, which lacked metadata information (e.g., artist name), are sufficient for predicting individual differences in personality. This finding replicated across- and within-genres, and demonstrates that preferences for the musical content itself, rather than the name of the artist, or a genre, contains sufficient information for personality inference.

Second, our study corroborates the MUSIC model's capacity to capture individual differences in preferences for Western music in a large and diverse population. Extensive research and discussions devoted to estimating the replicability of laboratory experiments in psychology, and social sciences in general, have highlighted the critical importance of replication efforts (Open Science Collaboration 2015; Nosek et al. 2015; Camerer et al. 2016; Simons 2014; Nave, Camerer, and McCullough 2015; Lane et al. 2016; Carter and McCullough 2013). Our results show that the 5-factor MUSIC model is highly replicable, providing a solid foundation for the future investigations of musical preferences and their links with other psychological constructs.

Third, we find that preferences for specific dimensions of the MUSIC model are associated with two of the Big Five traits. Preferences for sophisticated musical excerpts were related to openness to experience, whereas preferences for unpretentious excerpts were associated with extraversion.

Fourth, the present research helps to establish the external validity of the link between musical preferences and personality, by showing that personality traits can be reliably predicted both from liking ratings that follow actual listening, and also from digital records of naturally-occurring, real-world behaviors. Previous research has shown that personality can be inferred from Facebook Likes in general (Kosinski, Stillwell, and Graepel 2013), yet the mechanisms at work are poorly understood. Here, we have shown that focusing on musical preferences alone reveals valid information about users' personalities. With the growing presence of services for streaming and sharing music online, this finding has direct implications for the music industry, recommendation algorithms, and marketing practitioners (Ogden, Ogden, and Long 2011/3; Bruner 1990; Matz et al. 2017; Matz, Gladstone, and Stillwell 2016).

Our study has several important limitations, leaving open questions for future research. First, while our results demonstrated that musical preferences carry personality information that goes beyond age, gender, and general liking tendencies, we recognize that there are likely other unmeasured person-level variables (e.g., geo-location, socio-economic status, culture, and preferences for different leisure activities) that would capture at least some of this incremental variance. Moreover, our results are correlational, and therefore cannot address questions of causality. For example, it is possible that common environmental factors (e.g., peer influence) influence both personality and musical preferences.

Second, although Facebook Likes are active, naturally-occurring behaviors, they do not automatically reflect what music people actually listen to. Moreover, liking of an artist might be driven by factors other than musical taste, such as peer influence or self-image. The increasing use of music streaming services (e.g., Last FM, Spotify) is expected to allow further investigations of the links between personality and active ecologically-valid music listening behavior.

Third, although our study generalizes previous findings to populations that are beyond college students, our sample is composed of Facebook users. It is thus an open question whether our findings generalize to populations that are not represented in the current work, such as non-Western societies (Henrich, Heine, and Norenzayan 2010). Moreover, the musical pieces used in Study 1 were entirely Western in origin, so for now, the conclusions we can draw from the current findings are restricted to predominantly Western societies.

Finally, while our data-driven predictive analyses provide strong evidence supporting the link between musical preferences and personality, the results call for further development of theoretical models for identifying the mechanisms at work. One such candidate mechanism is based on self-identity motives (Abrams 2009). That is, people are drawn to musical styles that validate their self-perceptions and convey that information to others (e.g., listening to avant-garde music can serve to simultaneously reinforce and communicate the belief that one is creative and unconventional). A second mechanism is based on emotion regulation (Saarikallio and Erkkilä 2007). That is, people prefer styles of music that reinforce their mood or emotional state (e.g., listening to uplifting music may help to maintain a positive mood). A third possible mechanism is based on activity congruence, or the idea that people prefer auditory content that complements the activities they regularly pursue. For example, fast and upbeat music complements various energetic activities, from dancing to socializing, that are likely to appeal to extraverted people.

While some of our exploratory findings demonstrate associations between personality traits and components of the five-factor MUSIC model that are consistent with the above mechanisms (e.g., high openness is associated with liking of sophisticated music), and are also in accord with previous research (Schäfer and Mehlhorn 2017), the magnitudes of these associations are generally small in size. Thus, a considerable amount of variance in musical preferences remains unexplained. Future investigations concerned with musical preferences should illuminate the underlying mechanisms by investigating how preferences relate to identity motives, emotion regulation processes, and activity preferences, and also by exploring how preferences for particular auditory features (e.g, rhythm, time signature, frequency components) may correspond to different personality traits (see Logan and Others 2000; Lindenbaum et al. 2010).

In summary, we have shown that preference ratings for unfamiliar musical stimuli and naturally occurring statements of musical preferences in online social media, allow for making reliable inference of personality traits. These results corroborate that music - a form of self-expression that is ubiquitous across human cultures - communicates meaningful information about basic psychological characteristics.

Author Contributions

All authors developed the study concept and contributed to the study design. Data collection were performed by M.K and D.S, data analysis and interpretation was performed by G.N. and J.M under the supervision of J.R.; G.N. drafted the manuscript.

References

Abrams, Dominic. 2009. "Social Identity on a National Scale: Optimal Distinctiveness and Young People's Self-Expression through Musical Preference." *Group Processes & Intergroup Relations: GPIR* 12 (3 Spec). Sage Publications: 303–17.

Bonneville-Roussy, Arielle, Peter J. Rentfrow, Man K. Xu, and Jeff Potter. 2013. "Music through the Ages: Trends in Musical Engagement and Preferences from Adolescence through Middle Adulthood." *Journal of Personality and Social Psychology* 105 (4). psycnet.apa.org: 703–17.

Brown, R. A. 2012. "Music Preferences and Personality among Japanese University Students." *International Journal of Psychology: Journal International de Psychologie* 47 (4). Taylor & Francis: 259–68.

Bruner, Gordon C. 1990. "Music, Mood, and Marketing." *Journal of Marketing* 54 (4). American Marketing Association: 94–104.

Camerer, Colin F., Anna Dreber, Eskil Forsell, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, et al. 2016. "Evaluating Replicability of Laboratory Experiments in Economics." *Science* 351 (6280). science.sciencemag.org: 1433–36.

Camerer, Colin F., Gideon Nave, and Alec Smith. 2017. "Dynamic Unstructured Bargaining with Private Information and Deadlines: Theory, Experiment, and Outcome Prediction via Machine Learning." *Manuscript Submitted for Publication*.

Carter, Evan C., and Michael E. McCullough. 2013. "Is Ego Depletion Too Incredible? Evidence for the Overestimation of the Depletion Effect." *The Behavioral and Brain Sciences* 36 (6): 683–84; discussion 707–26.

Chong, Il-Gyo, and Chi-Hyuck Jun. 2005. "Performance of Some Variable Selection Methods When Multicollinearity Is Present." *Chemometrics and Intelligent Laboratory Systems* 78 (1). Elsevier: 103–12.

Collaboration, Open Science, and Others. 2015. "Estimating the Reproducibility of Psychological Science." *Science* 349 (6251). American Association for the Advancement of Science: aac4716.

Colley, Ann. 2008. "Young People's Musical Taste: Relationship With Gender and Gender-Related Traits." *Journal of Applied Social Psychology* 38 (8). Wiley Online Library: 2039–55.

Delsing, Marc J. M. H., Tom F. M. ter Bogt, Rutger C. M. E. Engels, and Wim H. J. Meeus. 2008. "Adolescents' Music Preferences and Personality Characteristics." *European Journal of Personality* 22 (2). John Wiley & Sons, Ltd.: 109–30.

George, Darren, Kelly Stickle, Faith Rachid, and Alayne Wopnford. 2007. "The Association between Types of Music Enjoyed and Cognitive, Behavioral, and Personality Factors of Those Who Listen." *Psychomusicology: A Journal of Research in Music Cognition* 19 (2). Music and Arts Publications: 32.

Goldberg, Lewis R., John A. Johnson, Herbert W. Eber, Robert Hogan, Michael C. Ashton, C. Robert Cloninger, and Harrison G. Gough. 2006/2. "The International Personality Item Pool and the Future of Public-Domain Personality Measures." *Journal of Research in Personality* 40 (1). Elsevier: 84–96.

Greenberg, David M., Simon Baron-Cohen, David J. Stillwell, Michal Kosinski, and Peter J. Rentfrow. 2015. "Musical Preferences Are Linked to Cognitive Styles." *PloS One* 10 (7). journals.plos.org: e0131151.

Greenberg, David M., Michal Kosinski, David J. Stillwell, Brian L. Monteiro, Daniel J. Levitin, and Peter J. Rentfrow. 2016. "The Song Is You Preferences for Musical Attribute Dimensions Reflect Personality." *Social Psychological and Personality Science*. SAGE Publications, 1948550616641473.

Henrich, Joseph, Steven J. Heine, and Ara Norenzayan. 2010. "The Weirdest People in the World?" *The Behavioral and Brain Sciences* 33 (2-3). Cambridge Univ Press: 61–83.

Kosinski, Michal, David Stillwell, and Thore Graepel. 2013. "Private Traits and Attributes Are Predictable from Digital Records of Human Behavior." *Proceedings of the National Academy of Sciences of the United States of America* 110 (15): 5802–5.

Lane, A., O. Luminet, G. Nave, and M. Mikolajczak. 2016. "Is There a Publication Bias in Behavioural Intranasal Oxytocin Research on Humans? Opening the File Drawer of One Laboratory." *Journal of Neuroendocrinology* 28 (4). Wiley Online Library. doi:10.1111/jne.12384.

Langmeyer, Alexandra, Angelika Guglhör-Rudan, and Christian Tarnai. 2012. "What Do Music Preferences Reveal About Personality?" *Journal of Individual Differences* 33 (2). econtent.hogrefe.com: 119–30.

Lindenbaum, Ofir, Shai Maskit, Ophir Kutiel, and Gideon Nave. 2010. "Musical Features Extraction for Audio-Based Search." In *Electrical and Electronics Engineers in Israel (IEEEI), 2010 IEEE 26th Convention of*, 000087–000091. IEEE.

Logan, Beth, and Others. 2000. "Mel Frequency Cepstral Coefficients for Music Modeling." In *ISMIR*. cs.swarthmore.edu. http://web.cs.swarthmore.edu/~turnbull/cs97/f09/paper/logan00.pdf.

Matz, Sandra C., Joe J. Gladstone, and David Stillwell. 2016. "Money Buys Happiness When Spending Fits Our Personality." *Psychological Science* 27 (5). pss.sagepub.com: 715–25.

Matz, S., M. Kosinski, G. Nave, and D. Stillwell. 2017. "Psychographic Targeting as an Effective Approach to Digital Mass Persuasion." Proceedings of the National Academy of Science, in press.

Miranda, Dave, and Michel Claes. 2008. "Personality Traits, Music Preferences and Depression in Adolescence." *International Journal of Adolescence and Youth* 14 (3). Taylor & Francis: 277–98.

Nave, Gideon, Colin Camerer, and Michael McCullough. 2015. "Does Oxytocin Increase Trust in Humans? A Critical Review of Research." *Perspectives on Psychological Science: A Journal of the Association for Psychological Science* 10 (6): 772–89.

Nosek, B. A., G. Alter, G. C. Banks, D. Borsboom, S. D. Bowman, S. J. Breckler, S. Buck, et al. 2015. "SCIENTIFIC STANDARDS. Promoting an Open Research Culture." *Science* 348 (6242). ncbi.nlm.nih.gov: 1422–25.

Ogden, James R., Denise T. Ogden, and Karl Long. 2011/3. "Music Marketing: A History and Landscape." *Journal of Retailing and Consumer Services* 18 (2). Elsevier: 120–25.

Palmer, Stephen E., and William S. Griscom. 2013. "Accounting for Taste: Individual Differences in Preference for Harmony." *Psychonomic Bulletin & Review* 20 (3). Springer: 453–61.

Reise, Steven P., Tyler M. Moore, and Mark G. Haviland. 2010. "Bifactor Models and Rotations: Exploring the Extent to Which Multidimensional Data Yield Univocal Scale Scores." *Journal of Personality Assessment* 92 (6). Taylor & Francis: 544–59.

Rentfrow, Peter J., Lewis R. Goldberg, and Daniel J. Levitin. 2011. "The Structure of Musical Preferences: A Five-Factor Model." *Journal of Personality and Social Psychology* 100 (6): 1139–57.

Rentfrow, Peter J., Lewis R. Goldberg, David J. Stillwell, Michal Kosinski, Samuel D. Gosling, and Daniel J. Levitin. 2012. "The Song Remains the Same: A Replication and Extension of the MUSIC Model." *Music Perception* 30 (2). mp.ucpress.edu: 161–85.

Rentfrow, Peter J., and Samuel D. Gosling. 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). psycnet.apa.org: 1236–56.

Saarikallio, Suvi, and Jaakko Erkkilä. 2007. "The Role of Music in Adolescents' Mood Regulation." *Psychology of Music* 35 (1). SAGE Publications: 88–109.

Schäfer, Thomas, and Claudia Mehlhorn. 2017. "Can Personality Traits Predict Musical Style Preferences? A Meta-Analysis." *Personality and Individual Differences* 116 (October). Elsevier: 265–73.

Simons, Daniel J. 2014. "The Value of Direct Replication." *Perspectives on Psychological Science: A Journal of the Association for Psychological Science* 9 (1): 76–80.

Steiger, James H. 1980. "Tests for Comparing Elements of a Correlation Matrix." *Psychological Bulletin* 87 (2). American Psychological Association: 245.

Stone, Mervyn. 1974. "Cross-Validatory Choice and Assessment of Statistical Predictions." *Journal of the Royal Statistical Society. Series B, Statistical Methodology*. JSTOR, 111–47.

Tibshirani, Robert. 1996. "Regression Shrinkage and Selection via the Lasso." *Journal of the Royal Statistical Society. Series B, Statistical Methodology*. JSTOR, 267–88.

Vukasović, Tena, and Denis Bratko. 2015. "Heritability of Personality: A Meta-Analysis of Behavior Genetic Studies." *Psychological Bulletin* 141 (4). psycnet.apa.org: 769–85.

Yarkoni, T., and J. Westfall. 2017. “Choosing Prediction over Explanation in Psychology: Lessons from Machine Learning.” *Perspectives on Psychological Science*, in press

Yeoh, Joanne P. S., and Adrian C. North. 2010. “The Effect of Musical Fit on Consumers’ Memory.” *Psychology of Music* 38 (3). SAGE Publications: 368–78.

Youyou, Wu, Michal Kosinski, and David Stillwell. 2015. “Computer-Based Personality Judgments Are More Accurate than Those Made by Humans.” *Proceedings of the National Academy of Sciences of the United States of America* 112 (4): 1036–40.