AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: Applied

ISSN: 1076-898X

https://doi.org/10.1037/xap0000447

AI Composer Bias: Listeners Like Music Less When They Think It Was Composed by an AI

Daniel B. Shank, Courtney Stefanik, Cassidy Stuhlsatz, Kaelyn Kacirek, and Amy M. Belfi
Department of Psychological Science, Missouri University of Science and Technology

The use of artificial intelligence (AI) to compose music is becoming mainstream. Yet, there is a concern that listeners may have biases against AIs. Here, we test the hypothesis that listeners will like music less if they think it was composed by an AI. In Study 1, participants listened to excerpts of electronic and classical music and rated how much they liked the excerpts and whether they thought they were composed by an AI or human. Participants were more likely to attribute an AI composer to electronic music and liked music less that they thought was composed by an AI. In Study 2, we directly manipulated composer identity by telling participants that the music they heard (electronic music) was composed by an AI or by a human, yet we found no effect of composer identity on liking. We hypothesized that this was due to the "AI-sounding" nature of electronic music. Therefore, in Study 3, we used a set of "human-sounding" classical music excerpts. Here, participants liked the music less when it was purportedly composed by an AI. We conclude with implications of the AI composer bias for understanding perception of AIs in arts and aesthetic processing theories more broadly.

Public Significance Statement
Artificial intelligence (AI)—computers making intelligent decisions or emulating humans—is revolutionizing the music industry. Yet, very little is known about how people emotionally respond to AI-generated music. The findings of the present work indicate that listeners tend to be biased against music that they think was created by an AI if the music itself does not fit expectations of what an AI could create.

Keywords: music, artificial intelligence, liking, aesthetics, algorithms

Supplemental materials: https://doi.org/10.1037/xap0000447.supp

Artificial Intelligence (AI)—widely defined as computer systems that make intelligent decisions or emulate humans—is a growing field that is changing the way humans live every day, including the ways we listen to music. For example, music streaming platforms such as Spotify and Pandora use AI algorithms to recommend new music to listeners based on their previous listening behavior. Generally, users welcome such applications of AI to improve their music listening experience (Jones & Pu, 2008; McCourt & Zuberi, 2016). In contrast, users may be hesitant about listening to music that was *created* or *composed* by an AI. For example, recent survey research suggests that both the general public and music professionals have negative perceptions of AI-created music and express a low likelihood of purchasing music created by AIs (Tigre Moura & Maw, 2021). In this prior work, participants were asked to rate their attitudes towards AI-composed music and their intention to purchase AI-composed music *in general,* but this study did not consider people's attitudes toward specific songs purportedly created by AIs. Their ratings reflected negative attitudes and low intention to purchase music in general that is created by AIs.

Amy M. Belfi https://orcid.org/0000-0002-1469-4117

The authors would like to acknowledge the following individuals for their assistance on this project: Carson Arnold, Emily Swaters, Lawrence Hierlmeier, Mallory North, and Claire Neill. This work was supported by funding from the Missouri S&T Intelligent Systems Center. The authors have no conflicts of interest to disclose.

Study design, hypotheses, and analytic plans were not preregistered. All data reported in the present article can be found at the following Open Science Framework (OSF) repository: https://osf.io/x8kqs/.

Daniel B. Shank played equal role in conceptualization, funding acquisition, investigation, methodology, project administration, supervision, writing of original draft and writing of review and editing. Courtney Stefanik played supporting role in investigation, writing of original draft and writing of review and editing. Cassidy Stuhlsatz played supporting role in investigation, writing of original draft and writing of review and editing. Kaelyn Kacirek played supporting role in formal analysis, software, writing of original draft and writing of review and editing. Amy M. Belfi played lead role in data curation, formal analysis, resources, software and visualization and equal role in conceptualization, funding acquisition, investigation, methodology, project administration, supervision, writing of original draft and writing of review and editing.

Correspondence concerning this article should be addressed to Amy M. Belfi, Department of Psychological Science, Missouri University of Science and Technology, 132 Humanities & Social Sciences, 500 West 14th Street, Rolla, MO 65409, United States. Email: amybelfi@mst.edu

Despite this hesitance from both music professionals and the general public, the field of AI is swiftly working towards creating music using AI. AI tools made by Google, OpenAI, and others are not only assisting musicians in creating new music, but are composing their own completely new music, including deepfakes that realistically emulate artists ranging from Bach to Nirvana. For example, DeepBach is an AI trained to emulate Bach chorales. When asked to identify whether a piece of music composed by DeepBach was created by a computer or Bach, around 50% of participants judge them as Bach-composed, with trained musicians being less likely to identify a DeepBach piece as composed by Bach than naïve listeners (Hadjeres et al., 2017). Bachbot is a similar AI designed to emulate Bach chorales. Participants were not able to distinguish a true Bach composition from a Bachbot composition at levels above chance (Liang et al., 2017). That is, listeners found Bachbot nearly indistinguishable from real Bach music. Thus, while the field of AI is swiftly improving on the technical challenges of programming AIs to compose music, an important question remains: Do people enjoy music less if they think it was created by an AI? First, we will situate this question in the broader literature on aesthetic judgments of music and of products created by AIs, before empirically investigating this question in a series of three studies.

Contextual Influences on Aesthetic Judgments of Music

One long-standing debate in the field of empirical aesthetics surrounds the relative contributions of stimulus properties to aesthetic judgments. That is, is it possible to distill down the sensory features of a stimulus that are universally considered aesthetically appealing, or are aesthetic judgments predominantly influenced by extrastimulus information? While historically, most empirical research on aesthetics has focused on stimulus properties and/or artworks in isolation, more recently, scholars have begun embracing the role that context and extrastimulus features play in aesthetic judgments (Leder & Pelowski, 2021). Several major theories of aesthetic appreciation and judgment include contextual information (for visual-arts specific theories, see Chatterjee & Vartanian, 2014; Leder & Nadal, 2014; Pelowski et al., 2017; for a music-specific theory, see Brattico et al., 2013). Empirical work on the subject has identified that contextual information, including information about the creator of an artwork, and whether an artwork is framed as “art” or “not art,” can influence a viewer or listener’s aesthetic judgments of that object (for a thorough review of framing effects in aesthetic judgments, see Leder & Pelowski, 2021).

Specifically regarding music, prior work has indicated that musical preferences can be influenced by many types of contextual or extramusical information (North, 2010), including social factors like group affiliation (Lonsdale & North, 2009, 2017) or environmental factors such as time of year (Krause & North, 2018). One particularly relevant contextual factor that can affect aesthetic judgments of music is knowing who composed or performed the music. Infamously, auditions for orchestras were highly biased against female musicians until the introduction of blind auditions mitigated that bias (Goldin & Rouse, 2000). Listeners have also been shown to like music more when they believe it was composed by Mozart versus an unfamiliar composer (Fischinger et al., 2018). Liking is also influenced by the psychological characteristics of the composer: that is, people prefer music that is composed by artists who have a similar personality to their own (Greenberg et al., 2021). In addition, a congruence between musical style and performer can impact liking. For example, listeners like U.S. patriotic music more when it is played by a U.S. military band rather than a university band (Belfi et al., 2021). Taken together, these results indicate that the aesthetic appeal of music can be influenced by the identity of the composer and performer of the music.

Recent theories of aesthetic judgments have also proposed that one’s preexisting schemas about contextual information interact with the perceived self-relevance of an artwork to influence aesthetic judgments. That is, if an artwork both fits with one’s prior schemas *and* is seen as self-relevant, this will result in positive aesthetic judgments of the work (Pelowski et al., 2017). Perhaps some of the aversion around AI-created music is that people have strong emotional responses to music and find music particularly central to their identities and sense of self (Lamont & Loveday, 2020; Lonsdale & North, 2017; Peck & Grealey, 2020). While seemingly obvious, prior research indicates that the primary reason people listen to music is because they enjoy it (Sanfilippo et al., 2020). Furthermore, almost everyone experiences strong emotional responses to music, with only around 3% of the population reporting that they do not experience pleasure from music (for review, see Belfi & Loui, 2020). Perhaps it is the strong connections people feel towards music that causes skepticism about AI-generated music. Listeners may have preexisting expectations about what type and quality of music an AI could create, and high-quality AI-generated music challenges their preexisting schemas.

Aesthetic Judgments of Products Created by AI

Across many domains, people tend to judge AIs differently than humans, even when in identical situations. For example, negative outcomes produced by the actions of AIs are judged less harshly and forgiven more than humans (Bigman et al., 2020; Furlough et al., 2021; Malle et al., 2015; Shank et al., 2019). However, when an algorithm and a human both make the same mistake, people are more likely to rely on the human than the algorithm moving forward (Bigman et al., 2020; Dietvorst et al., 2015). This may be partly due to the fact that AIs are not viewed as having the ability to feel emotion or have experiences (Gray et al., 2007). Despite this perceived lack of emotion, AIs are still expected to be agentic in completing tasks (Gray et al., 2007). However, whether or not an AI is seen as suitable for a given task may depend in part on whether the task involves a subjective, aesthetic judgment. Some people believe that AIs are incapable of completing more subjective tasks (e.g., providing dating advice) and therefore people are less likely to accept AI recommendations on such tasks (Castelo et al., 2019). In a study about AI recommendations on shopping websites, people preferred AI over human recommenders when choosing products based on their utilitarian attributes (e.g., recommending winter coats based on warmth) but preferred human recommenders when choosing products based on aesthetic attributes (e.g., recommending winter coats based on appearance; Longoni & Cian, 2020).

In addition to differences in the perceived utility of AI for *recommending* products based on aesthetics, differences are found in responses to aesthetic products *created* by AIs. For example, prior work has indicated that people find music, cooking recipes, and paintings created by AIs to be less authentic than those created by humans (Jago, 2019). In the advertising industry, people prefer advertisements created by AIs when they believe that the creation of the advertisements is more objective than creative (Wu & Wen, 2021). Individuals are also more likely to view poems and paintings created by an AI as less favorably than ones created by humans (Wu et al., 2020). People also prefer human-created poetry, even though individuals were unable to reliably distinguish between human and AI-created poems (Köbis & Mossink, 2021). Additionally, pieces of visual artwork labeled as being created by a computer were rated as less aesthetically appealing than those labeled as sourced from an art gallery (Kirk et al., 2009). Notably, the response to AI-created aesthetics also depends on people's perception of the AI. For example, when an AI system has been described as a tool, it is assigned less responsibility for the creation of a piece of artwork than when it is anthropomorphized and made to seem more human (Epstein et al., 2020). In sum, this work suggests that people judge both the actions and products of AIs more negatively than those of humans.

The Present Study

Taken together, this prior work on aesthetics and AI has led to what we deem as the "AI composer bias." We hypothesize that people will like music less if they believe that the music was composed by an AI. It is important to note that our hypothesis is not founded on music that is *actually composed* by an AI versus that which is composed by a human. Instead, we predict that, if given an identical piece of (human-composed) music, a listener will like it less if they are *simply told* it was composed by an AI. In a series of three studies, we probe this hypothesis in different ways. In Study 1, participants listened to excerpts of music (classical and electronic) and made two judgments. They first identified whether they thought the music was composed by an AI or by a human and then they rated how much they liked the music. In Studies 2 (electronic music) and 3 (classical music), we directly manipulated composer identity by telling participants that the music they heard was either composed by an AI or by a human. The overall goal of the present work was to identify whether the presumed identity of a composer (e.g., AI or human) influences a listener's aesthetic judgments of that music, and if so, under what conditions.

Study 1

The goals of Study 1 were to identify whether listeners reliably perceive certain pieces of music as sounding like they were composed by an AI versus a human and to assess whether these assumed composer identities were related to how much listeners liked the pieces of music. Participants listened to either 20 electronic or 20 classical music excerpts. We hypothesized that listeners would agree on pieces that were AI-sounding and others that were human-sounding, and that listeners would like human-sounding pieces more than AI-sounding pieces.

Study 1: Method

Transparency and Openness

Data. All data reported in the present article can be found at the following Open Science Framework (OSF) repository: https://osf.io/x8kqs/

Analytic Methods and Materials. Analytic code and materials will be made available upon request.

Participants

Participants were recruited using Prolific, a website designed to recruit participants for research studies (Palan & Schitter, 2018; Peer et al., 2017). We recruited 150 participants in each group (electronic and classical). Participants were grouped such that each person only rated a single genre to prevent cross-genre comparisons. We based this sample size on our prior work using the same stimuli and a similar task (Belfi, 2019; Belfi et al., 2018). Our prior work using the same stimuli in the present study (Belfi et al., 2018) found a large effect size ($\eta^2 = 0.20$) when comparing aesthetic ratings for classical versus electronic musical excerpts, with a sample size of 20 participants. We are therefore confident that our number of participants has sufficient power to detect our predicted effects.

Of the 150 participants in the electronic group, we excluded four participants for the following reasons: Two participants were removed for submitting the survey entirely blank, and two participants for correctly reporting that they recognized the composer of one of the pieces of music. The remaining 146 participants (59 men, 86 women, 1 nonbinary) were an average of 31.52 years old ($SD = 11.26$), had an average of 15.12 years of education ($SD = 2.72$), and had an average of 3.29 years of formal musical training ($SD = 4.67$). Of the 150 participants in the classical group, we excluded one participant for failing the attention/sound check. The remaining 149 participants (66 men, 82 women, 1 nonbinary) were an average of 30.65 years old ($SD = 10.51$), had an average of 14.92 years of education ($SD = 2.42$), and had an average of 3.01 years of formal musical training ($SD = 4.24$). The participants in the electronic and classical groups did not significantly differ in terms of gender ($\chi^2 = 0.45$, $p = .49$); age, $t(293) = -0.68$, $p = .49$; education, $t(292) = 0.65$, $p = .51$; or years of musical training, $t(293) = -0.54$, $p = .58$. Additionally, both the classical and electronic groups had relatively little prior familiarity with AI-composed music (11% of classical 6% of electronic participants indicated "I had heard of it and have read some about it and/or listened to it before"), with no differences between the groups in their levels of familiarity with AI-composed music ($\chi^2 = 1.56$, $p = .21$).

Stimuli

Stimuli, selected from our prior work (Belfi, 2019; Belfi et al., 2018; Kasdan & Belfi, 2020), consisted of 15s excerpts from electronic and classical music. Twenty excerpts that contained no lyrics or human vocalizations were chosen for each genre. No particularly well-known pieces were selected in an effort to ensure the unfamiliarity of the pieces. The electronic pieces consisted of electronic dance music with a distinctive beat structure (i.e., no ambient music). All electronic pieces were contemporary and ranged in tempo from 60 to 150 beats per minute ($M = 111.05$, $SD = 23.41$). The classical pieces consisted of 19th century small ensemble music of the Romantic era. All pieces were from the European tradition and ranged in tempo from 69 to 160 beats per minute ($M = 117.15$, $SD = 28.00$). A list of all musical pieces used in the present study can be found in the Supplemental Tables S1–S2.

Procedure

All procedures were conducted in compliance with the American Psychological Association Ethical Principles and were approved by the institutional review board of the University of Missouri. Before the study began, participants were given a brief description of the task and gave their consent to participate. Participants were instructed to wear headphones in order to participate in the study. As an attention/sound check, participants were told to listen to an audio clip that stated, "Please type the word 'banana' in the box below." Participants who did not correctly type this word were excluded from the study. Participants were then given instructions for the task and were presented with the following definition of AI:

> An AI performs behaviors which are considered intelligent if performed by humans, learns or changes based on new information or environments, generalizes to make decisions based on limited information, or makes connections between otherwise disconnected people, information, or other agents. AI is sometimes used to compose music.

Each participant listened to the 20 musical excerpts belonging to their musical genre group (electronic or classical). Excerpts were presented one at a time in a random order. On each trial, participants were first required to listen to the entirety of the musical excerpt before advancing to the next page, which contained the questions. Participants had the option to listen to the excerpts more than once. After listening to each excerpt, participants were then asked to choose whether they thought the excerpt was composed by a human or by an AI, in a two-alternative forced-choice question. Participants were asked to rate their confidence in their composer choice from *not at all confident* (coded 1) to *completely confident* (coded 5). Finally, participants were asked to rate how much they liked each clip on a 7-point Likert scale, with responses ranging from *dislike a great deal* (coded −3) to *like a great deal* (coded 3). Participants did not see the numerical ratings on the Likert scales.

After rating all 20 clips, participants answered a brief demographic survey which included questions about their age, gender, years of education, and years of formal musical training. For musical training, participants were asked: "How many years of formal musical training (e.g., music lessons) do you have?" They were asked about their familiarity level with AI-composed music: the exact question was: "How familiar were you with AI-composed music prior to this survey?" with three options: (a) I had never heard of it, (b) I had heard of it, but don't know much about it, and (c) I had heard of it and have read some about it and/or listened to it before. They were also given an opportunity to provide any additional thoughts about AI-composed music and to indicate if they were familiar with any of the pieces (a free-response question). After completing the survey, a debriefing screen was shown which included a description of the study and how the data will be used. The study took approximately 15 min, and each participant was paid $2.50 for completing the study.

Study 1: Results

We performed all statistical analyses using *R* Version 3.6.2. For each genre, we conducted a generalized linear mixed effects model using the glmer function from the lme4 package (Bates et al., 2015). These models included a fixed effect of liking ratings and random intercepts for participants and items. The outcome variable was the composer attribution ratings (AI or human); given that the dependent variable is a binomial response, we used a generalized linear mixed model. For the electronic group, this model was significant ($\beta = -0.92$, $SE = 0.06$, $z = -16.12$, $p < .001$), indicating that liking was predictive of AI versus human composer judgments. That is, musical excerpts with a high proportion of AI ratings were liked less than excerpts with a high proportion of human ratings (Figure 1A). We repeated this analysis using confidence ratings as the predictor variable. This analysis was also significant, indicating that participants were more confident when they rated pieces as human, rather than AI-composed ($\beta = -0.13$, $SE = 0.04$, $z = -3.04$, $p < .001$; Figure 1B).

For the classical group, the mixed-effect model analyses replicated those of the electronic group: pieces with a higher proportion of AI ratings were liked less ($\beta = -0.98$, $SE = 0.06$, $z = -16.53$, $p < .001$; Figure 1C), and participants were more confident when identifying pieces as having a human-sounding composer ($\beta = -0.78$, $SE = 0.06$, $z = -13.35$, $p < .001$, Figure 1D). To investigate potential effects of musical training, these analyses were repeated with years of musical training included as a fixed effect, and there were no effects of musical training (see Supplemental Materials for details).

We further explored the AI composer bias effect by categorizing the excerpts as either AI-sounding or human-sounding if over 60% of participants rated them as such. For the electronic group, this resulted in seven AI-sounding excerpts and five human-sounding excerpts (Figure 2A). A paired-samples *t* test revealed that participants liked the AI-sounding excerpts significantly less (0.04, $SD = 0.96$) than the human-sounding excerpts, 0.85, $SD = 0.85$; $t(145) = -11.203$, $p < .001$, 95% CI [−0.95, −0.66], $d = -0.92$. In contrast to the electronic group, in the classical group, only one piece was identified as AI-sounding and seven pieces were identified as human-sounding (Figure 2B). Due to the lack of AI-sounding excerpts, we were unable to conduct the liking comparison across AI versus human-sounding classical excerpts.

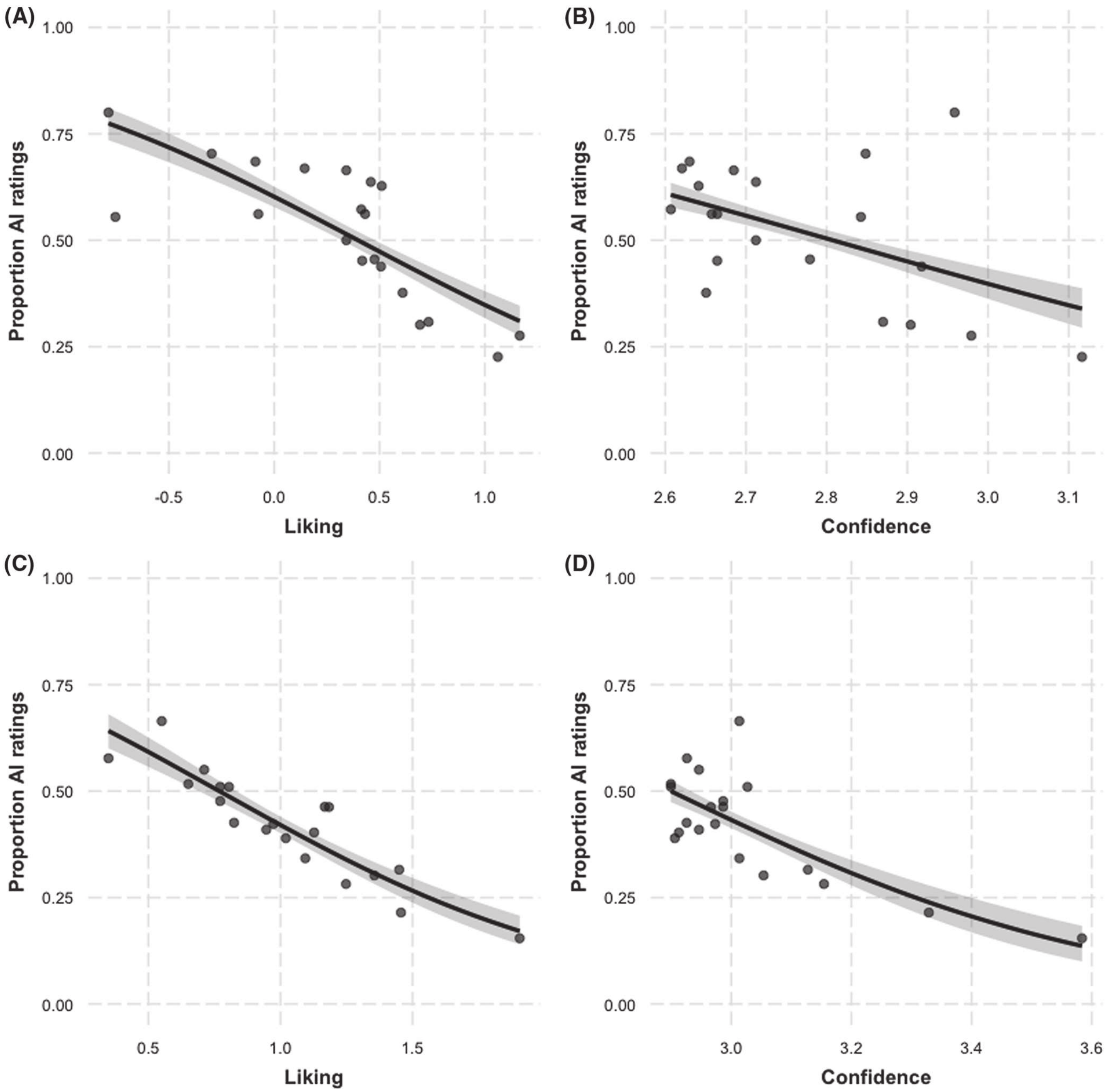
Finally, we sought to directly investigate differences between liking ratings of the electronic and classical groups. To do this, we conducted an independent-samples *t* test to investigate whether overall, participants in one group liked the excerpts more than the other. We found that the classical music was liked significantly more than the electronic music (classical mean = 1.01, $SD = 0.85$; electronic mean = 0.31, $SD = 0.80$; $t(293) = 7.28$, $p < .001$, 95% CI [0.51, 0.89], $d = 0.84$; Figure 3A). Next, we conducted an independent-samples *t* test to investigate whether confidence ratings differed between the two groups. We found that participants were significantly more confident in their composer identity ratings of the classical pieces than the electronic pieces (classical mean = 3.02, $SD = 0.80$; electronic mean = 2.77, $SD = 0.68$; $t(293) = 2.92$, $p = .003$, 95% CI [0.08, 0.42], $d = 0.34$; Figure 3B).

Musical Features Analysis

Our results from Study 1 show that certain musical pieces are consistently rated as sounding as though they were composed by an AI or by a human. We next sought to analyze the musical features of these clips to identify whether certain features were associated with more AI-sounding or human-sounding music. We chose nine musical features based on prior research evaluating the features associated with emotions in different musical genres (Eerola, 2011), which identified the most predictive features in the categories of *dynamics, rhythm, timbre,* and *tonality*. These features are described in detail elsewhere (e.g., Eerola, 2011) and are well documented in the MIR Toolbox (Lartillot et al., 2008), but we will briefly explain the features used here.

Root-mean-square (RMS) energy is a measure of the dynamics in a piece of music, calculated by computing the root average of the square of the amplitude in the signal. We calculated both the slope (RMSl) and

Figure 1
Relationships Between Liking, Confidence, and Perceived Composer Identity in Study 1

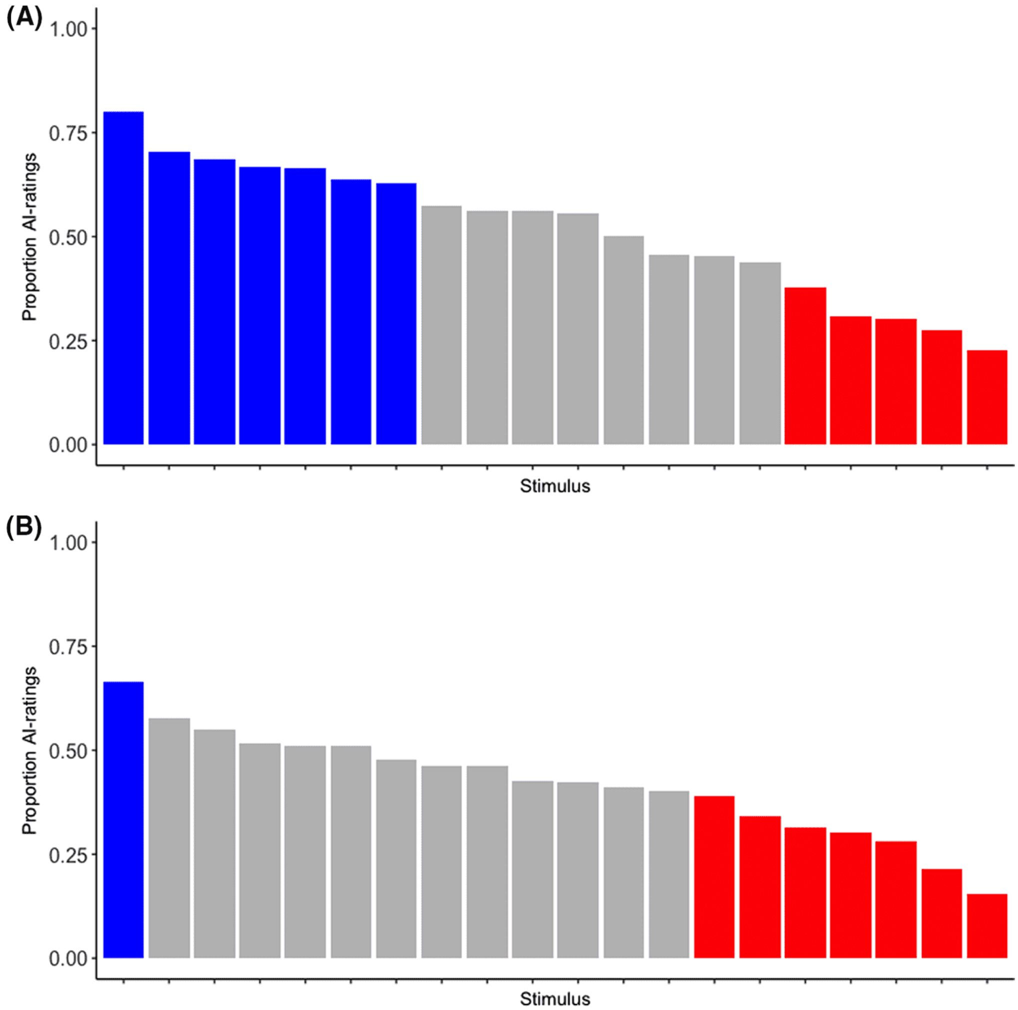


Note. (A) Electronic liking ratings, (B) electronic confidence ratings, (C) classical liking ratings, and (D) classical confidence ratings. *Y*-axis is the proportion of AI ratings. Each dot represents an individual musical excerpt. AI = artificial intelligence.

the standard deviation (RMSd) of the RMS energy. Pulse clarity (PC) estimates the rhythmic clarity, indicating the strength of the beats in a piece of music. This is estimated by the global characteristic of the autocorrelation function of the amplitude envelope (Lartillot et al., 2008). We calculated the standard deviation of PC (PCd). Tempo estimates the tempo by detecting periodicities from the event detection curve. We calculated the mean (Tm) and standard deviation of the tempo (Td). Fluctuation peak magnitude (FM) measures the periodicity over different spectral bands. We calculated the mean of fluctuation peak (FMm). Spectral flux (SF) computes the distance between the spectrum of each successive frame. We computed the mean of spectral flux (SFm). Key mode is a measure of modality (M), that is, major or minor. The closer a value is to +1, the more major the excerpt is predicted to be. We calculated the mean of modality (Mm). Pitch (P) estimates pitches based on the centroid and deviation of the unwrapped chromagram. We calculated the mean of pitch (Pm). Therefore, we

Figure 2

Proportion of AI-Ratings for Each Musical Excerpt in Study 1



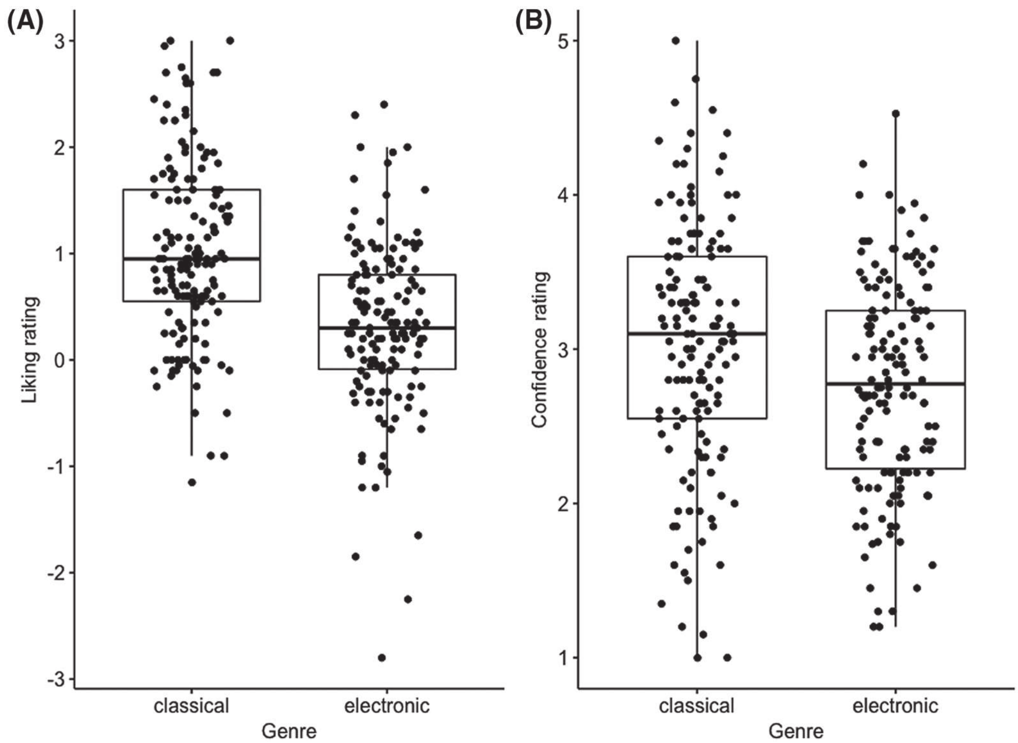
Note. (A) Electronic excerpts and (B) classical excerpts. Each musical excerpt is depicted by a separate bar. The *y*-axis depicts the proportion of AI-ratings and the excerpts are sorted based on this proportion. Excerpts were categorized as AI-sounding (blue) or human-sounding (red) if over 60% of participants rated them as such. AI = artificial intelligence. See the online article for the color version of this figure.

extracted nine features from each piece of music which can be grouped according to the musical properties they represent: dynamics (RMSl, RMSd), rhythm (Td, Tm, PCd, FMm), timbre (SFm), and tonality (Mm, Pm).

We extracted each of these nine features from the 40 musical clips used in Study 1 (20 electronic, 20 classical) using the MIRtoolbox in MATLAB (Lartillot et al., 2008). As in prior work (Eerola, 2011), for high level features (pulse clarity and mode), we used a 2s frame with an overlap of 50%. For low level features (all other features), we used a frame of 42 ms and a 50% overlap between frames. The results from the frames were then summarized either using the mean, the standard deviation, or the slope function as described above (selected based on prior research). Features were extracted for each of the 40 pieces used in Study 1. Therefore, each piece has a single value for each feature. We then sought to investigate whether these nine features could predict the proportion of AI versus human-sounding ratings from Study 1. We combined both electronic and classical clips into a single model to identify musical features that predicted AI or human ratings across genres. We then conducted a binomial regression using the nine features as predictors and the proportion of AI ratings as the outcome variable. The overall model was significant $\chi^2(9) = 155.14$, $p < .001$. The following features were significantly predictive of the proportion of AI ratings: RMSd, Tm, PCd, and Mm. See Table 1 for the full results of the regression and Figure 4 for a graphical depiction of the results.

To interpret what these significant predictors may mean in terms of what types of music are perceived as more AI or human-like, we can consider what the features represent. First, RMS(d) positively predicted the proportion of AI ratings. This suggests that music that sounds like it was composed by an AI tends to have a wider range (i.e., standard deviation) of RMS energy, or greater dynamic fluctuations. Next, tempo(m) negatively predicted the proportion of AI ratings. This suggests that pieces with slower average tempos

Figure 3
Comparison of Ratings Between Classical and Electronic Groups for Study 1



Note. (A) Liking ratings and (B) confidence ratings. Individual points indicate individual participants.

are more likely to be perceived as composed by an AI. Pulse clarity(d) also negatively predicted the proportion of AI ratings. This suggests that pieces with stronger pulse clarity are more likely to be perceived as composed by a human. Finally, mode(m) positively predicted the proportion of AI ratings. This suggests that pieces in a major mode are more likely to be perceived as composed by an AI.

Study 1: Discussion

Across both electronic and classical music, we found support for the AI composer bias. Although overall, participants liked classical music more than electronic (replicating our prior work; Belfi et al., 2018) and were more willing to attribute an AI composer to electronic music, there was still a strong relationship between perceived composer and liking for both genres. That is, music that was perceived as being composed by an AI was liked less than music that was perceived as being composed by a human. Additionally, our analysis of musical features suggests that the perception of an AI versus a human composer may be related to acoustic features of the music itself: Music that participants perceived as having an AI composer tended to have different musical features than music that was perceived as having a human composer. However, since participants rated both liking and perceived composer of the music, we cannot disentangle whether the perceived AI composer led to less liking, or whether less liked music was believed to be created by an AI, or both. To address this, rather than having participants rate the perceived composer identity, Study 2 directly manipulated composer identity to investigate the influence of purported composer identity on aesthetic judgments of music.

Table 1
Regression Results From Musical Feature Analysis Predicting Proportion of AI Ratings

Feature	β	*SE*	*z*	*p*	Sig.
Intercept	−0.13	0.03	−4.86	<.001	***
Root-mean-square (*d*)	0.28	0.06	4.42	<.001	***
Root-mean-square (*l*)	0.04	0.03	1.20	.23	
Tempo (*d*)	−0.04	0.05	−0.87	.38	
Tempo (*m*)	−0.10	0.03	−3.34	<.001	***
Pulse clarity (*d*)	−0.20	0.03	−5.97	<.001	***
Fluctuation magnitude (*m*)	0.04	0.06	0.77	.44	
Spectral flux (*m*)	−0.01	0.06	−0.22	.83	
Mode (*m*)	0.12	0.03	4.02	<.001	***
Pitch (*m*)	−0.05	0.03	−1.41	.16	

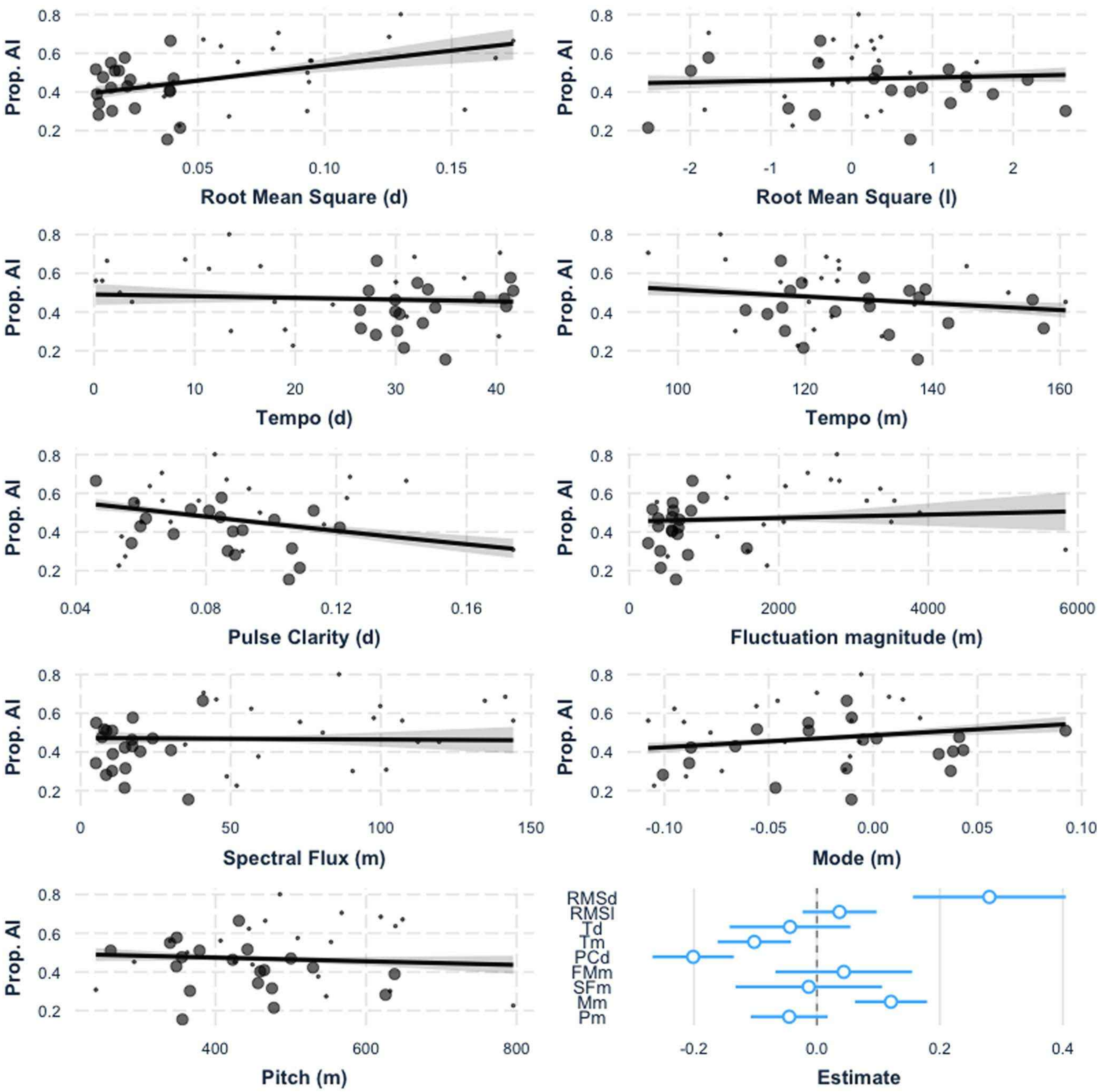
Note. *d* = standard deviation; *l* = slope; *m* = mean; *SE* = standard error. *** $p < .001$.

Study 2

In Study 2, we sought to investigate whether the *sound* of the music (i.e., does it sound like it was composed by an AI or a human) or the purported *identity* of the composer influences aesthetic judgments of music, or whether there is an interaction between the two. In Study 1, we identified music that was rated as more “AI-sounding” or “human-sounding,” but only for electronic music (i.e., there was only one “AI-sounding” classical excerpt). In Study 2,

Figure 4

Regression Results for Musical Features Predicting Proportion of AI Ratings



Note. Each panel depicts the regression results for each feature. *Y*-axis indicates the proportion of AI ratings. Large circles indicate classical pieces, small circles indicate electronic pieces. Bottom right panel indicates β estimates and 95% confidence intervals for the estimates for each predictor. RMSd = root-mean-square energy, standard deviation; RMSl = root-mean-square energy, slope; Td = tempo, standard deviation; Tm = tempo, mean; PCd = pulse clarity, standard deviation; FMm = fluctuation peak magnitude, mean; SFm = spectral flux, mean; Mm = mode, mean; Pm = pitch, mean; AI = artificial intelligence. See the online article for the color version of this figure.

we selected the four most "AI-sounding" and four most "human-sounding" electronic clips as stimuli (see Figure 2A). Participants were either told that the pieces were composed by an AI or by a human. Thus, we sought to assess whether knowing the identity of the composer has an impact on liking for music, and whether this identity information interacts with the way the music sounds (human-sounding or AI-sounding). In Study 2a, composer identity was a between-subjects manipulation: participants were randomly assigned to one of three groups and were told that *all* of the musical clips were either composed by (a) an AI, (b) a human, or (c) were not given information on composer at all. In Study 2b, composer identity was a within-subjects manipulation where all participants were told that half the pieces were composed by an AI and half the pieces were composed by a human.

Study 2a

Study 2a: Method

Participants. As in Study 1, 150 participants were recruited per group (with one additional participant who completed the study), which resulted in a total of 451 participants. Participants were recruited online using Prolific. Of these, 52 participants were excluded for the following reasons: failing the attention check ($n = 32$), correctly identifying one of the musical pieces ($n = 17$) or producing nonsensical responses ($n = 3$). This left a total of 399 participants in the following groups: AI ($n = 142$; 57 men, 83, women, 2 nonbinary), Human ($n = 130$; 63 men, 64 women, 3 nonbinary), Control ($n = 127$; 62 men, 62 women, 3 nonbinary).

Participants in the AI group were an average of 35.5 years old ($SD = 12.2$), had an average of 15.8 years of education ($SD = 2.29$), and had an average of 2.18 years of musical training ($SD = 3.54$). Participants in the Human group were an average of 34.0 years old ($SD = 12.0$), had an average of 15.2 years of education ($SD = 2.38$), and had an average of 2.12 years of musical training ($SD = 3.20$). Participants in the Control group were an average of 35.9 years old ($SD = 12.2$), had an average of 15.3 years of education ($SD = 2.56$), and had an average of 2.60 years of musical training ($SD = 4.71$). The three groups did not significantly differ on any of the demographic variables (all ps $> .10$). Overall, participants reported relatively little familiarity with AI-composed music (based on the percentage of participants rating a 3 on the AI familiarity scale: AI group = 11%, human group = 8%, control group = 11%), with no differences between the groups ($\chi^2 = 0.87$, $p = .64$).

Stimuli. We selected eight 15s musical excerpts based on the average ratings from Study 1: the four electronic pieces that had the highest proportion of AI ratings (“AI-sounding” pieces) and the four electronic pieces that had the highest proportion of human ratings (“human-sounding” pieces). See Table S1, for the titles and composers in the Supplemental Materials.

Procedure. Ethical compliance, requirement for headphones, general instructions, and participant compensation were identical to Study 1. In Study 2a, however, participants were instructed that they would hear eight musical clips that had either been composed by an AI (AI group), by a human composer (Human group), or composer identity was not mentioned (Control group). Exact instructions were as follows:

> In this survey you will hear eight musical clips [AI: created by an artificial intelligence (AI) composing software/ Human: created by various composers/ Control: no additional text]. After each clip, you will be asked a few questions about your thoughts about the music. Please listen to the entire clip and answer the questions as accurately as possible.

Participants in the AI group were also presented with the same definition of AI as in Study 1. Participants heard each of the musical excerpts presented in a random order. As in Study 1, each musical excerpt was presented individually, and participants were able to replay the excerpt if desired. Participants were unable to move on to the next page until the entire 15s excerpt had finished playing once.

Measures. Participants rated each musical excerpt on four dimensions: liking, musical quality, arousal, and valence. To assess how much a participant liked the music, they were asked “How much did you like the music you just heard?” and responded on a 7-point scale ranging from *dislike a great deal* (coded −3) to *like a great deal* (coded 3). Musical quality was measured by asking “How would you rate the quality of this piece?” with ratings on a 5-point scale ranging from *very poor* (coded 1) to *excellent* (coded 5). Emotional arousal was assessed by asking “How relaxing or stimulating did you find this piece?” and providing ratings on 7-point scale ranging from *very relaxing* (coded −3) to *very stimulating* (coded 3). Lastly, to measure emotional valence participants were asked “How positive or negative are the emotions this piece evokes for you?” measured on a 7-point scale with responses from *very negative* (coded −3) to *very positive* (coded 3).

At the end of the experiment, participants were asked three questions about the musical pieces *overall*. First, they were asked “How purposeful did you perceive these clips as a whole?” with a 4-point Likert scale ranging from *not purposeful* (coded 1) to *very purposeful* (coded 4). Next, they were asked “How much effort do you think the composer put into creating these clips?” with a 4-point Likert scale ranging from *no effort* (coded 1) to *a lot of effort* (coded 4). Finally, they were asked to “Please rate your perception of the authenticity of these clips as a whole” with a 4-point Likert scale ranging from *not authentic at all* (coded 1) to *very authentic* (coded 4). Participants did not see the numerical ratings on any of the scales.

Finally, an attention check question inquiring about the identity of the composer was included to ensure that participants were attentive during the duration of the survey. Participants were asked “Who was the composer of the music you just heard?” and were given the choices: “an AI,” “various composers,” or “not specified.” A brief demographic questionnaire was included prior to debriefing, which asked participants to report their age, gender, years of education, years of formal musical training, and familiarity with AI-composed music. Last, participants were asked if they recognized any of the songs played and if so to list which songs and artists they recognized.

Study 2a: Results

To investigate whether the sound of the music itself or the identity of the composer influenced listeners’ judgments of music, we conducted a series of linear mixed effect models, one for each of the four dependent variables. For these analyses, “composer sound” and “composer identity” were included as fixed effects (as well as the interaction between composer sound and composer identity), with random intercepts for participants and stimuli. As a categorical variable with three levels, the variable “composer identity” was set using treatment contrasts which were dummy coded using the Control group as the reference (to conduct all pairwise comparisons between the three groups, analyses were repeated with the Human group as reference, which did not produce any significant results). The variable “composer sound” was coded as AI = −0.5, human = 0.5. The four dependent variables were the four ratings made after each piece: liking, quality, arousal, and valence.

These models revealed main effects of composer sound on liking, quality, and valence; in all cases, the human-sounding pieces were rated significantly higher (more liked, higher quality, and more positively valenced) than the AI-sounding pieces. There was no main effect of composer sound on arousal. There were no main effects of composer identity in any of our analyses. The only significant interaction between composer sound and composer identity was for liking ratings: the difference in liking ratings between human-sounding and AI-sounding pieces was greater in the AI composer identity group than the control composer identity group. See Table 2 for full statistical results and Figure 5 for a graphical depiction of the data. To investigate potential effects of musical training, these analyses were repeated with years of musical training included as a fixed effect, and there were no effects of musical training (see Supplemental Materials, for details).

We also investigated whether composer identity (AI, human, or no composer) influenced participants' overall ratings of the entire collection of musical clips. We conducted three one-way analyses of variance (ANOVAs) to look for composer identity differences in the overall ratings of purpose, effort, and authenticity. There was no difference in the three conditions in terms of ratings of purposefulness of the composers, $F(2, 396) = 1.80$, $p = .16$, $\eta_p^2 = 0.009$.

For ratings of effort, there was a significant difference between the groups, $F(2, 396) = 17.85$, $p < .001$, $\eta_p^2 = 0.08$; Figure 6A. Post hoc pairwise comparisons, Bonferroni corrected for multiple comparisons, indicated that participants in the AI composer group rated the pieces as requiring significantly less effort than the human group, $t(257) = -4.94$, $p < .001$, and control group, $t(262) = -4.98$, $p < .001$. There were no differences in effort ratings between the human and control groups, $t(255) = -0.12$, $p = .99$. For authenticity ratings, there was a significant difference between groups, $F(2, 396) = 4.28$, $p = .01$, $\eta_p^2 = 0.02$; Figure 6B. Post hoc pairwise comparisons, Bonferroni corrected for multiple comparisons, indicated that participants in the AI composer group rated the music as significantly less authentic than the human group music, $t(263) = -2.65$, $p = .02$. Yet, there were no differences in the authenticity ratings of the music between the AI and control groups, $t(267) = -2.21$, $p = .08$, or the human and control groups, $t(246) = -0.68$, $p = .99$.

Table 2
Results of Mixed Effects Models for Study 2a

Fixed effects	β	SE	t	p	Sig.
Liking					
Composer sound	0.79	0.20	3.96	<.001	***
Identity—AI vs. control	−0.15	0.12	−1.26	.21	
Identity—Human vs. control	−0.12	0.12	−1.00	.32	
Sound × Identity—AI	0.25	0.11	2.39	.02	*
Sound × Identity—Human	0.12	0.11	1.12	.26	
Quality					
Composer sound	0.47	0.10	4.69	<.001	***
Identity—AI vs. control	−0.14	0.07	−1.94	.12	
Identity—Human vs. control	−0.05	0.07	−0.74	.57	
Sound × Identity—AI	0.08	0.06	1.33	.18	
Sound × Identity—Human	0.03	0.06	0.55	.59	
Arousal					
Composer sound	0.06	0.57	0.11	.91	
Identity—AI vs. control	−0.08	0.08	−1.05	.29	
Identity—Human vs. control	−0.07	0.08	−0.81	.42	
Sound × Identity—AI	0.10	0.09	1.17	.24	
Sound × Identity—Human	0.07	0.09	0.80	.42	
Valence					
Composer sound	0.71	0.19	3.64	.01	*
Identity—AI vs. control	−0.10	0.11	−0.90	.37	
Identity—Human vs. control	0.02	0.11	0.17	.87	
Sound × Identity—AI	0.13	0.10	1.31	.19	
Sound × Identity—Human	0.01	0.10	0.06	.95	

Note. *SE* = standard error; AI = artificial intelligence.
* $p < .05$. *** $p < .001$.

Study 2b

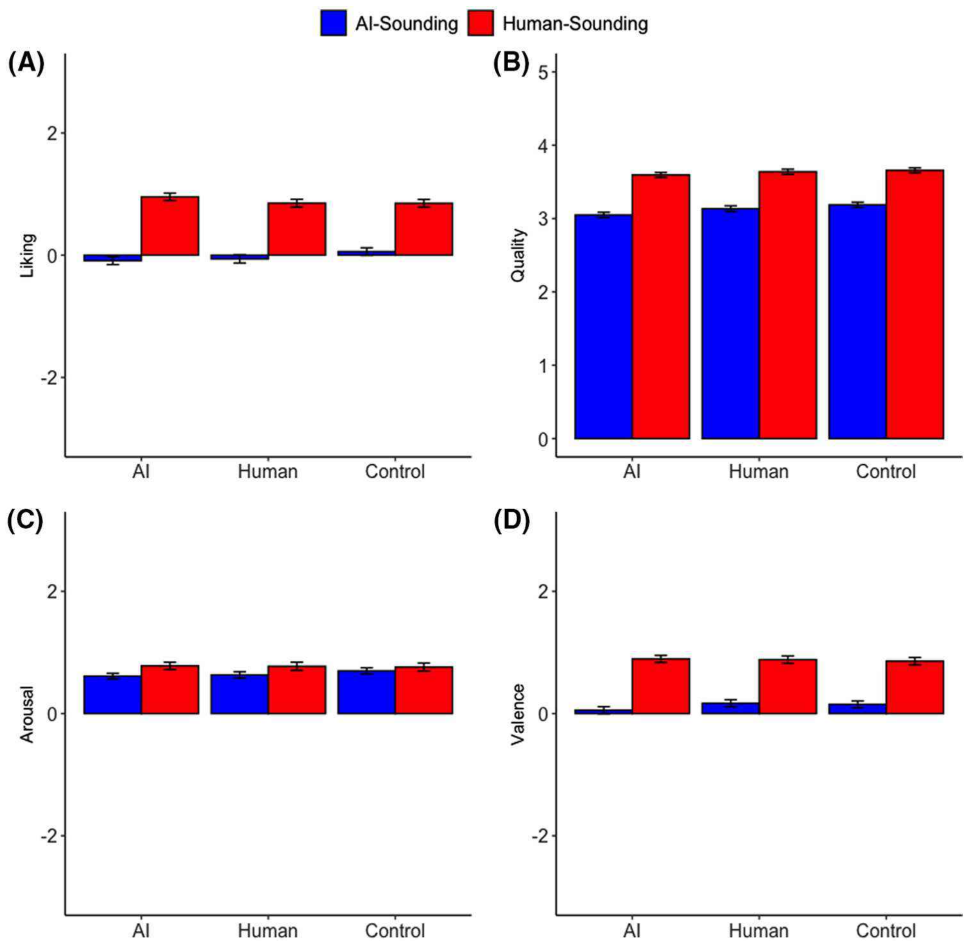
In Study 2a, we found a significant effect of composer sound (i.e., AI- vs. human-sounding), but no effect of composer identity. We suspect that this lack of composer identity effect may have been due to the nature of our between-subjects manipulation. That is, it is possible that a composer identity effect would only be seen if participants were to make *direct* comparisons between music purportedly composed by an AI versus a human. Therefore, in Study 2b, we conducted a near-identical study to that of Study 2a but used a within-subjects manipulation for composer identity. Unlike Study 2a, there were no "control" trials where the composer was unspecified. We also strengthened the identity manipulation, by stating that each clip was composed by a specifically named human or AI. Therefore, Study 2b was a two (composer identity: AI vs. human) by two (composer sound: AI-sounding vs. human-sounding) within-subjects design.

Study 2b: Method

Participants. As in Study 2a, participants were recruited using Prolific. As this was a within-subjects manipulation, 150 participants were recruited and completed the study. Of those, 14 participants were excluded: two for correctly recognizing the composer of one of the pieces of music, and 12 for failing the attention check. This resulted in a final total of 136 participants in Study 2b. Participants (80 men, 53 women, 2 nonbinary, 1 did not disclose gender) were an average of 37.04 years old (*SD* = 12.59), had 15.50 years of education (*SD* = 2.66) and 2.22 years of formal musical training (*SD* = 4.16). Overall, participants had relatively little prior familiarity with AI-composed music (9% of participants reported familiarity with AI-composed music).

Procedure. Ethical approval, consent, stimuli, measures, and general instructions were identical to Study 2a except when mentioned. Participants were instructed that they would hear eight musical clips, some of which had been composed by an AI or by a human composer. Participants heard each of the eight 15s excerpts (four AI-sounding and four human-sounding, the same as Study 2a) presented in a random order. Participants were told that the composer of four of the pieces was a human composer and for the other four pieces was an AI composer. Pieces were balanced such that participants heard half AI-sounding pieces with a purported AI composer, half AI-sounding pieces with a purported human composer, half human-sounding pieces with a purported human composer, and half human-sounding pieces with a purported AI composer. Within this balancing, specific pieces were randomly assigned to a purported composer type. AI composer names were as follows: TuneSoft, MelodyBot, RoboBeats, SoundGen, IntelligentMedia, FutureBeats, BotStudio, GenSoft. Human composer names were as follows: Victoria Moore, Christopher Thompson, Anna Jones, Jason Miller, Morgan Walker, Jack Martin, Vanessa Johnson, and Matthew Hall. Composer identity was counterbalanced across

Figure 5
Results of Study 2a



Note. (A) Liking ratings, (B) quality ratings, (C) arousal ratings, and (D) valence ratings. Error bars indicate standard error of the mean. Composer sound is represented by color, composer identity is depicted on the *x*-axis. AI = artificial intelligence. See the online article for the color version of this figure.

participants such that the same pieces were not always given the same composer identity.

Since participants saw a mix of AI and human composers, they were not asked to make any "overall" ratings of the pieces as a whole as they did in Study 2a. The attention check question was similar to Study 2a (i.e., Who was the composer of the pieces you heard?) but the choices were modified: all humans, all AIs, some humans and some AIs, and the composers were not listed.

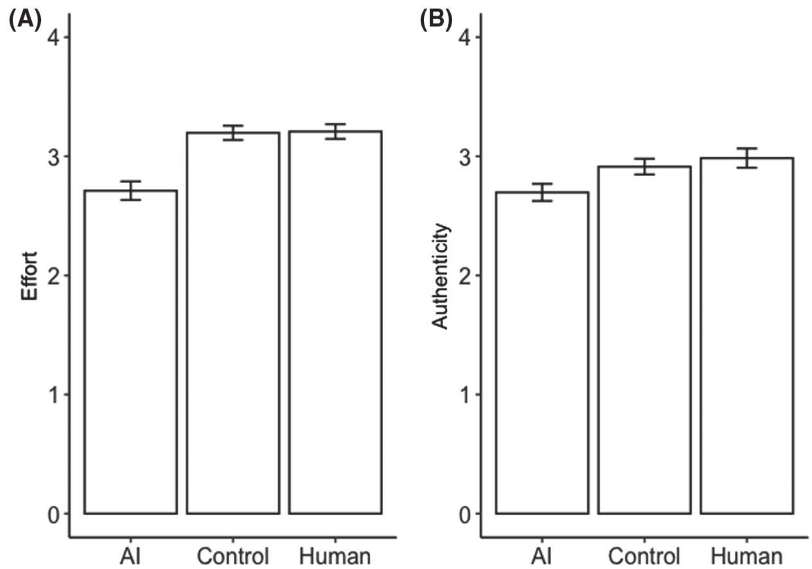
Study 2b: Results

To investigate whether the sound of the music itself or the identity of the composer influenced listeners' judgments of music, we conducted a series of linear mixed effects models, one for each of the four dependent variables. For these analyses, fixed effects included "composer sound" and "composer identity," as well as their interaction; random intercepts for participants and items were also included. As categorial predictors, composer sound was coded as: AI = −0.5, human = 0.5 and composer identity was coded as: AI = −0.5, human = 0.5. The four dependent variables were the four ratings made after each piece: liking, quality, arousal, and valence. The four models revealed no significant effects of composer sound, composer identity, or interactions between the two (for the full results, see Table 3). To investigate potential effects of musical training, these analyses were repeated with years of musical training included as a fixed effect, and there were no effects of musical training (see Supplemental Materials, for details).

Study 2: Discussion

In both Studies 2a and 2b, we aimed to further investigate whether composer identity influences aesthetic judgments of music. In Study 2a, when we manipulated composer identity in a between-subjects manner, we found a significant effect of composer sound. That is, human-sounding excerpts were liked more than AI-sounding excerpts. However, when we manipulated composer identity in a within-subjects manner, in Study 2b, we did not see this effect. The only difference in Studies 2a and 2b was the way

Figure 6
Overall Rating Results From Study 2a



Note. (A) Overall effort ratings and (B) overall authenticity ratings. The *x*-axis is the composer identity conditions. Error bars indicate standard error of the mean. AI = artificial intelligence.

this composer identity manipulation was conducted. Therefore, it could be that when confronted with both AI and human composers (as in Study 2b), participants focused less on the music sound *itself* and more on the composer identity, which is why there was no effect of composer sound in this study—that is, composer sound and composer identity may have directly influenced each other in ways that "canceled out" the effects seen in Study 2a. Additionally, it may have been the case that our choice of electronic music influenced the results of both Studies 2a and 2b: Listeners may be less biased toward or against AI composers when the music fits their expectations of what an AI can create (i.e., that electronic music has computerized elements and therefore could plausibly be composed by an AI). Similarly, electronic music was not liked as much as classical, therefore potentially making an effect of composer identity on liking harder to detect. Therefore, in Study 3, we sought to determine whether composer identity has an effect on aesthetic judgments of music when all pieces are homogenous in terms of their composer sound (i.e., human sounding) and also well liked.

Table 3
Results of Linear Mixed Effects Models for Study 2b

Fixed effects	β	*SE*	*t*	*p*
Liking				
Composer sound	−0.06	0.12	−0.45	.65
Composer identity	−0.13	0.12	−1.05	.29
Sound × Identity	0.21	0.17	1.21	.23
Quality				
Composer sound	−0.06	0.07	−0.74	.47
Composer identity	−0.12	0.07	−1.84	.07
Sound × Identity	0.15	0.10	1.58	.12
Arousal				
Composer sound	−0.07	0.10	−0.66	.51
Composer identity	0.07	0.10	0.73	.46
Sound × Identity	0.07	0.14	0.47	.64
Valence				
Composer sound	−0.06	0.11	−0.56	.58
Composer identity	−0.02	0.11	−0.19	9.85
Sound × Identity	0.14	0.16	0.89	.38

Note. *SE* = standard error.

Study 3

To examine whether the AI composer bias exists when musical pieces are homogenous in terms of their composer sound (i.e., all human-sounding) and overall liking (i.e., generally liked), we conduced Study 3. Like Study 2, we selected eight clips, but this time from the classical clips in Study 1. We selected the eight clips that had the highest proportion of human-sounding ratings from Study 1 (see Figure 2B), which were also highly liked pieces. Therefore, the pieces in Study 3 differed from those in Study 2 in that they were both stylistically and aesthetically homogenous—that is, all were well-liked human-sounding classical excerpts. Also as in Study 2, we conducted two versions of Study 3—Study 3a manipulated composer identity in a between-subjects manner, while Study 3b manipulated composer identity in a within-subjects manner.

Study 3a

Study 3a: Method

Participants. As in Study 2a, 150 participants were recruited per group (with one additional participant who completed the study), which resulted in a total of 451 participants. Participants were recruited online using Prolific. Of the 451 participants, 58 were excluded for failing the attention check. This left a total of

393 participants in the following groups: AI (n = 144; 58 men, 82, women, 4 nonbinary), Human (n = 128; 54 men, 71 women, 3 nonbinary), Control (n = 121; 56 men, 64 women, 1 nonbinary).

Participants in the AI group were an average of 34.8 years old (SD = 12.3), had an average of 14.8 years of education (SD = 2.31), and had an average of 2.80 years of musical training (SD = 4.86). Participants in the Human group were an average of 37.6 years old (SD = 13.8), had an average of 15.0 years of education (SD = 2.01), and had an average of 2.00 years of musical training (SD = 3.51). Participants in the Control group were an average of 36.9 years old (SD = 11.9), had an average of 15.6 years of education (SD = 2.41), and had an average of 2.28 years of musical training (SD = 3.59). The three groups did not significantly differ in age (p = .15) or years of musical training (p = .25), but the control group did have more years of education than both the human group (p = .04) and the AI group (p = .005). Overall, participants reported relatively little familiarity with AI-composed music (based on the percentage of participants rating a 3 on the AI familiarity scale: AI group = 5%, human group = 8%, control group = 9%), with no differences between the groups (χ^2 = 1.29, p = .52).

Stimuli. Stimuli were selected from the classical stimulus set in Study 1. Eight classical stimuli were chosen that had the highest proportion of “Human” ratings in Study 1. See Supplemental Table S2, for titles and composers of the stimuli.

Procedure. Ethical compliance, requirement for headphones, general instructions, measures, and all procedures were identical to those used in Study 2a. The only aspect that differed between Study 2a and Study 3a were the stimuli.

Study 3a: Results

To investigate whether the identity of the composer influenced listeners’ judgments of music, we conducted a series of linear mixed-effect models, one for each of the four dependent variables. For these analyses, “composer identity” was included as a fixed effect with random intercepts for participants and stimuli. As a categorical variable with three levels, the variable “composer identity” was set using treatment contrasts which were dummy coded using the Control group as the reference (to conduct all pairwise comparisons between the three groups, analyses were repeated with the Human group as reference). In contrast to Study 2, there was no “composer sound” variable as all stimuli were human-sounding. The four dependent variables were the four ratings made after each piece: liking, quality, arousal, and valence. To investigate potential effects of musical training, we also included musical training as a fixed effect. Since there were significant effects of musical training, we have included these results here.

There was a significant effect of composer identity on *quality* ratings: Participants in the AI group rated the music as being significantly lower quality than participants in both the Human group and the Control group. There was also a significant effect of composer identity on *arousal* ratings, such that participants in the Human group rated the pieces as more arousing than participants in the Control group. Additionally, we saw significant effects of musical training on liking and valence ratings, such that participants with more musical training liked the pieces more and rated them as more positively valenced. See Table 4 for the full results of the models.

Table 4
Results of Mixed Effects Models for Study 3a

Fixed effects	β	SE	t	p	Sig.
Liking					
AI vs. control	−0.02	0.08	−0.27	.79	
Human vs. control	0.10	0.08	1.31	.19	
AI vs. human	−0.12	0.08	−1.63	.10	
Musical training	0.09	0.03	2.80	.005	**
Quality					
AI vs. control	−0.14	0.07	−2.07	.03	*
Human vs. control	0.03	0.07	0.41	.68	
AI vs. human	−0.16	0.07	−2.39	.01	*
Musical training	0.05	0.03	1.73	.08	
Arousal					
AI vs. control	0.12	0.09	1.30	.19	
Human vs. control	0.25	0.09	2.66	.01	*
AI vs. human	−0.13	0.09	−1.45	.15	
Musical training	0.08	0.04	2.06	.04	*
Valence					
AI vs. control	0.05	0.10	0.51	.61	
Human vs. control	0.06	0.10	0.57	.57	
AI vs. human	−0.01	0.10	−0.07	.94	
Musical training	0.08	0.04	2.01	.04	*

Note. SE = standard error; AI = artificial intelligence.
$^{*}p < .05$. $^{**}p < .01$.

Study 3b

Study 3b: Method

Participants. As in Study 2b, 150 participants were recruited and completed the study using Prolific. Of those, 14 were excluded for failing the attention check, leaving a total of 136 participants (75 men, 60 women, 1 genderfluid). Participants were an average of 33.63 years old (SD = 11.61), had 15.16 years of education (SD = 2.33), and 2.47 years of musical training (SD = 4.06). Overall, participants had relatively little prior familiarity with AI-composed music (9% of participants reported familiarity with AI-composed music).

Stimuli. Stimuli were the same as those used in Study 3a.

Procedure. Ethical compliance, requirement for headphones, general instructions, measures, and all procedures were identical to those used in Study 2b. The only aspect that differed between Study 2b and Study 3b were the stimuli.

Study 3b: Results

To investigate whether the identity of the composer influenced listeners’ judgments of music, we conducted a series of linear mixed effects models, one for each of the four dependent variables. For these analyses, “composer identity” was a fixed effect, with random effects for participants and stimuli. Composer identity as a categorical predictor was coded as AI = −0.5, human = 0.5. In contrast to Study 2, there was no “composer sound” variable as all stimuli were human-sounding. The four dependent variables were the four ratings made after each piece: liking, quality, arousal, and valence.

There was a significant effect of composer identity on *liking* ratings, such that the pieces with a purported human composer were rated as more liked than the pieces with a purported AI composer (β = 0.15, SE = 0.06, t = 2.56, p = .009). There was also a significant effect of composer identity on *quality* ratings, such that the pieces with a purported human composer were rated as higher quality than the pieces with a purported AI composer ($\beta = 0.15$, $SE = 0.03$, $t = 4.46$, $p < .001$). There was not a significant effect of composer identity on the arousal of the pieces ($\beta = -0.05$, $SE = 0.08$, $t = -0.64$, $p = .52$) or the valence of the pieces ($\beta = 0.04$, $SE = 0.06$, $t = 0.64$, $p = .52$). To investigate potential effects of musical training, these analyses were repeated with years of musical training included as a fixed effect, and there were no effects of musical training (see Supplemental Materials, for details).

Study 3: Discussion

For this well-liked, human-sounding set of classical musical excerpts, we identified an AI composer bias. In our between-subjects manipulation of composer identity (Study 3a), participants in the AI group rated the music as significantly lower quality than participants in the Human and Control groups. In our within-subjects manipulation of composer identity (Study 3b) participants rated the music as both lower quality and liked it less if it were purportedly composed by an AI. These results replicate our findings from Study 1 and extend them to show that participants like music less and find it lower quality when the purported composer is an AI, whether or not they are making direct comparisons to music composed by a purported human. These results provide evidence to support our hypothesis of an AI composer bias for classical, generally well liked, and human-sounding music, such that the identity of the composer has an impact on the degree to which listeners enjoy a piece of music.

General Discussion

In the present series of experiments, we found support for the AI composer bias in certain situations: People liked music less when they believed it was composed by an AI. Specifically, for both classical and electronic music, pieces that were liked less were also judged as more likely to be composed by an AI (Study 1). When we directly manipulated composer identity, the AI composer label decreased how much people liked and perceived the quality of the music (Study 3). However, we did not find evidence for the AI composer bias when using a set of stimuli that varied in the degree to which the stimuli sounded AI-like or human-like (Study 2). That is, when we found support for the AI composer bias (in Study 3), we used stimuli that were homogenous both in terms of their style (i.e., all human-sounding classical pieces) and in their aesthetic appeal (i.e., all well-liked pieces).

We believe that there are two nonexclusive potential reasons for this lack of support for our hypothesis in Study 2. First, it may be that people are only biased against AI composers for music that is not congruent with their expectations of what AIs can produce (i.e., if the music sounds “human like”), and the bias is nullified or even in the opposite direction for more congruent pieces (i.e., AI-sounding). This is consistent with the idea that the bias in liking is due to a mismatch in expectations, not a general bias against all music composed by an AI. In the case of Study 2, participants heard electronic music, much of which was rated as AI-sounding. Therefore, listeners may be less biased against AI composers when the musical genre contains computerized elements. This aligns with our prior research, which indicates that listeners like music more when it is congruent with their expectations of the musical artist/performer (Belfi et al., 2021).

Additionally, similar work found that there was no bias against a known AI composer when the music heard was *actually composed* by an AI (Tigre Moura & Maw, 2021). That is, this prior study asked participants to rate several attitudes towards musical pieces labeled as human- or AI-composed, but those pieces were actually composed by AIs (vs. our approach here, which was to use all human-composed music). Tigre Moura and Maw *did not* find any differences in attitudes between the pieces labeled as AI-created versus human-created, similar to what we found in Study 2. Similar work has also used atonal (dodecaphonic) music and labeled it as either human- or computer-generated. In this case, researchers also found no difference in the affective ratings of music purportedly generated by a human and a computer (Steinbeis & Koelsch, 2009). Atonal and computerized music may be more likely to fit one’s prior expectations as to what types of music could presumably be composed by an AI. Therefore, it could be the case that, when presented with electronic music (in Study 2), listeners found this music congruent with their expectations of an AI composer and therefore did not show the AI composer bias.

Second, it may be that a bias against AI composers only occurs for music that is highly liked (and potentially highly disliked). The electronic clips used in Study 2 were not highly liked or disliked, as liking ratings were, on average, near the midpoint of the scale. It is easy to imagine that participants did not look to composer identity as a source to attribute their enjoyment, when in fact they were not experiencing much enjoyment. That is, if a listener has neutral feelings towards a piece of music, the identity of the composer (or any other contextual factors) may not influence their feelings about it. In many cases, especially those dealing with aesthetics or human domains, judgments of AIs are similar to those of humans, only weaker (Jago, 2019; Longoni & Cian, 2020; Shank, 2014) potentially due to AIs being perceived as liminal minds without the full range of agency and experience (Gamez et al., 2020; Gray et al., 2007; Shank et al., 2021). For AI’s aesthetic compositions specifically, research has found that people have different schemas related to whether an AI can actually create true “art” (Hong, 2018), and this significantly alters their evaluation of the composition (Hong & Curran, 2019). Future research could examine the effect of composer identity on music of a wider variety of genres, or, for example, identify whether the composer identity decreases liking for an individual’s favorite genre.

Both explanations, however, point to the same general conclusion: Enjoyable music is perceived as human-sounding (i.e., composed by a human) and a purported AI composer makes music less enjoyable. This has important real-world implications: If people like music less simply because it has been composed by an AI, this suggests that the public will be less likely to readily accept music created by AI and enjoy that music less when they hear it. It also suggests that knowing the identity of the composer ahead of time sets this bias in place. That would mean that hearing the music without knowing the composer combats this bias in the same way that blind orchestra auditions allow judges to fully appreciate the music quality without regards to the gender of the musician (Goldin & Rouse, 2000). Yet, there may be systematic stereotypes—at least for the present time—on what types of music are congruent with AI composers. As shown here, it seems likely that listeners will more readily adopt an AI composer for more “AI-sounding” music, that is, electronic or computerized music. This would suggest that as AI-composed music becomes more acceptable and the genres and styles of AI-composed music becomes more diverse, the stereotype and bias against AIs may diminish.

In addition to differences in aesthetic judgments of music based on composer identity, we also found that AI-sounding pieces differed from human-sounding pieces in terms of their musical features. Prior work also found that liked and disliked musical excerpts significantly differ in their musical features, in terms of pitch, articulation, rhythm (including fluctuation peak, tempo, and pulse clarity), and timbre (including spectral flux; Brattico et al., 2016). This coincides with our finding, suggesting that certain musical features are both (a) liked more than others and (b) more likely to be associated with a perceived AI-composer. Other work has looked more specifically at the musical features present in electronic music: RMS energy was found to be critical to the "break" routine (a sudden large decrease and increase in a track's intensity) and was associated with emotional responses when listening to electronic music (Solberg & Dibben, 2019; Solberg & Jensenius, 2016). Therefore, it is not entirely surprising that we found a positive association between RMS and ratings of AI-sounding music, as the electronic music was rated as sounding more AI-composed. RMS seems to be a particularly defining feature of electronic music, which listeners are more likely to attribute as being composed by an AI.

Contributions to the Broader Fields of AI Creativity and Aesthetic Judgments

Overall, our work has important implications for the growing field of AI and the likelihood of individuals adopting new AI technologies when it comes to music. While people have readily accepted AI technology in various aspects of everyday life (e.g., smart home assistants, recommender systems), people are still hesitant to accept creative products produced by AIs, including visual arts and advertisements (Jago, 2019; Kirk et al., 2009; Wu & Wen, 2021). Our findings suggest not only that this applies to musical composition, but also that this goes beyond acceptance and affects aesthetic judgment. That is, our work suggests that individuals may be less likely to accept creative works ostensibly produced by AIs because they *like them less* than creative works produced by humans. Musical preferences, tastes, and choices are highly individual and often a central part of a person's identity (Lamont & Loveday, 2020; Lonsdale & North, 2017; Peck & Grealey, 2020). It may be that connection to the personal that makes the bias against AIs so important. While people may accept AIs and machines for many utilitarian tasks and conveniences, they may be hesitant, resistant, or—in our case—biased against AIs engaged in aesthetic creations perhaps because aesthetic tastes are considered especially meaningful, personal, and key to an individual's identity.

Additionally, our work has important implications for theories of aesthetic judgments more generally. One of the major topics of debate in empirical aesthetics surrounds the role of stimulus features and their influence on aesthetic judgments—that is, is it possible to distill down the precise features that make an object aesthetically pleasing? Are certain features universally beautiful, or are their other, nonstimulus features that contribute to aesthetic judgments? Our current work suggests that aesthetic judgments are not solely the result of stimulus features alone—here, in Study 3 we found that when given the *same piece of music*, listeners make different aesthetic judgments based on the purported identity of the composer.

Therefore, the present work supports theories of aesthetic judgments that include the importance of context and influences beyond the stimuli themselves. While more specifically focused on visual aesthetics, Leder and Nadal's (2014) model of aesthetic appreciation and aesthetic judgments highlights the importance of context in making aesthetic judgments, considering both "cultural, institutional, and physical" contexts. Similarly, Chatterjee and Vartanian (2014) propose an "aesthetic triad" model, by which aesthetic experiences are the result of interactions between knowledge-meaning, sensory-motor, and emotion-valuation systems, with contextual effects included in the knowledge-meaning system. Their model also highlights the importance of expectations when making aesthetic judgments of artwork. Our work supports and extends the claim that context is a critical component of aesthetic judgments: Here, we show that contextual information about the composer of a piece of music can influence aesthetic judgments of that music. Focusing specifically on aesthetic experience of music, Brattico et al. (2013) differentiate between external context (the situation) and internal context (the listener) as influences on aesthetic judgments. In particular, they focus on the potential role of attitudes in aesthetic judgments of music. Here, listeners may have biased attitudes against AIs as creative agents, which may influence their aesthetic judgments of music purportedly created by AIs.

More recently, the Vienna integrated model of top-down and bottom-up processes in art perception (Pelowski et al., 2017) integrates both top-down (i.e., context, extramusical features) and bottom-up (i.e., stimulus features) in aesthetic judgments. One particularly relevant component of this model is the concept of "schema congruence" when evaluating works of art. This model proposes that an important component of aesthetic judgments is that viewers (or, in our case, listeners) evaluate whether the artwork fits with their prior schemas. Our work provides support for the idea that schema congruence is important for positive aesthetic judgments of an artwork. That is, listeners may have a preexisting schema for what AI-created music sounds like. When the music matches with that schema (in our case, electronic music), the aesthetic judgments are not negatively biased by the contextual information. However, when the music is not congruent with the schema (e.g., classical music), the contextual information does have a negative influence on aesthetic judgments. Overall, the present work fits within several key theories of aesthetic judgments by further highlighting the role of extramusical features, including the framing and context (Leder & Pelowski, 2021) and their influence on aesthetic judgments of music.

Limitations and Future Directions

Of course, this work is not without limitations. One important aspect to note is that the excerpts used here were 15 s long, which is most likely not representative of the way individuals listen to music in a more naturalistic setting. However, our prior work using these same musical excerpts has indicated that listeners make stable and reliable aesthetic judgments of music in as little as 750 ms (Belfi et al., 2018). Even when listening to longer (60 s) excerpts, listeners make an initial judgment about how much they like the piece that tends to be reinforced over time while listening to the duration of the piece (Belfi et al., 2018). Therefore, we feel confident that the excerpts used here were long enough to induce a stable aesthetic judgment in a listener.

Another potential limitation of the present work is our exclusive use of participants recruited through Prolific. As these participants are likely more technologically savvy and familiar with AI and algorithms in general, the results of the present study may not generalize to other populations. Another aspect of generalizability is the music itself. While here we chose two musical genres that are quite different from one another, both styles of music are part of the Western musical tradition. It is important to note that historically, the field of music cognition has often ignored non-Western music (Baker et al., 2020) and Western music of different styles. Therefore, it is important to not generalize to "music" overall from such a restricted set of musical cues as were used here. A final consideration related to the musical stimuli is that we used naturalistic excerpts of "real" music. While this approach allows better generalizability, it also reduces experimental control. Future research could attempt to create musical pieces that are either AI-sounding or human-sounding to further explore what precise musical features influence both perceived composer identity and liking.

An additional factor that is known to influence aesthetic judgments of music is familiarity: for example, familiarity with musical pieces is associated with increased pleasure while listening (van den Bosch et al., 2013). While we sought to eliminate the potential effects of familiarity here by choosing uncommon musical pieces and excluding participants who recognized any of the pieces, there may indeed be effects of familiarity on a genre level. For example, participants may have been more familiar with classical music (and therefore have a preexisting preference for it) than electronic music, or vice versa. We did find that participants overall liked the classical music more than electronic, and familiarity could be influencing this effect. Additionally, in Study 3a, we found effects of musical training on liking and valence. It could be the case that individuals with more musical training are also more familiar with classical music and that could be influencing the relationship between musical training and liking. However, we did not collect data on participants' preexisting musical preferences or listening habits. For a stronger test of the role of familiarity and prior exposure on composer attributions and aesthetic judgments, future research could collect this information to better select stimuli that match (or do not match) participants' preexisting preferences. One might predict, for example, that when confronted with unfamiliar musical styles, listeners like the music less and therefore are more likely to attribute it as created by an AI.

Conclusion

Using AI, machines can now compose music. The use of AIs to compose and create music, both alone and alongside human composers, continues to grow. The AI composer bias—which we introduced and explored in the present work—may therefore be an important issue as AI-composed music becomes more prevalent. While new AI technology might compete with human musicians, it has been an open question as to when listeners will accept and enjoy this music. Our findings here suggest that listeners may be hesitant to accept AI-composed music, in part because they are biased to like music less when told it was composed by an AI. However, this AI composer bias may be influenced by musical features and whether the music in question is seen as being congruent with creation by a machine.

References

Baker, D. J., Belfi, A., Creel, S., Grahn, J., Hannon, E., Loui, P., Margulis, E. H., Schachner, A., Schutz, M., Shanahan, D., & Vuvan, D. T. (2020). Embracing anti-racist practices in the music perception and cognition community. *Music Perception*, *38*(2), 103–105. https://doi.org/10.1525/mp.2020.38.2.103

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

Belfi, A. M. (2019). Emotional valence and vividness of imagery predict aesthetic appeal in music. *Psychomusicology: Music, Mind, and Brain*, *29*(2–3), 128–135. https://doi.org/10.1037/pmu0000232

Belfi, A. M., Kasdan, A., Rowland, J., Vessel, E. A., Starr, G. G., & Poeppel, D. (2018). Rapid timing of musical aesthetic judgments. *Journal of Experimental Psychology: General*, *147*(10), 1531–1543. https://doi.org/10.1037/xge0000474

Belfi, A. M., & Loui, P. (2020). Musical anhedonia and rewards of music listening: Current advances and a proposed model. *Annals of the New York Academy of Sciences*, *1464*(1), 99–114. https://doi.org/10.1111/nyas.14241

Belfi, A. M., Samson, D. W., Crane, J., & Schmidt, N. L. (2021). Aesthetic judgments of live and recorded music: Effects of congruence between musical artist and piece. *Frontiers in Psychology*, *12*, Article 618025. https://doi.org/10.3389/fpsyg.2021.618025

Bigman, Y., Gray, K., Waytz, A., Arnestad, M., & Wilson, D. (2020). Algorithmic discrimination causes less moral outrage than human discrimination. *PsyArXiv*. https://doi.org/10.31234/osf.io/m3nrp

Brattico, E., Bogert, B., Alluri, V., Tervaniemi, M., Eerola, T., & Jacobsen, T. (2016). It's sad but I like it: The neural dissociation between musical emotions and liking in experts and laypersons. *Frontiers in Human Neuroscience*, *9*, Article 676. https://doi.org/10.3389/fnhum.2015.00676

Brattico, E., Bogert, B., & Jacobsen, T. (2013). Toward a neural chronometry for the aesthetic experience of music. *Frontiers in Psychology*, *4*, Article 206. https://doi.org/10.3389/fpsyg.2013.00206

Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-dependent algorithm aversion. *Journal of Marketing Research*, *56*(5), 809–825. https://doi.org/10.1177/0022243719851788

Chatterjee, A., & Vartanian, O. (2014). Neuroaesthetics. *Trends in Cognitive Sciences*, *18*(7), 370–375. https://doi.org/10.1016/j.tics.2014.03.003

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, *144*(1), 114–126. https://doi.org/10.1037/xge0000033

Eerola, T. (2011). Are the emotions expressed in music genre-specific? An audio-based evaluation of datasets spanning classical, film, pop and mixed genres. *Journal of New Music Research*, *40*(4), 349–366. https://doi.org/10.1080/09298215.2011.602195

Epstein, Z., Levine, S., Rand, D. G., & Rahwan, I. (2020). Who gets credit for AI-generated art? *iScience*, *23*(9), Article 101515. https://doi.org/10.1016/j.isci.2020.101515

Fischinger, T., Kaufmann, M., & Schlotz, W. (2018). If it's Mozart, it must be good? The influence of textual information and age on musical appreciation. *Psychology of Music*, *48*(4), 579–597. https://doi.org/10.1177/0305735618812216

Furlough, C., Stokes, T., & Gillan, D. J. (2021). Attributing blame to robots: I. The influence of robot autonomy. *Human Factors*, *63*(4), 592–602. https://doi.org/10.1177/0018720819880641

Gamez, P., Shank, D. B., Arnold, C., & North, M. (2020). Artificial virtue: The machine question and perceptions of moral character in artificial moral agents. *AI & Society*, *35*(4), 795–809. https://doi.org/10.1007/s00146-020-00977-1

Goldin, C., & Rouse, C. (2000). Orchestrating impartiality: The impact of "blind" auditions on female musicians. *The American Economic Review*, *90*(4), 715–741. https://doi.org/10.1257/aer.90.4.715

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, *315*(5812), 619. https://doi.org/10.1126/science.1134475

Greenberg, D. M., Matz, S. C., Schwartz, H. A., Fricke, K. R., Greenberg, D. M., Matz, S. C., Schwartz, H. A., & Fricke, K. R. (2021). The self-congruity effect of music. *Journal of Personality and Social Psychology*, *121*(1), 137–150. https://doi.org/10.1037/pspp0000293

Hadjeres, G., Pachet, F., & Nielsen, F. (2017). *DeepBach: A steerable model for bach chorales generation* [Conference session]. 34th International Conference on Machine Learning, ICML 2017, 3, 2187–2196.

Hong, J. W. (2018). Bias in perception of art produced by artificial intelligence. In M. Kurosu (Ed.), Lecture notes in computer science: Vol. 10902. *Human–computer interaction. Interaction in context. HCI 2018* (pp. 290–303). Springer. https://doi.org/10.1007/978-3-319-91244-8_24

Hong, J. W., & Curran, N. M. (2019). Artificial intelligence, artists, and art. *ACM Transactions on Multimedia Computing Communications and Applications*, *15*(2s), 1–16. https://doi.org/10.1145/3342227

Jago, A. S. (2019). Algorithms and authenticity. *Academy of Management Discoveries*, *5*(1), 38–56. https://doi.org/10.5465/amd.2017.0002

Jones, N. & Pu, P. (2008). *User acceptance issues in music recommender systems*. EPFL Technical Report HCI-REPORT-2009-001.1-10.

Kasdan, A., & Belfi, A. M. (2020). Investigating a self-reference effect in musical aesthetics. *Experimental Results*, *9*, 1–8. https://doi.org/10.1017/exp.2020.6

Kirk, U., Skov, M., Hulme, O., Christensen, M. S., & Zeki, S. (2009). Modulation of aesthetic value by semantic context: An fMRI study. *NeuroImage*, *44*(3), 1125–1132. https://doi.org/10.1016/j.neuroimage.2008.10.009

Köbis, N., & Mossink, L. D. (2021). Artificial Intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry. *Computers in Human Behavior*, *114*, Article 106553. https://doi.org/10.1016/j.chb.2020.106553

Krause, A. E., & North, A. C. (2018). ’Tis the season: Music-playlist preferences for the seasons. *Psychology of Aesthetics, Creativity, and the Arts*, *12*(1), 89–95. https://doi.org/10.1037/aca0000104

Lamont, A., & Loveday, C. (2020). A new framework for understanding memories and preference for music. *Music & Science*, *3*, 1–14. https://doi.org/10.1177/2059204320948315

Lartillot, O., Toiviainen, P., & Eerola, T. (2008). A Matlab toolbox for music information retrieval. In C. Preisach, H. Burkhardt, L. Schmidt-Thieme, & R. Decker (Eds.), *Data analysis, machine learning and applications* (pp. 261–268). Springer.

Leder, H., & Nadal, M. (2014). Ten years of a model of aesthetic appreciation and aesthetic judgments : The aesthetic episode—Developments and challenges in empirical aesthetics. *British Journal of Psychology*, *105*(4), 443–464. https://doi.org/10.1111/bjop.12084

Leder, H., & Pelowski, M. (2021). Empirical aesthetics: Context, extra information, and framing. In M. Nadal & O. Vartanian (Eds.), *The Oxford Handbook of Empirical Aesthetics*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198824350.013.43

Liang, F., Gotham, M., Johnson, M., & Shotton, J. (2017). *Automatic stylistic composition of Bach chorales with deep LSTM* [Conference session]. 18th International Society for Music Information Retrieval Conference.

Longoni, C., & Cian, L. (2020). Artificial intelligence in utilitarian vs. hedonic contexts: The “Word-of-Machine” effect. *Journal of Marketing*, *86*(1), 91–108. https://doi.org/10.1177/0022242920957347

Lonsdale, A. J., & North, A. C. (2009). Musical taste and ingroup favouritism. *Group Processes & Intergroup Relations*, *12*(3), 319–327. https://doi.org/10.1177/1368430209102842

Lonsdale, A. J., & North, A. C. (2017). Self-to-stereotype matching and musical taste: Is there a link between self-to-stereotype similarity and self-rated music-genre preferences? *Psychology of Music*, *45*(3), 307–320. https://doi.org/10.1177/0305735616656789

Malle, B. F., Scheutz, M., Arnold, T., Voiklis, J., & Cusimano, C. (2015). *Sacrifice one for the good of many? People apply different moral norms to human and robot agents* [Conference session]. Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction. https://doi.org/10.1145/2696454.2696458

McCourt, T., & Zuberi, N. (2016). Music and discoverer. *Popular Communication*, *14*(2), 123–126. https://doi.org/10.1080/15405702.2016.1199025

North, A. C. (2010). Individual differences in musical taste. *The American Journal of Psychology*, *123*(2), 199–208. https://doi.org/10.5406/amerjpsyc.123.2.0199

Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22–27. https://doi.org/10.1016/j.jbef.2017.12.004

Peck, L. S. L., & Grealey, P. (2020). Autobiographical significance of meaningful musical experiences: Reflections on youth and identity. *Music & Science*, *3*, 1–12. https://doi.org/10.1177/2059204320974221

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, *70*, 153–163. https://doi.org/10.1016/j.jesp.2017.01.006

Pelowski, M., Markey, P. S., Forster, M., Gerger, G., & Leder, H. (2017). Move me, astonish me . . . delight my eyes and brain: The Vienna integrated model of top-down and bottom-up processes in art perception (VIMAP) and corresponding affective, evaluative, and neurophysiological correlates. *Physics of Life Reviews*, *21*, 80–125. https://doi.org/10.1016/j.plrev.2017.02.003

Sanfilippo, K. R. M., Spiro, N., Molina-Solana, M., & Lamont, A. (2020). Do the shuffle: Exploring reasons for music listening through shuffled play. *PLOS ONE*, *15*(2), Article e0228457. https://doi.org/10.1371/journal.pone.0228457

Shank, D. B. (2014). Impressions of computer and human agents after interaction: Computer identity weakens power but not goodness impressions. *International Journal of Human–Computer Studies*, *72*(10–11), 747–756. https://doi.org/10.1016/j.ijhcs.2014.05.002

Shank, D. B., DeSanti, A., & Maninger, T. (2019). When are artificial intelligence versus human agents faulted for wrongdoing? Moral attributions after individual and joint decisions. *Information, Communication & Society*, *22*(5), 648–663. https://doi.org/10.1080/1369118X.2019.1568515

Shank, D. B., North, M., Arnold, C., & Gamez, P. (2021). Can mind perception explain virtuous character judgments of artificial intelligence? *Technology, Mind, and Behavior*, *2*(3). https://doi.org/10.1037/tmb0000047

Solberg, R. T., & Dibben, N. (2019). Peak experiences with electronic dance music: Subjective experiences, physiological responses, and musical characteristics of the break routine. *Music Perception*, *36*(4), 371–389. https://doi.org/10.1525/mp.2019.36.4.371

Solberg, R. T., & Jensenius, A. R. (2016). Pleasurable and intersubjectively embodied experiences of electronic dance music. *Empirical Musicology Review*, *11*(3–4), 301–318. https://doi.org/10.18061/emr.v11i3-4.5023

Steinbeis, N., & Koelsch, S. (2009). Understanding the intentions behind man-made products elicits neural activity in areas dedicated to mental state attribution. *Cerebral Cortex*, *19*(3), 619–623. https://doi.org/10.1093/cercor/bhn110

Tigre Moura, F., & Maw, C. (2021). Artificial intelligence became Beethoven: How do listeners and music professionals perceive artificially composed music? *Journal of Consumer Marketing*, *38*(2), 137–146. https://doi.org/10.1108/JCM-02-2020-3671

van den Bosch, I., Salimpoor, V. N., & Zatorre, R. J. (2013). Familiarity mediates the relationship between emotional arousal and pleasure during music listening. *Frontiers in Human Neuroscience*, *7*, Article 534. https://doi.org/10.3389/fnhum.2013.00534

Wu, L., & Wen, T. J. (2021). Understanding AI advertising from the consumer perspective. *Journal of Advertising Research*, *61*(2), 133–146. https://doi.org/10.2501/JAR-2021-004

Wu, Y., Mou, Y., Li, Z., & Xu, K. (2020). Investigating American and Chinese subjects’ explicit and implicit perceptions of AI-generated artistic work. *Computers in Human Behavior*, *104*, Article 106186. https://doi.org/10.1016/j.chb.2019.106186

Received November 12, 2021
Revision received July 7, 2022
Accepted July 8, 2022 ■