

# Investigating the relationship between liking and belief in AI authorship in the context of Irish traditional music

Ken Déguernel<sup>1,\*</sup>, Bob L. T. Sturm<sup>1,\*</sup> and Hugo Maruri-Aguilar<sup>2</sup>

<sup>1</sup>Royal Institute of Technology KTH, Lindstedtsvägen 24 SE-100 44 Stockholm, Sweden

<sup>2</sup>Queen Mary University of London, Mile End Road, London E1 4NS, UK

## Abstract

Past work has investigated the degree to which human listeners may be prejudiced against music knowing that it was created by artificial intelligence (AI). While these studies did not find a statistically significant relationship, the listening experiments were performed with music genres such as contemporary classical music or free jazz which are fairly welcoming of technology. In this work, we explore this prejudice in a context where strong opinions on authenticity and technology are typical: Irish traditional music (ITM). We conduct a listening experiment with practitioners of ITM asking each subject to first listen to a human performance of music generated by a computer in the style of ITM (this provenance is unknown to the listener), and then rate how much they like the piece. After rating all six pieces, each subject listens to each again but rates how likely they believe it is composed by a computer. The results of our pilot study suggest ITM practitioners tend to rate belief in AI authorship lower the more they rate liking a tune.

## Keywords

Creative AI systems, Appreciation bias, Liking, Expertise, Listening test, Irish traditional music

## 1. Introduction

One's experience of music can involve numerous factors, some of which are related to what one senses (e.g., skill and effort, materials, setting), and some related to what one knows (e.g., programmatic information, historic context, authenticity). Music appreciation is influenced by musical properties [1, 2, 3] modulated by personal [4, 5] and contextual factors [6, 7], such as socio-cultural contexts [8, 9]. Prejudice and expectations also play a role in one's engagement with music. For instance, Canonne [10] showed that the listening experience is drastically different whether someone thinks they are listening to a composition or an improvisation. And Kroger and Margulis [11] showed that music appreciation can be biased by a listener's belief that a performance is of a renowned musician or a student.

When applying artificial intelligence (AI) to music creation, how does one's knowledge about the involvement of AI impact their appreciation of the resulting music? Moffat and Kelly [12] investigated such bias by having music listeners rate their liking of particular pieces of music

---

CREAI 2022, Workshop on Artificial Intelligence and Creativity, Nov.28–Dec.02, 2022, Udine, Italy

\*Corresponding author.

✉ kende@kth.se (K. Déguernel); bobs@kth.se (B. L. T. Sturm); h.maruri-aguilar@qmul.ac.uk (H. Maruri-Aguilar)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



CEUR Workshop Proceedings (CEUR-WS.org)

in several styles, such as contemporary classical music or free jazz, and asking whether they thought each music excerpt was composed by a human or by a computer. This study was unable to find any bias for any of the styles. Pasquier et al. [13] performed more extensive listening experiments in these directions with many more subjects, but test only one kind of music (contemporary string quartets). Their results also suggest that bias against AI in music is not significant. Moura and Maw [14] investigated attitudes about the involvement of AI in music using a survey and a behavioral experiment. Interestingly, they found contradictory results, where the survey respondents display negative attitudes toward AI in music, but the experiment participants do not show any significant differences in their responses based on knowing whether the music comes from a human or AI.

In this paper, we investigate the extent to which bias against AI is present in a context where computer authorship is considered at odds in the musical practice. To what extent does the culture and context of the music to which AI is being applied matter when it comes to a listener's perception of the results? Inspired by these prior studies, we investigate the relationship between liking for a musical piece and the belief that it is AI-composed in the context of *Irish traditional music* (ITM). Participants of our experiments are active practitioners of ITM, and are drawn from traditional music programs in Ireland. Our hypothesis is that the context of ITM is one in which authenticity is so heavily human-centred that practitioners will show a bias against liking music they believe is authored by AI, or alternatively against believing music that they like is AI-composed.

In the next section, we review the details of past experiments in this area [12, 13, 14]. Section 3 briefly presents the context of ITM and its relationship to innovations, both social and technological. Both of these help motivate the decisions we make in the design and analysis of our listening experiment, which are described in Sec. 4 and 5, respectively. Section 6 discusses the results of our experiment. Finally, we conclude with a look towards future iterations of this experiment.

## 2. Previous studies investigating bias against AI in music

One of the challenges of studying bias against AI involvement in music is that there is no standard methodology on how to proceed. Different experimental designs have been proposed, as described for the following three studies.

Moffat and Kelly [12] presents a two-stage listening experiment conducted with 20 participants. Each of six stimuli of the experiment is a one-minute audio recording excerpt of a musical piece: three are designated computer-generated, and the other three human-composed. In the first stage, participants rate their liking of the stimulus on a 5-point Likert scale, and indicated whether they believed it was composed by a human or a computer. In the second stage, the participant are given information about the authorship of each stimulus, and then answer a written questionnaire asking how much they enjoyed the music, and whether they would buy the music, download it, or recommend it to a friend. The authors conclude from their results that their participants preferred music thought to be human-composed rather than computer-composed music, but the participants did not change their liking of a music piece after being told about its origins. That is, there was no evidence that a listener likes a piece less

or more after being told it was created by a computer or a human.

Pasquier et al. [13] presents an experiment reproducing and extending that of Moffat and Kelly. They use six video-recorded stimuli, three of which are from the same human composer, and the other three are generated by an AI system designed by that composer [15]. A trial of this experiment involves watching a video and then giving a 50-point rating along each of four dimensions: “Good–Bad”, “Like–Dislike”, “Emotional–Unemotional” and “Natural–Artificial”. All participants did this twice for each stimulus (randomized), but under three different conditions. In one condition, a participant is never told about human/computer authorship, and in another condition, a participant is told about authorship. In the third condition (“informed”) a participant is told about the authorship only after they first rate the six stimuli. Pasquier et al. [13] reports results from 122 participants, and conclude that there is no significant difference in any of the four dimensions between conditions, or even within the informed condition.

Moura and Maw [14] explores listener attitudes to the involvement of AI in music creation. They conducted an online survey with one group of people consisting of 72 music professionals and another of 374 non-professionals. Both groups showed some minor degree of questioning the credibility of musicians using AI, and reported a low likelihood of purchasing music created by AI. They also conducted a listening test with 86 university students split into two groups. In one group a subject reads a narrative describing the music they will hear as AI-generated. In the other group a subject reads a narrative describing human emotions and experiences reflected in the music. Both groups listened to the same music: two 1.5-minute excerpts of an AI (co-)composed work. Moura and Maw found no significant differences in responses between the two groups, and thus concluded that a listener’s perception of a song they like is not affected by knowing it was generated by AI.

### 3. The Context of Irish Traditional Music

Irish traditional music (ITM) is a complex genre encompassing many practices, and is an important part of Irish culture and identity [16, 17, 18]. Due in great part to the folk music revivals of the mid-20th century, as well as culture-focused organizations of Ireland, not to mention waves of immigration resulting from dire economic and environmental conditions in the 19th and 20th centuries, ITM has spread around the globe, and is actively practiced today by enthusiasts of the music [19].

The practice of ITM is accompanied by values that emphasize authenticity, etiquette and often nationalism [18, 20, 21]. For instance, there are strong opinions about the ways in which tunes should be learned and taught, how tunes should be performed, which tunes and instruments are acceptable, and so on. Several of these aspects of ITM are implicitly and explicitly codified in the local and national summer schools and music competitions organized throughout Ireland each year by cultural organizations, such as *Comhaltas*<sup>1</sup> and *Oireachtas na Gaeilge*,<sup>2</sup> founded to preserve and promote Irish culture [22, 23].

Considering the context of ITM and its values, there exists tension around innovation, as

---

<sup>1</sup><https://comhaltas.ie>

<sup>2</sup><https://www.antoireachtas.ie/>

well as how ITM should be used and presented [24]. In a speech delivered at a 1996 academic conference focused on traditional music in Ireland, the musician and Irish music advocate Tony MacMahon [25] relayed concerns about the loss of authenticity due to external forces of innovation by commercialization. Hillhouse [18] highlights the contradiction between community ownership and authorship inherent to ITM and the importance of intellectual property in national and international marketplaces. More recently, computer science research applying AI to modeling and imitating stylistic elements of ITM revealed friction around notions of authorship and the (im)proper treatment of the tradition [24]. That authenticity is so important to ITM suggests that there should be a bias on the part of the ITM practitioner against inauthentic forces, such as AI, coming to play. However, this has yet to be explored to the best of our knowledge.

## 4. Method

The hypothesis we want to test is the following: an ITM practitioner will exhibit an inverse relationship between liking a piece of music and believing it is authored by AI in the context of ITM.

### 4.1. Participants

We drew participants from traditional music programs at the University of Limerick, Ireland. The first cohort (E1) includes 20 participants (12 women and 8 men, aged 18–64 years,  $M=36.75$ ,  $SD=16.02$ ), who are students, teachers and technicians from the BLAS International Summer School of Irish Traditional Music and Dance.<sup>3</sup> The second cohort (E2) includes 26 participants (20 women and 6 men, aged 19–60 years,  $M=25.73$ ,  $SD=10.14$ ), who are students of degree programs offered at the Irish World Academy of Music and Dance.<sup>4</sup> Each participant was compensated with a €20 gift card.

### 4.2. Stimuli

Since we expected participants to be very knowledgeable about ITM, we decided not to use existing traditional tunes as stimuli for this experiment to prevent familiarity issues. Instead, we hired professional Irish accordionist Pádraig O'Connor<sup>5</sup> to select six tunes that he likes from large collections generated using a particular AI system [26, 27], including one consisting of 58,105 tunes [28].<sup>6</sup> The six double jigs selected by O'Connor all have an AABB form, with each section comprising eight 6/8 bars. Figure 1 shows the notation of one of these tunes.

O'Connor recorded himself playing each tune on solo accordion, with stylistic ornamentation, variation, bass, and harmonic content added as he saw fit. We permitted O'Connor to make minor changes to the notation of each tune as he wanted, but very few changes were made.

---

<sup>3</sup><https://www.blas.ie>

<sup>4</sup><https://www.irishworldacademy.ie/>

<sup>5</sup><http://www.paudieoconnor.com/>

<sup>6</sup>O'Connor participated as one of four judges in the AI Music Generation Challenge 2020, which focused on generating plausible Irish double jigs [27].



**Figure 1:** Notation of AI-composed double jig No. 8091 [27].

Having a professional ITM practitioner select and perform the stimuli ensures that they are presented in a realistic setting with an authentic performance. The resulting six stimuli are about the same duration ( $M=75.5$  seconds,  $SD=1.51$ ) and tempo ( $M=106$  bpm,  $SD=1.84$ ). All stimuli were recorded by O'Connor at his home with the same accordion and single microphone audio setup to avoid discrepancies in the audio quality. The stimuli were encoded with mp3 format (MPEG ADTS, layer III, v1, 128 kbps, 44.1 kHz, stereo) and are available online.<sup>7</sup>

### 4.3. Procedure

For both cohorts (E1 and E2), the experimental sessions took place in a media lab at the University of Limerick. We designed and hosted the experiments with a web-based interface built using jsPsych [29]. All participants were briefed and provided their informed consent before taking part. Participants used the same computer and headset models. A soundcheck using a recording of an Irish traditional tune allowed the participants to adjust the volume setting to a comfortable level.

The experimental paradigm comprised two tasks in series: the “liking task” and then the “authorship task”.

· In the **“Liking task”**, the participant is asked, “*How much do you like the tune?*”, and is given a 5-point Likert scale anchored by the labels (from left to right) “Don’t like it at all”, “Don’t like it”, “Neutral”, “Like it” and “Like it a lot”.

· In the **“Authorship task”**, the participant is asked, “*How likely do you believe that the tune is composed by a computer?*”, and is given a 5-point Likert scale anchored by the labels (from left to right) “Not likely at all”, “Not likely”, “Neutral”, “Likely” and “Very likely”.

These tasks were completed in this order to avoid liking ratings being influenced by a prior mention of AI. For each task, on-screen instructions were provided. In both tasks, the participants were encouraged to use the full range of the scale. The rating scale only appeared once the stimulus had finished playing. The order of stimuli in each task was randomized for each participant.

After completing the two tasks, the participants filled out a short questionnaire about demographics, musical practice, and familiarity with ITM. They answered the following questions:

- age: “*How old are you in years?*”
- gender: “*What is your gender?*”
- nationality: “*What is your nationality?*”

<sup>7</sup><https://www.kth.se/profile/bobs/page/research-data>

- education: “What is the highest level of education you have achieved?”
- mus\_pro: “Are you or have you been a professional musician?”
- irish\_fam: “How many Irish traditional tunes can you play/sing from memory?”, with options “0–10”, “11–50”, “51–100” and “100+”
- instrument: “What is (are) your main instrument(s)”

As a final question, the participant was asked to freely describe any strategies they used to determine whether a tune was composed by a human or a computer.

Each trial of this experiment lasted about 20 minutes. The experiment was reviewed and approved by the Ethics Committee of KTH (V-2021-0615).

#### 4.4. Data analysis

We code the demographic data as follows: age is left as an interval variable; gender and mus\_pro are nominal with two levels; nationality is nominal with two levels based on being Irish; instrument is nominal with two levels based on whether accordion is specified; education and irish\_fam are both ordinal with four levels.

A two-sample  $t$ -test on demographic data collected during the experiment shows that E1 and E2 differ significantly in age ( $t = 2.79$ ,  $df = 44$ ,  $p < .008$ ). A Mann-Whitney U test shows a significant difference between E1 and E2 only in mus\_pro ( $U = 171.0$ ,  $p < 0.03$ ), but does not show a significant difference in education ( $U = 315.5$ ,  $p > 0.16$ ), irish\_fam ( $U = 293.5$ ,  $p > 0.44$ ), gender ( $U = 304.0$ ,  $p > 0.22$ ), nationality ( $U = 196.0$ ,  $p > 0.06$ ), or instrument ( $U = 269.0$ ,  $p > 0.74$ ).

Our analytical approach involves two stages. In the first stage we model the bivariate responses of the participants using linear mixed-effects models [30]. In the second stage we regress on the individual coefficients of the model in the first stage using the demographic covariates mentioned above.

Denote by  $x_{jt}$  the value of the “Liking” by the  $j$ -th participant in the trial they evaluate the  $t$ -th tune, and  $y_{jt}$  the value of their belief that the same tune was composed by an AI (AIC). One model relating these responses is

$$x_{jt} = \mu + m_j y_{jt} + \beta_t + b_j + \varepsilon_{jt} \quad (1)$$

where  $\mu$  and  $m_j$  are the intercept a participant-based slope, respectively;  $\beta_t$  is a fixed effect of tune and  $b_j$  is a random effect of participant; and  $\varepsilon_{jt}$  is the residual error. An alternative mixed-effects model of the bivariate responses casts AIC as a function of reported liking:

$$y_{jt} = \mu + m_j x_{jt} + \beta_t + b_j + \varepsilon_{jt}. \quad (2)$$

All random quantities are considered independent with zero means and variances  $\sigma_j^2$  and  $\sigma^2$ , respectively. Each model considers tune to be a factor with six levels (and thus a fixed effect), because we only wish to draw conclusions about the population of participants. In other words, we wish to generalize our conclusions to ITM practitioners for these six stimuli, and not to the population of tunes O’Connor would curate and perform from a large collection.

The differences in interpretation of these models is important. Model (1) poses liking as a function of the belief of being AI-composed, which can be motivated by current thinking about

how aesthetic appreciation is a function of many factors, such as the value a stimulus has for a participant, the context of the perception of the stimulus, and the physiological state of the participant [31]. Model (2) seeks to determine how the factors considered by a participant in rating their belief a tune is AI-composed relate to or modulate factors they considered in rating their liking of the tune. If  $m_j$  is significantly different from zero then one might conclude there to be a significant overlap of these factors. However, since the number of stimuli are few and the time between the two tasks is short, we expect there to be some contribution of memory informing AIC for a participant. In other words, their memory of having rated their liking of a tune a particular way could inform their AIC rating of the tune. Model (2) can thus be motivated by a more machine-learning oriented goal where one wants to predict a participant’s AIC rating from their rating of their liking of the tune.

In the second stage, we attempt to explain the participant coefficients  $m_j$  of the first stage using participant covariates. To this end, denote by  $z_j$  the vector of covariate measurements, available for the  $j$ -th participant. The model is a standard regression

$$m_j = \theta_0 + \theta^T z_j + \varepsilon_j \quad (3)$$

where  $\theta_0$  and  $\theta$  are the model coefficients and  $\varepsilon_j$  is the usual independent error terms assumed normal with variance  $\sigma^2$ . In the above formula, for simplicity, we write  $m_j$  to mean the fitted coefficient stemming from the first stage of the analysis.

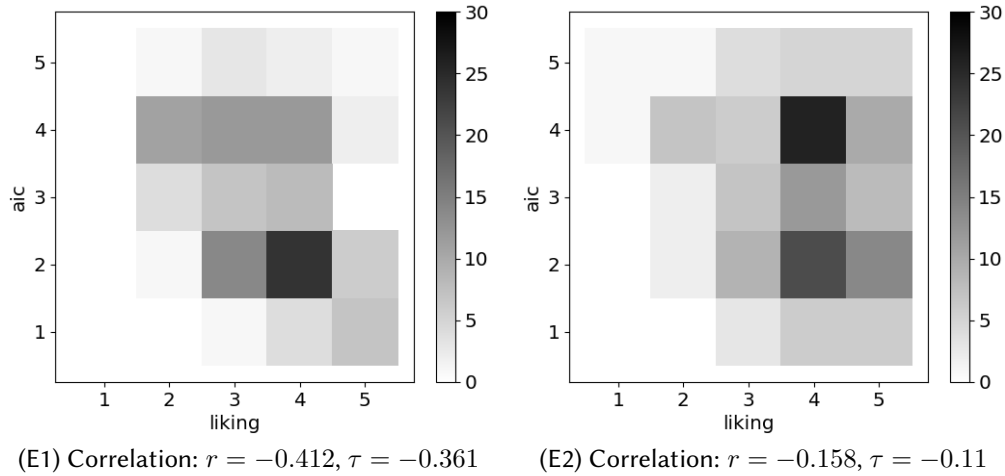
Finally, we analyze the free-form response about strategies using the method of constant comparison [32] to identify recurring themes and convergences in the strategies of participants. This allows us to explore potential explicit bias and identify which elements of ITM practice are expected to be different when composed by a computer. These, in turn, might reflect a listening focus in ITM practice that may inform future research in empirical musicology and generative systems.

## 5. Results

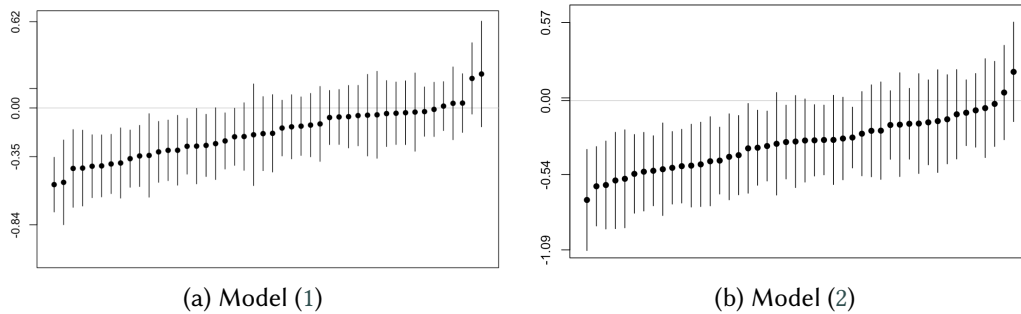
### 5.1. Quantitative analysis: Relationship between liking and belief in AI-authorship

Figure 2 illustrates a cross-tabulation of bivariate responses for each cohort and the numerical values of Pearson’s and Kendall’s  $\tau$  correlations. Exploratory analysis shows that high “Liking” scores are associated with low belief in AI “Authorship” (AIC). We find a negative correlation moderately strong in E1 while in E2 we find a weaker but still negative correlation.

Considering that the two cohorts are samples of the population of interest drawn from the same location, we first model the pooled data. The results are shown in the rows “E1+E2” in Table 1, and the estimates of the participant coefficients are shown in Fig. 3. The number of significant negative coefficients of Model (1) is 19 while that for Model (2) is 24. When regressing the  $m_j$  coefficients of Model (1) the only significant factors we find are cohort ( $\theta = 0.163, t = 2.62, p < 0.013$ ) and education ( $\theta = -0.073, t = -2.17, p < 0.04$ ). For the regression of the coefficients of Model (2), we find three significant factors: education ( $\theta = -0.116, t = -3.46, p < 0.002$ ), gender ( $\theta = -0.171, t = -2.60, p < 0.014$ ) and instrument ( $\theta = -0.388, t = -4.29, p < 0.0002$ ).



**Figure 2:** Tabulation of cases involving “Liking” and “Authorship” for cohorts E1 and E2.  $r$  and  $\tau$  correspond respectively to Pearson’s and Kendall’s correlation.



**Figure 3:** Estimated fixed effects  $m_j$  with 95% confidence intervals for all participants (ordered by effect size).

We now fit our models to data of each cohort individually, which is motivated first by the regression on the coefficients of Model (1), and second a difference in sampling participants between them at the University of Limerick. More specifically, E1 consists of students of a two-week-long summer school focused on Irish traditional music, as well as several staff (professors, administrators, etc.). E2 consists of students enrolled in longer-term educational programs about Irish traditional music. These cohorts may differ in other unmeasured ways as well, e.g., motivations to attend a two-week summer school are different from those to attend a longer educational program at a university. This could explain why cohort is a significant factor in the regression of the effects of Model (1) for the pooled data.

For Model (1) we see that E1 has a much larger number of significant negative  $m_j$  terms (14 of 20) relative to the number for E2 (4 of 26). For Model (2), E1 has 18  $m_j$  terms that are significantly less than zero, of which E2 has 7. For E1, when regressing on the coefficients of



Effect	Cohort	$r_m^2$	$r_c^2$	AkaikeIC	BIC	Lik	$d_{res}$	$m$ [95% CI]
liking~AIC (Model 1)	E1+E2	0.316	0.461	858.04	1042.23	-375.02	224	-0.18 [-0.28,-0.09]
	E1	0.263	0.263	388.59	459.80	-166.30	94	-0.35 [-0.50,-0.20]
	E2	0.332	0.505	485.13	581.02	-208.56	124	-0.05 [-0.17,0.07]
AIC~liking (Model 2)	E1+E2	0.253	0.566	1005.34	1189.57	-448.67	224	-0.31 [-0.47,-0.15]
	E1	0.350	0.350	426.91	498.12	-185.46	94	-0.46 [-0.68,-0.23]
	E2	0.206	0.732	588.29	684.17	-260.14	124	-0.20 [-0.43,0.03]

**Table 1**

Analysis results for Models 1 and 2 for both cohorts pooled and separately. The quantities  $r_m^2$  and  $r_c^2$  are the marginal and the conditional coefficients of determination, respectively. We also report the Akaike Information Criterion, Bayesian Information Criterion and (log)likelihood in columns AkaikeIC, BIC and Lik, respectively, as well as the degrees of freedom for the residual  $d_{res}$ . The column  $m$  has the estimate of average of  $m_j$  with a confidence interval.

Model (1), we find significant positive contribution from `irish_fam` ( $\theta = 0.049$ ,  $t = 2.69$ ,  $df = 12$ ,  $p < 0.02$ ) and `age` ( $\theta = 0.003$ ,  $t = 2.45$ ,  $df = 12$ ,  $p < 0.031$ ). The adjusted  $R^2$  of this model is 0.301. When regressing on the coefficients of Model (2) for E1, we find significant contributions from all covariates except `mus_pro`: `age` ( $\theta = 0.005$ ,  $t = 3.72$ ,  $df = 12$ ,  $p < 0.003$ ), `education` ( $\theta = -0.048$ ,  $t = -2.30$ ,  $df = 12$ ,  $p < 0.04$ ), `irish_fam` ( $\theta = 0.078$ ,  $t = 4.42$ ,  $df = 12$ ,  $p < 0.001$ ), `gender` ( $\theta = -0.12$ ,  $t = -2.28$ ,  $df = 12$ ,  $p < 0.042$ ), `nationality` ( $\theta = -0.145$ ,  $t = -2.87$ ,  $df = 12$ ,  $p < 0.015$ ), and `instrument` ( $\theta = -0.248$ ,  $t = -3.9$ ,  $df = 12$ ,  $p < 0.003$ ). The adjusted  $R^2$  of this model is 0.676. When regressing on the coefficients of either model for E2, we find no significant contributions from the covariates (adjusted  $R^2 < 0.18$ ).

The explanatory variables had interesting imbalances between cohorts which we briefly describe. Concerning variable `age`, E2 was younger than E1, but both cohorts covered about the same range of age; that is, E2 has a more skewed distribution than E1 in age. Although we can test equality of means and reject it strongly between cohorts ( $p = 4 \times 10^{-10}$ ), the shape of distributions is such that the influence in the analysis is much deeper than just the location. Concerning binary variables, `mus_pro` had opposing proportions between cohorts; whereas variables `gender`, `nationality` had less pronounced differences between cohorts, only variable `instrument` had similar patterns between E1 and E2.

## 5.2. Qualitative analysis: Biases in self-reported strategies

At the conclusion of the listening test, each participant is asked “*What strategies did you use to determine if a tune was composed by a human or a computer?*”. Our qualitative analysis of the free-form responses done across both cohorts find five main strategies:

- 17 participants (37.0%) (5 from E1, 12 from E2) reported *listening to repetitions and patterns*. Tunes that were deemed “overly repetitive” were associated with computer authorship.
- 12 participants (26.1%) (8 from E1, 4 from E2) reported *listening to structure and harmony*. In particular, participants listened to how phrases were linked together as well as chord progressions and cadences. “Unnatural” harmonic and melodic structures or chords following a

Adjectives/Descriptives used	
for computers/AI	for humans
<i>inorganic, unnatural, uncanny valley feeling, simple, weird, rigid, out of place, robotic, unusual, logical, generic, predictable, algorithmic</i>	<i>catchy, flowed, clarity, fluid, like speaking language, alive, emotive, organic, usual, finicky, surprising, sustaining interest, with purpose, familiar, creative, thought out, natural</i>

**Table 2**

List of adjectives/descriptives used by participants in their self-reported strategies to rate the authorship of a tune.

structure deemed “very rigid” were associated with computer authorship.

- 11 participants (23.9%) (4 from E1, 7 from E2) reported *listening to variation and ornamentation*. Tunes with more variations in the melody and with more ornamentation/embellishment were associated with human authorship.

- 10 participants (21.7%) (4 from E1, 6 from E2) reported *listening to familiarity*. Tunes that were deemed to fit with the style of Irish music and sounded familiar in that aspect were associated with human authorship. By contrast, tunes deemed too similar to existing tunes or too “generic” were associated with computer authorship.

- 7 participants (15.2%) (5 from E1, 2 from E2) reported *listening to instrumental technique*. Tunes with elements deemed “unnatural” to play on the accordion, such as “unusual” phrasing or range were associated with computer authorship.

It is also noteworthy that 5 participants tried to use strategies based on audio quality, thinking that some of the tunes were synthesized (although all tunes were human-performed in the same conditions). This shows some potential confusion on what is meant by “computer-composed”.

We also looked at the adjectives/descriptives used by the participants when talking about the tunes they believed to be AI- or human-composed (see Table 2). in order to have a basic sentiment analysis of their report. When talking about what they believe to be computer generated, the descriptives used by participants were almost exclusively negative, with a few neutral ones. Conversely, when talking about what they believe to be human-composed, the descriptives used were for the large majority positive. This difference points to a potential conscious prejudice regarding what a subject believes the capacities of AI or computers are compared to humans when it comes to ITM composition. Anecdotal but still amusing, one of the participants even claimed that their strategy was assuming that “*the tunes [they] liked better were composed by humans and the ones [they] disliked were composed by the computer.*”

## 6. Discussion

Our quantitative analysis points to a plausible bias among ITM practitioners for these six AI-composed tunes: they tend to like more the tunes they deem hardly likely to be composed by an AI. Alternatively, the more they report liking a tune the less they report believing the tune is AI-composed. The difference in results with the previous studies could validate our hypothesis on the importance of the context (both in terms of musical culture and participants)

when it comes to observing such a bias. However, as it stands, this study can only be considered a pilot study before performing a power analysis of the experiment to determine its likelihood of producing a Type-I error. Nonetheless, this pilot study provides us with valuable information regarding further testing of our hypothesis. In particular, expertise and professionalism must be explicitly defined, accounting for differences between instrumental or vocal traditions within ITM. Our wording of the professionalism question (`mus_pro`) makes no distinction between career performers and a musician who has been paid to play at some event.

The analysis of the self-reported strategies of the participants suggests a conscious bias against AI authorship with more limited expectations on computer capabilities when it comes to some musical criteria, and an overall negative sentiment when describing tunes they believed to be computer composed. It is interesting to compare that attitude with the reception of technology in general in ITM. Cawley [23] studied ethnographically the use and reception of technology as a cultural process in the enculturation of Irish traditional musicians. From the use of music notation to audio and video recording and the numerous websites providing ITM educational resources, technology has changed the way ITM is learned, played and discussed. These technologies are now well-accepted overall and an integral part of the day-to-day life of Irish traditional musicians. However, a couple of caveats are noted: the “information overload” arising from the amount of resources available, and the alienation of the tradition that can come with the use of technology – especially if used without engaging in the “traditional” social and musical interactions with other musicians. A good illustration of this relationship with technology can be observed with Tunepal [33], a service enabling people to retrieve the name and music notation of a tune from a short recording. Tunepal is the most downloaded traditional music software, is widely used during sessions and classes, and generally well-received by musicians, with some criticism from a minority pointing out that it sometimes hinders engagement with ITM practitioners [33].

It is therefore difficult to argue that a bias against AI authorship emerges only from a general negative attitude toward technology. The difference in attitude that arises in the reception of AI music generation shows that this practice might raise some specific issues. A potential explanation, suggested by our qualitative results, could simply be the belief that current AI systems are not skilled enough to reach human capabilities and we are observing a similar bias as in Kroger and Margulis [11]. Alternatively, such a negative attitude could also stem from a more Romantic and human-centered notion of creativity as an unconscious process associated with the notions such as “inspiration” or “genius” [34, 35].

It is also worth noting that AI music generation is not the only use of technology where tension becomes apparent in ITM. For instance, Tunetracker [36] – a software designed to “surveil” local sessions in a pub by documenting which tunes were played which days and in which configuration with other tunes – was met with concerns from practitioners going from the sheer presence of a microphone in a pub (although it was explicitly not used to store or broadcast recordings), to more complex issues such as the idea of the homogenization of the practice, or being caught by the Irish Music Rights Organisation seeking to collect royalties on modern, copyright protected tunes.

## 7. Conclusion and future work

Overall, the results of our pilot study are encouraging. First, for these six tunes and our ITM practitioners, our quantitative analysis shows evidence that they tend to have a negative association between liking a tune and believing that it is AI-composed. Second, our qualitative analysis points to a conscious prejudice against the application of AI systems to ITM. The results from this pilot study encourages us to continue our investigation, and study larger populations. It would be interesting to conduct an online version of this listening test in order to test our hypothesis in the general population, and observe how bias might appear in non-practitioners of ITM. It would also be interesting to test whether such differences hold in other musical traditions that value authenticity in ways similar to ITM. However, given the potential importance of the cultural context for the observation of bias, it is difficult to define a standardised experimental design. Appropriate decisions should be taken to adapt to the different musical traditions and to the participants. We also want to conduct a more in-depth qualitative study involving ITM professionals and students using process-oriented research methodology such as Think-Aloud Protocols involving introspection and retrospection [37] as well as participant-oriented research with semi-guided interviews [38]. This research would help us have a better understanding on the prejudice ITM practitioners may have with the involvement of AI in music composition or AI-assisted co-composition, and help inform future work on the ethics of this field.

## Acknowledgments

This paper is an outcome of MUSAiC, a project that has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation program (Grant agreement No. 864189). We would like to thank P. Cotter, S. Joyce, N. Keegan, and A. Dormer for letting us conduct our study at the BLAS Summer School at the Irish World Academy of Music and Dance, University of Limerick, Ireland. We would like to thank P. O'Connor for taking part in the creation of the corpus of stimuli and for his interpretation of the tunes. We would like to thank A. Clemente for discussions about the psychological aspects of this work.

## References

- [1] G. Ilie, W. F. Thompson, A comparison of acoustic cues in music and speech for three dimensions of affect, *Music Perception* 23(4) (2006) 319–330.
- [2] V. Salimpoor, D. Zald, R. Zatorre, A. Dagher, A. McIntosh, Predictions and the brain: How musical sounds become rewarding, *Trends in Cognitive Sciences* 19(2) (2015) 86–91.
- [3] A. J. Milne, S. A. Herff, The perceptual relevance of balance, evenness, and entropy in musical rhythms, *Cognition* 203 (2020) 104233.
- [4] I. Lahdelma, T. Eerola, Cultural familiarity and musical expertise impact the pleasantness of consonance/dissonance but not its perceived tension, *Scientific Reports* 10 (2020) 8693.
- [5] M. Orr, S. Ohlsson, Relationship between complexity and liking as a function of expertise, *Music perception* 22(4) (2005) 583–611.
- [6] E. Brattico, From pleasure to liking and back: Bottom-up and top-down neural routes to the

- aesthetic enjoyment of music, in: J. Huston, M. Nadal, F. Mora, L. Agnati, C. Cela-Conde (Eds.), *Art, aesthetics and the brain*, Oxford University Press, 2015, pp. 303–318.
- [7] A. North, D. Hargreaves, J. Hargreaves, Uses of music in everyday life, *Music perception* 22(1) (2004) 41–77.
- [8] A. North, D. Hargreaves, *The social and applied psychology of music*, Oxford University Press, 2008.
- [9] A. Greasley, A. Lamont, Musical preferences, in: S. Hallam, I. Cross, M. Thaut (Eds.), *The Oxford Handbook of Music Psychology*, Oxford University Press, 2016, pp. 263–284.
- [10] C. Canonne, Listening to improvisation, *Empirical Musicology Review* 13(1–2) (2018).
- [11] C. Kroger, E. H. Margulis, “But they told me it was professional”: Extrinsic factors in the evaluation of musical performance, *Psychology of Music* 45(1) (2016) 49–64.
- [12] D. C. Moffat, M. Kelly, An investigation into people’s bias against computational creativity in music composition, *Assessment* 13(11) (2006).
- [13] P. Pasquier, A. Burnett, N. Gonzalez Thomas, J. B. Maxwell, A. Eigenfeldt, T. Loughin, Investigating listener bias against musical creativity, in: *Proc. of the 7th International Conference on Computational Creativity*, 2016, pp. 42–51.
- [14] F. T. Moura, C. Maw, Artificial intelligence became Beethoven: how do listeners and music professionals perceive artificially composed music, *Journal of Consumer Marketing* 38(2) (2021) 137–146.
- [15] A. Eigenfeldt, Corpus-based recombinant composition using a genetic algorithm, *Soft Computing* 16(12) (2012) 2049–2056.
- [16] M. Ó Súilleabháin, Irish music defined, *The Crane Bag* 5(2) (1981) 83–87.
- [17] M. Trachsel, Oral and literate constructs of “authentic” irish music, *Éire-Ireland* 30(3) (1995) 27–46.
- [18] A. N. Hillhouse, *Tradition and innovation in Irish instrumental folk music*, Master’s thesis, The University of British Columbia, 2005.
- [19] H. O’Shea, *The making of Irish Traditional Music*, Cork University Press, Cork, Ireland, 2008.
- [20] G. Ó hAllmhuráin, *A pocket history of Irish Traditional Music*, The O’Brien Press, 1998.
- [21] M. D. Nichol森, Francis O’Neill, music collection, and Irish traditional musicians in Chicago, 1898-1921, in: R. T. Cornish, M. Quintelli-Neary (Eds.), *Crafting Infinity: Reworking Elements of Irish Culture*, Cambridge Scholars Publishing, 2012.
- [22] S. Spencer, Traditional Irish Music in the twenty-first century: Networks, technology, and the negotiation of authenticity, in: S. Brady, F. Walsh (Eds.), *Crossroads: Performance studies and Irish culture*, Palgrave Macmillan, 2009, pp. 58–70.
- [23] J. Cawley, *The musical enculturation of Irish traditional musicians: An ethnographic study of learning processes*, Ph.D. thesis, National University of Ireland, Cork, 2013.
- [24] R. Huang, B. L. T. Sturm, Reframing “aura”: Authenticity in the application of AI to Irish Traditional Music, in: *Proc. of the AI Music Creativity conference*, 2021.
- [25] T. MacMahon, The language of passion, in: *Crossroads Conference*, 1996.
- [26] B. L. Sturm, J. F. Santos, O. Ben-Tal, I. Korshunova, Music transcription modelling and composition using deep learning, in: *Proc. of the Conference on Computer Simulation of Musical Creativity*, Huddersfield, UK, 2016.
- [27] B. L. T. Sturm, H. Maruri-Aguilar, *The Ai Music Generation Challenge 2020: Double jigs*

- in the style of O'Neill's "1001", *Journal of Creative Music Systems* (2021).
- [28] B. L. T. Sturm, 2021, 58,105 irish-style double jigs, URL: <http://kth.diva-portal.org/smash/record.jsf?pid=diva2%3A1562396>.
  - [29] J. R. de Leeuw, A javascript library for creating behavioral experiments in a web browser, *Behavior research methods* 47(1) (2015) 1–12.
  - [30] J. J. Hox, M. Moerbeek, R. Van de Schoot, *Multilevel analysis: Techniques and applications*, Routledge, 2017.
  - [31] M. Skov, Aesthetic appreciation: The view from neuroimaging, *Empirical Studies of the Arts* 37(2) (2019) 220–248.
  - [32] K. Henwood, N. Pidgeon, Grounded theory in psychological research, in: P. Camic, J. Rhodes, L. Yardley (Eds.), *Qualitative Research in Psychology: Expanding Perspectives in Methodology and Design*, American Psychological Association, 2003, pp. 131–156.
  - [33] B. Duggan, B. O'Shea, Tunepal: searching a digital library of traditional music scores, *OCLC Systems & Services: International digital library perspectives* 27(4) (2011) 284–297.
  - [34] R. Pope, *Creativity: Theory, History, Practice*, Routledge, 2005.
  - [35] C. G. Johnson, The creative computer as Romantic hero? Computational creativity systems and creative personae, in: *Proc. of the International Conference on Computational Creativity*, 2012, pp. 57–61.
  - [36] B. Duggan, N. M. Su, TuneTracker: Tensions in the surveillance of traditional music, in: *Proc. of the ACM Conference on Designing Interactive Systems*, 2014, pp. 845–854.
  - [37] T. Boren, J. Ramey, Thinking aloud: Reconciling theory and practice, *IEEE Transactions on Professional Communication* 43(3) (2000) 261–278.
  - [38] G. Saldanha, S. O'Brien, *Research methodologies in translation studies*, Routledge, 2014.