EPJ Data Science

a SpringerOpen Journal

REGULAR ARTICLE   **Open Access**

Large scale analysis of gender bias and sexism in song lyrics

Lorenzo Betti[1,2]* , Carlo Abrate[3,4] and Andreas Kaltenbrunner[1,5]

*Correspondence:
lrn.betti@gmail.com
[1] ISI Foundation, Via Chisola 5, 10126
Turin, Italy
[2] Department of Network and Data
Science, Central European
University, Quellenstraße 51-55,
1100 Vienna, Austria
Full list of author information is
available at the end of the article

Abstract

We employ Natural Language Processing techniques to analyse 377,808 English song lyrics from the "Two Million Song Database" corpus, focusing on the expression of sexism across five decades (1960–2010) and the measurement of gender biases. Using a sexism classifier, we identify sexist lyrics at a larger scale than previous studies using small samples of manually annotated popular songs. Furthermore, we reveal gender biases by measuring associations in word embeddings learned on song lyrics. We find sexist content to increase across time, especially from male artists and for popular songs appearing in Billboard charts. Songs are also shown to contain different language biases depending on the gender of the performer, with male solo artist songs containing more and stronger biases. This is the first large scale analysis of this type, giving insights into language usage in such an influential part of popular culture.

Keywords: Song lyrics; Gender; Natural language processing; Word embeddings; Language bias; Sexism

1 Introduction

Music allows the expression of emotions, feelings and ideas through verbal and auditory language. Multiple messages can be conveyed by means of this language both at an emotional level, mostly through sounds, and at a verbal level through song lyrics [1]. In particular, musical lyrics can contain verbal messages of various nature, through which ideas of individuals and social groups have been spread in society. Song lyrics are also a popular expression of culture, and they reflect or maybe even emphasize many social phenomena. They hold ideas that have occurred over time about different social issues, such as gender discrimination and sexism [2–4]. As stated by Davis in [5], song lyrics "are more than mere mirrors of society; they are a potent force in the shaping of it". As such, song lyrics are an important albeit underexplored data source to observe and measure societal changes.

Natural Language Processing (NLP) techniques are an ideal tool for analyzing song lyrics due to their textual nature. These techniques have proven to be effective in improving results for companies, producers, and songwriters in the music industry [6]. However, there is growing concern that NLP models may inadvertently learn and perpetuate existing biases encoded in the training data when trained on large text corpora [7]. This raises the possibility of biases being amplified in language models and systems that use them [8, 9].

Springer

However this undesired property of NLP models can also be turned into an opportunity when it comes to the measurement of linguistic biases present in text corpora. Previous studies have demonstrated that word embeddings, which represent words as high-dimensional vectors [10], are capable of capturing linguistic biases related to human biases [11], as well as reflecting stereotypes towards women and ethnic minorities [12, 13]. The magnitude of such biases can vary depending on the domain of the text corpus [14, 15]. These findings provide evidence of the effectiveness of measuring linguistic biases in text through word embeddings.

NLP techniques have also shown promise in identifying hate speech and other forms of discriminatory language in text corpora [16, 17]. These techniques involve models trained on annotated datasets that have been labeled for various concepts related to hate speech, such as sexist content. The resulting models can then be used to automatically detect and quantify for example the presence of sexist content in new text data.

In this study we take advantage of some of these NLP techniques and analyse English song lyrics regarding two aspects receiving increasing attention: gender-related language bias and sexism in language. We consider two main research questions:

RQ1: Is there evidence of sexist content in song lyrics? Can we find differences with respect to artist gender and musical genres across time?

RQ2: Are gender and social role stereotypes reflected in song lyrics? Are there differences in the songs with respect to artist gender?

To answer these research questions, we use *Song Lyrics Data* from the "Two Million Song Database" of the WASABI project [18] to obtain a large corpus of song lyrics and song related information. We enrich the songs' metadata with Billboard chart performance to get the popularity trend of songs that spans from 1960 to 2010.

Then, we apply and adapt the sexism classifier of Samory et al. [19] to explore the presence of *sexism in song lyrics.* This approach allows to analyze the evolution of sexism in song lyrics of the WASABI dataset across time. We find an increase over time of sexist content in popular song lyrics by male solo artists, and that Hip hop and R&B and Soul songs have a higher fraction of sexist lyrics when compared to the other analyzed genres.

Finally, we use different word embedding association tests to detect *(gender) language bias in song lyrics* [11, 20, 21]. Differentiating the analysis by gender, we show the significant presence of language biases in the WASABI dataset, in particular for male solo artists, in contrast to more gender neutral bias of female solo artists.

The main contributions of this work consist in providing (to our best knowledge):

- the first large scale and longitudinal exploratory data analysis employing an automatic method to identify song lyrics containing sexist content,
- the first extensive study of language bias in song lyrics segregating by artist gender.

2 Related work

The analysis of gender stereotypes and sexism expressed in language with NLP methods has been a growing area of research in recent years [22]. These methods are closely entangled with the detection and mitigation (i.e. removal or reduction) of gender bias in NLP models and their output (see for example Sun et al. [23] for a review on mitigation). A promising approach for both, detection and mitigation, relies on measuring gender bias from the association between word vectors, for example through their cosine similarity. Bolukbasi et al. [12] proposed to first identify a gender direction defined by the subspace

spanned by gendered words (e.g., she and he, women and man). Then, they quantify gender bias as the projection of ideally gender-neutral words (e.g., job and profession names) onto the gender subspace.

This idea, combined with averaging methods and hypothesis testing has later led to the development of the Word Embedding Association Test (WEAT) [11]. The significance and sensitivity of WEAT has been analyzed by [24] and it has been successfully employed for example in [20] to quantify gender stereotypes in language corpora, adding a single category version (SC-WEAT) to the analysis. WEAT has later been extended further to SWEAT [21] to compare the relative polarization of two corpora (like male and female authored lyrics in our case). We will exploit WEAT, SC-WEAT and SWEAT here and describe them in more detail in Sect. 3.3. Besides word embeddings, similar approaches have been used to measure gender bias in large language models [15, 25].

Part of our work furthermore builds on the results of a sexism classifier developed by Samory et al. [19]. Since the exact definition of misogyny and sexism may be under discussion [26], the authors of [19] took into account different dimensions of sexism to increase model validity and furthermore improved the model reliability through including adversarial examples. Other noteworthy approaches to sexism detection use support vector machines, sequence-to-sequence models and a FastText classifier [27]; a BERT-based architecture to detect misogyny and aggression simultaneously on social media [28] or compare different models for identifying misogyny across languages and domains on Twitter data [29]. We refer to the following reviews for a more comprehensive overview about automatic misogyny and hate speech detection [16, 17].

In relation to our research questions, most of the works on gender stereotypes and sexism in song lyrics have targeted popular songs using manual content analysis, which allows studying fine-grained constructs related to sexism such as objectification and sexualization [30, 31]. These studies analysed popular songs from the 1960s to the 2000s showing that sexualization and mentions of sexual desire increased dramatically only after the 1990s, while mentions of love decreased. This is interpreted as a signal of the increase of sexism in songs because lust in the absence of love is likely to objectify the object of the desire [30]. Other works consider both the differences between male and female artists and the gender of the target being objectified. In [32], authors analyze the same time period considered in our work, finding female artists more likely to sing about love, while men are more likely to objectify others (both men and women), with a stronger emphasis towards women. Flynn et al. [33] confirmed these findings, adding that women were more likely to objectify themselves than men do. Moreover, Rap, Hip hop and R&B songs are the genres with the most objectification. These works rely on different aspects of the concept of sexism, which is possible to code through manual content analysis. Although our methodology can not distinguish between such nuances, it allows to analyse song lyrics at a larger scale.

Here we are more interested in studies using large datasets like the one from [34], which analyses bias in half a million song lyrics using WEAT scores [11]. This study does not segregate its results neither by gender nor in the temporal dimension and finds that bias in songs is strongest in relation to gender stereotypes and career paths. However, also gender biases in relation to Math vs Arts and Science vs Arts are found, meaning that both math and science words are more closely related to male terms while females terms are more closely related to art. All these biases are similar (albeit a bit smaller) to what can be

observed in a large internet crawl of texts [11], thus indicating that biases present in song lyrics mirrors the biases that exist in society. Recently, in a study performed in parallel to ours it was found that in song lyrics men are more likely than women to be associated to traits depicting them as competent, even though this bias becomes weaker going forward in time [35]. The authors of this work, albeit incorporating a temporal dimension, used a simpler not standardized metric of gender bias and measured only a single trait, whereas our analysis investigates several traits and uses different association tests used as well in other published works [11, 20, 21].

Other tasks for which song lyrics have been used for is song mood [36] and sentiment [37] classification, or together with audio features for genre classification [38] and song popularity [39]. The performance in the later two tasks with lyrics alone has been explored in [40].

3 Data collection and methods

3.1 Data collection and filtering

The WASABI database is a knowledge base that includes data about 77K artists, 200K albums, and more than two million songs [18]. We queried the database for all solo artists and groups having published more than 10 songs, and collected all their English[1] song lyrics published between 1960 and 2009. The database contains information about solo artists and band members, including their gender. We assigned the gender label "male" ("female") to bands composed of only male (female) members, and "mixed" to bands with both male and female members. We discarded solo artists without gender information, as well as groups with at least one member without it. The database does not include other gender identities, except for 9 artists whose gender is labeled as "Other" and accounting for a total of 44 songs. We thus decided to consider artists' gender as binary, although we acknowledge this limitation that prevents us from taking into account other gender identities. Furthermore, we filtered out all the songs whose lyrics or publication year are unavailable, whose lyrics are shorter than 10 words or composed of less than 4 lines. After removing duplicate lyrics (details in Appendix B), our final dataset consists of 377,808 song lyrics: 244,146 are performed by 7131 solo artists and the remaining 133,662 by 4294 groups. Finally, we retrieved also the song genres, which was not available for 46,482 songs (12.3%). The number of unique genres was reduced by replacing them with their corresponding top-level genres.[2]

To identify popular songs in the WASABI dataset, we retrieved the Billboard Hot 100 weekly charts, composed of all-genre weekly song charts released since 1958 [41]. We furthermore extracted the top 10 songs from all the charts. We were able to map 10,798 out of 24,180 unique songs (44.7%) of the Billboard, and 2608 out of 4348 unique songs (60.0%) of the Billboard top 10 charts by using approximate string matching to match both the title and the artist of songs (details in Appendix C). In the following, when referring to songs in Billboard or in Billboard top 10 charts, we refer to these sets of songs that we mapped to the main WASABI dataset.

---

[1] The WASABI database provides the language of song lyrics through the "language_detect" field.

[2] We used the taxonomy of the Wikipedia page of popular genres https://en.wikipedia.org/wiki/List_of_music_genres_and_styles.

3.2 Sexism detection

We fine tuned a BERT[3] classifier to detect sexist passages in texts using the dataset and the code provided by Samory et al. [19]. The dataset contains texts manually labeled following a scale-based codebook that operationalizes the concept of sexism. The codebook consists of four non-overlapping categories resulting from the review of psychological scales measuring sexism and related constructs. These include a broad range of sexist content that can be present in song lyrics such as behavioral expectations, stereotypes & comparisons, endorsements of inequality, and denying inequality & rejection of feminism. In addition to sexist content, the codebook also includes categories that take into account sexist phrasing, in order to distinguish sexist content from texts containing only uncivil content or common profanity. This makes the classifier perform well on out-of-domain data, also thanks to the inclusion of human-written adversarial examples. We refer to Samory et al. [19] for further details on the codebook.

Since the original dataset consists of short texts, we adapted the representation of the input to classify song lyrics. In detail, lyrics are divided into batches composed of groups of four lines, each of them sharing two lines with the previous and following group. A batch is considered to contain sexist content if the model outputs a probability higher than a certain threshold. Whenever a song lyric contains at least one batch identified as sexist, we propagate the sexist label to the song. To verify the performance on song lyrics, we evaluated the classifier on an external dataset [42] of 190 lyrics,[4] 40.5% of which are considered to contain sexist content. For the optimal classification threshold of 0.725, the classifier achieves a precision of 0.73. We discuss in Appendix D the results obtained for different classification thresholds and requiring a larger number of batches labeled as sexist to propagate the label to the whole song, showing that our main findings are not affected by these choices.

3.3 Language biases from word embeddings

Words can be represented by vectors by leveraging the co-occurrence of nearby words in such a way that vectors that are close to each other represent words sharing a similar semantic meaning [43]. This representation of words is able to encode word analogies (e.g., "King" is to "Man" as "Queen" is to "Woman"), but it may also encode language biases and stereotypical analogies that are present in the training corpus [11]. For example, from a higher co-occurrence of pleasant words next to music instrument than to weapon words, we can likely expect word embeddings of unpleasant words to be closer to weapon than musical instrument words. We measured language biases in three different subsets of the lyrics corpus: all lyrics of solo artists, lyrics of male solo artists, and lyrics of female solo artists. This allowed us to compare the biases that are present in song lyrics performed by male and female artists. In the following, we list the methodologies we used to measure these kinds of associations in word embeddings.

WEAT: The Word Embedding Association Test (WEAT) measures the association between two sets of target words X and Y (e.g., music instruments and weapons), and two sets of attribute words A and B (e.g., pleasant and unpleasant words). The WEAT tests the null hypothesis that there is no difference between the two sets of target words in terms of

---

their relative similarity to the two sets of attribute words [11]. The similarity between two words a and b is defined as the cosine similarity $\cos(\vec{a}, \vec{b})$ between their representation in a vector space. The effect size is defined as:

$$\frac{\text{mean}_{x \in X} s(x, A, B) - \text{mean}_{y \in Y} s(y, A, B)}{\text{pooled\_std\_dev}_{w \in X \cup Y} s(w, A, B)},$$

where

$$s(w, A, B) = \text{mean}_{a \in A} \cos(\vec{w}, \vec{a}) - \text{mean}_{b \in B} \cos(\vec{w}, \vec{b}) \tag{1}$$

and pooled_std_dev is the pooled standard deviation [44]. [5] From this definition, positive values of the effect size indicate that the words in X are more similar to the words in A than in B, and the words in Y are more similar to the words in B than in A. Recalling the above example, we would expect a positive effect size. The significance is computed in the following way. The test statistic is defined as:

$$S(X, Y, A, B) = \sum_{x \in X} s(x, A, B) - \sum_{y \in Y} s(y, A, B).$$

Given 1000 random partitions of equal size for the union of the two target sets $\{(X_i, Y_i)\}_{i=1,\dots,1000}$, the one-sided P value is computed as

$$\text{Pr}_i\big[S(X_i, Y_i, A, B) > S(X, Y, A, B)\big].$$

SC-WEAT: Since the WEAT condensates into one effect size the relative associations of attribute sets against two target sets, the similarity against one single target set is lost. We used the Single Category Word Embedding Association Test (SC-WEAT) to disentangle this relative association [20]. The SC-WEAT measures the association between one set of target words W and two sets of attribute words A and B. If the SC-WEAT score is positive (negative), words in W are more (less) similar to the words in A compared to B. The procedure to compute the significance is analogous to the one described for the WEAT. Note that the WEAT and the SC-WEAT are not redundant because the former measures the relative association of X and Y with A and B, whereas the SC-WEAT evaluates the association of one target set per time, and their combination can add information to the language biases analysis.

SWEAT: The methods described above measure associations between word sets whose embeddings are learnt from a single text corpus. To compare the association between two corpora, we employed the Sliced Word Embedding Association Test (SWEAT) [21]. Given the word vectors learnt on two corpora $\mathcal{D}_1$ and $\mathcal{D}_2$, the SWEAT score is defined as:

$$S(W, A, B, \mathcal{D}_1, \mathcal{D}_2) = \sum_{w \in W} s(w, A, B, \mathcal{D}_1) - \sum_{w \in W} s(w, A, B, \mathcal{D}_2),$$

---

[5]We found different implementations of the WEAT, which compute the effect size using either the standard deviation or the pooled standard deviation (i.e. a weighted average of standard deviations). We opted for the latter solution because it reflects the definition of the effect size in terms of Cohen's D defined in Caliskan et al. [11].

where $s(w, A, B, \mathcal{D})$ has the same form as in Equation (1), but word vectors are taken from the distributional representation $\mathcal{D}$. A positive SWEAT score indicates that the word vectors of a target set in $\mathcal{D}_1$ are relatively more associated to the words in A than in B, and the word vectors in $\mathcal{D}_2$ are relatively more associated to the words in B than in A. Again, significance is computed in the same way as for the WEAT. Here, $\mathcal{D}_1$ and $\mathcal{D}_2$ refer to the male and female corpus respectively.

Selection of attribute and target words: Table 7 in the Appendix reports the sets of target and attribute words used, borrowed from [11] and [14]. We slightly modified the word sets to account for rare or missing words. For this, we first removed words occurring less than five times in at least one of the three corpora. Then, whenever pairs of attribute (target) sets contain a different number of words, we removed the least frequent words from the larger set until the two sets have the same size. In doing this, we defined the frequency of a word as the minimum of its frequencies in the three corpora under analysis. For the list of male and female proper names, we used the most frequent proper names in the lyrics corpora.

Learning word vectors: We used the Gensim implementation of Word2Vec [45] to learn word embeddings from scratch separately for the three corpora of song lyrics. To account for the potential variability that arises from different initializations of the word vectors, we conducted the association tests on word embeddings obtained from five independent runs of the Word2Vec algorithm. We then report the score obtained from the first iteration and deem the result to be statistically significant only if all five iterations yielded a significant score at a certain level. In other words, we conducted the association tests five times using different initializations to ensure the robustness of our results.
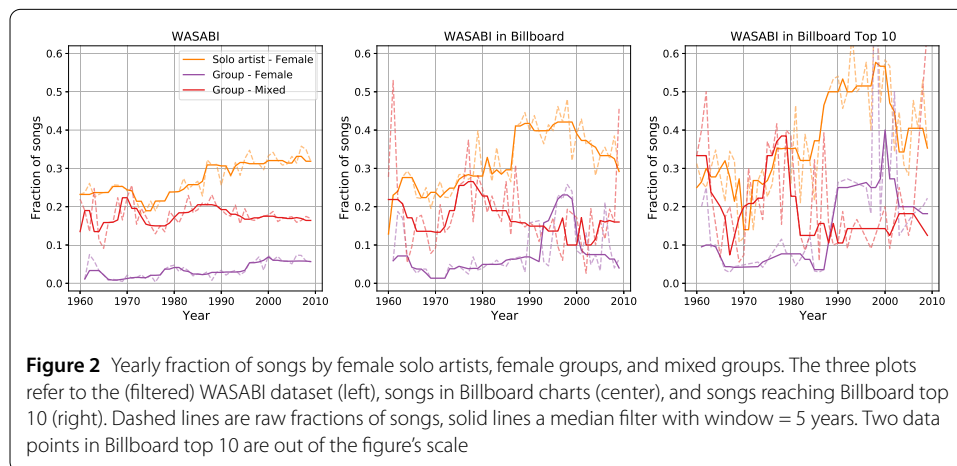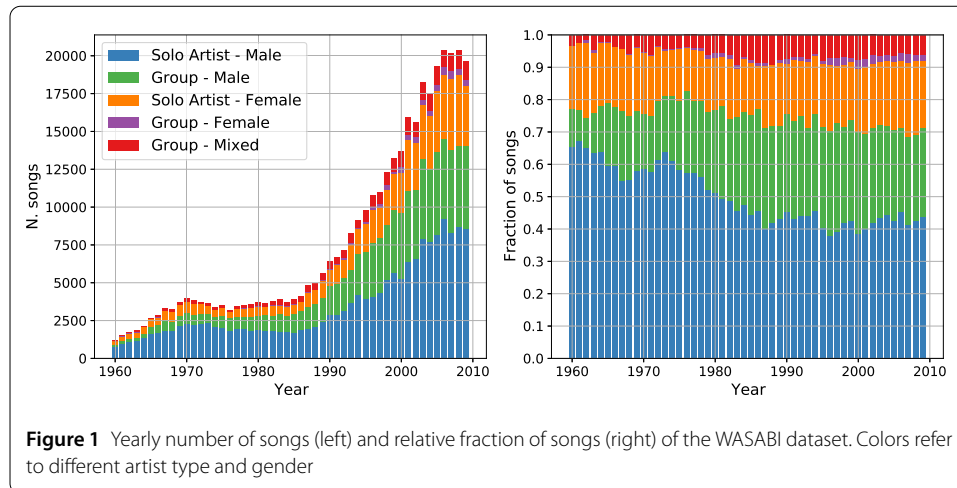
4 Results

4.1 Basic statistics of the dataset

Table 1 reports the number of songs for each combination of artist type and gender. Solo artists are more represented by males than females, with more than the double of the number of songs. This unbalance is exacerbated for groups, where female and mixed groups account for 1.7% and 6.2% of the songs. This observation holds for the songs in Billboard and Billboard top 10 charts as well, where the gender inequality in English song charts is well documented in the literature [46, 47]. The distribution of publication years of the songs in our dataset is shown in the left plot of Fig. 1. The total number of songs per year is

Table 1 Basic statistics of the dataset

Artist type	Gender	Artists	Songs	Billboard	Billboard (top 10)
Solo artist	Male	4770 41.8%	171,324 45.3%	5071 47.0%	1103 42.3%
	Female	2361 20.7%	72,822 19.3%	2412 22.3%	664 25.5%
Group	Male	3281 28.7%	103,729 27.5%	2493 23.1%	629 24.1%
	Female	306 2.7%	6416 1.7%	212 2.0%	56 2.1%
	Mixed	707 6.2%	23,517 6.2%	610 5.6%	156 6.0%
Total		11,425	377,808	10,798	2608

Figure 1 Yearly number of songs (left) and relative fraction of songs (right) of the WASABI dataset. Colors refer to different artist type and gender

Figure 2 Yearly fraction of songs by female solo artists, female groups, and mixed groups. The three plots refer to the (filtered) WASABI dataset (left), songs in Billboard charts (center), and songs reaching Billboard top 10 (right). Dashed lines are raw fractions of songs, solid lines a median filter with window = 5 years. Two data points in Billboard top 10 are out of the figure's scale

not uniformly distributed across time. There are about 3000 songs per year from the 1970s until the 1990s with a subsequent steady increase reaching around 20,000 songs per year in the late 2000s. However, the relative fraction of songs performed by male and female artists remains relatively constant, as can be seen in the right plot of Fig. 1. Indeed, male artists (in groups or as solo artist) contribute with a fraction of songs between 70% and 80% over the whole time span.

Figure 2 shows the increase of the fraction of songs by female solo artists across time in all the subsets of the dataset. In particular, the Billboard and Billboard top 10 charts show a sharp increase between 1980 and 1990 where the fraction of female artist songs increases by around 10%, reaching 40% and 50% respectively. Then, during the 2000s this fraction decreased again to values of the mid-80s. We refer to Appendix A for the fraction of songs segregated by genre across time.

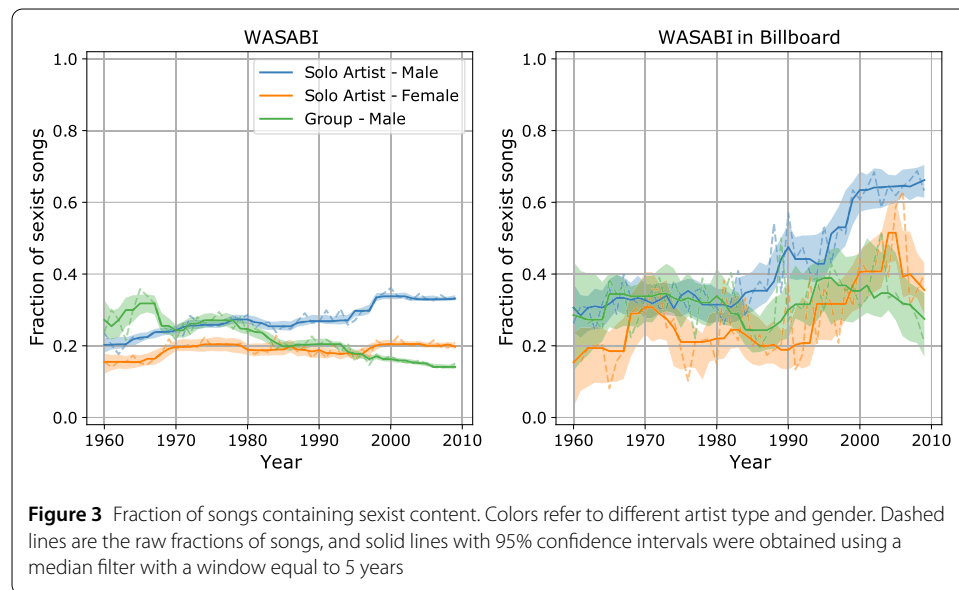## 4.2 Sexist song lyrics

Table 2 shows the proportion of songs containing sexist passages identified for each artist type and gender. The classifier identifies 89,462 (23.7%) lyrics in the WASABI dataset to contain sexist passages. Artists and groups have different fractions of sexist lyrics: 30% of male solo artist songs are classified as sexist, compared to 16% to 20% of songs of the other

Table 2 Percentage of songs containing sexist passages identified for each artist type and gender. Percentages correspond to the fraction of sexist lyrics within the artist type and gender

Artist type	Gender	Songs	Billboard	Billboard (top 10)
Solo artist	Male	30.0%	43.4%	48.4%
	Female	19.6%	29.8%	30.9%
Group	Male only	18.0%	32.0%	33.2%
	Female only	20.0%	34.4%	42.9%
	Mixed	16.2%	26.4%	30.1%

Figure 3 Fraction of songs containing sexist content. Colors refer to different artist type and gender. Dashed lines are the raw fractions of songs, and solid lines with 95% confidence intervals were obtained using a median filter with a window equal to 5 years

groups of artists. The fraction of sexist songs in Billboard and Billboard top 10 charts is at least 10% higher than the one of the whole WASABI dataset. This observation suggests that popular songs are more likely to contain sexist content than an average song, even more so if they reach a top position in the charts. This finding is consistent for different classifications thresholds (details in Appendix D).

 Furthermore, a larger fraction of male solo artist songs contains sexist content when compared to those of female solo artists. We observe that male and female artists display different trends for what concerns their percentage of songs classified as sexist over time, but male solo artists consistently publish relatively more sexist songs in the whole time span. Figure 3 shows the fraction of songs containing sexist passages for each artist type and gender. The share of female solo artist songs in the WASABI dataset with sexist lyrics remains relatively constant over time at around 20%, but, even though the share of male solo artist sexist songs becomes larger going forward in time, this trend is even stronger for popular songs in which we can identify a sharp increase starting around the mid-80s. At the end of the time span under analysis, more than 60% of male solo artist songs on Billboard are found to contain sexist passages. The share of female solo artist songs on Billboard with sexist lyrics increases over time as well, but during the 2000s this share is 20% lower when compared to male solo artists. Differently, group artists, regardless of members' gender, do not display a relevant increase of sexist songs. Figure 3 shows

only the group male artists to not overload the figure, but the female and mixed group artists display a similar trend. Interestingly, male solo artists perform relatively more sexist songs than male group artists. This might signal a difference between male solo artists and groups in terms of subjects and themes covered in their songs. As observed previously, the figure shows also that Billboard charts, compared to the WASABI dataset, attracted more songs with sexist content starting from the 1990s.

After this aggregate analysis we now inspect how different genres contribute to the observed trends.[6] In particular, we focus on the four most popular genres in our data-sets: Pop, Rock, Hip hop, R&B and soul. The changes of their frequency over time in the data set is depicted in Fig. 5 in the Appendix, while Fig. 4 shows the fraction of songs containing sexist passages of male and female solo artists as well as male only groups for Pop, Rock, Hip hop, and R&B and soul.[7] A majority of the Hip hop songs are classified as sexist, regardless of artists' gender. Pop and Rock songs do not show a substantial difference between male and female solo artists, but male artists perform relatively more sexist songs than female artists for the Pop genre. In both cases, these fractions are lower than the Hip hop genre. Lastly, R&B and soul songs display a steady increase of sexist lyrics over time, with male solo artists having a higher fraction of sexist songs than female artists.

4.3 Language bias in song lyrics

Here we investigate the corpora of song lyrics for potential language bias, showing the results of the different word embedding association tests defined in Sect. 3.3: SC-WEAT (measures the difference between two sets of attribute words and a target set), WEAT (measures the relative similarity of the two target sets with two sets of attribute words), and SWEAT (compares the differences for a target set in two corpora, i.e. male and female corpora).
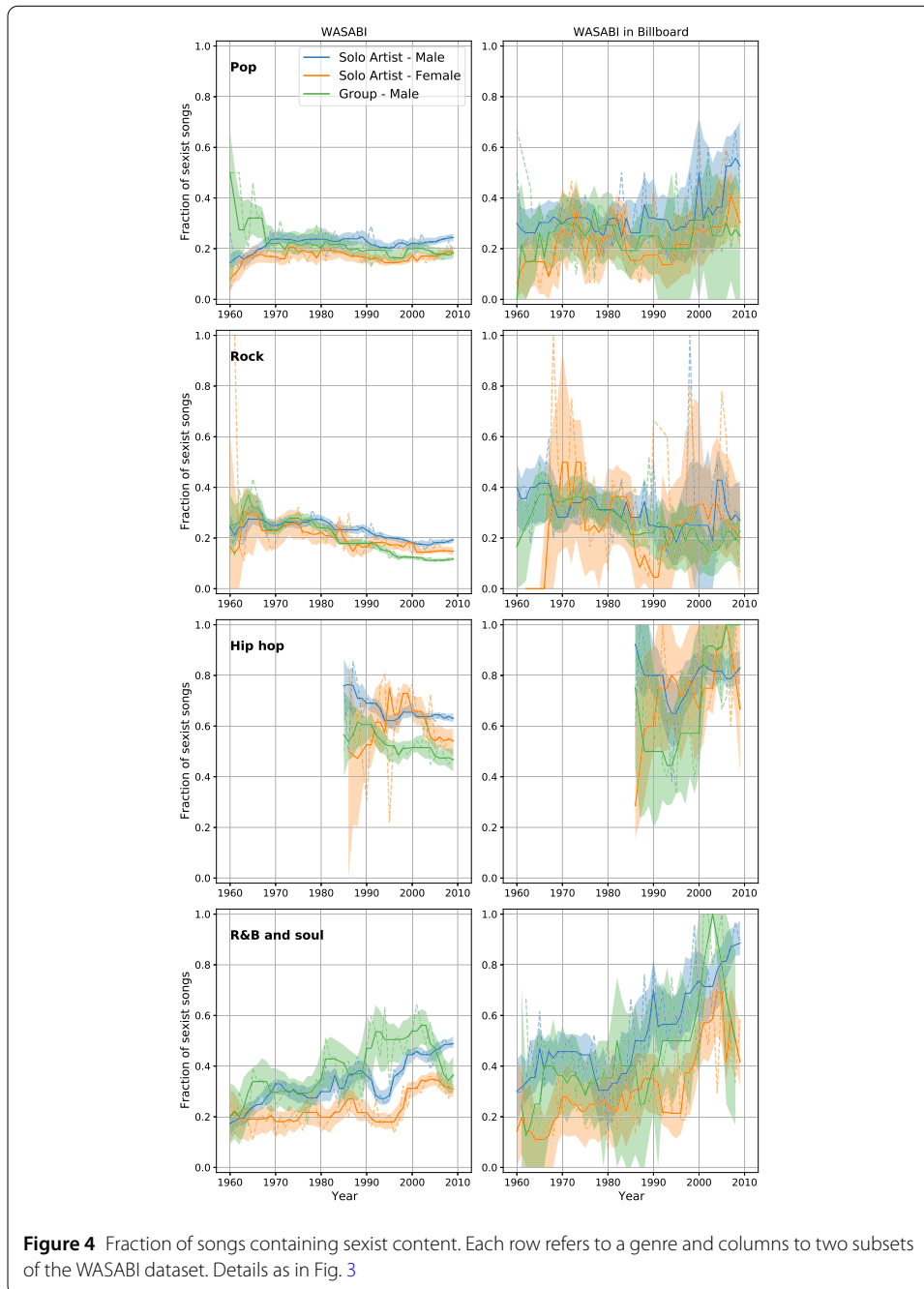
The corresponding results are shown in Table 3 with positive WEAT effect sizes (last column) for all the tests, indicating that words in the target set X (e.g., male names) are more similar to words in A (e.g., career) than B (e.g., family), and words in the target set Y (e.g., female names) are more similar to words in B than A (e.g., family vs career). However, the magnitudes and significance levels of these paired associations are different. We now explain the detailed results grouped by the tested sets of words.

Pleasant vs. Unpleasant words: These associations are considered to be universally accepted stereotypes for humans [48]: pleasant words to flowers and musical instruments, and unpleasant words to insects and weapons. Experiments involving human subjects also found these associations [49]. We thus expect to find the same in song lyrics as well, regardless of the gender of the artist. Indeed, we found all these associations to be statistically significant for both pairs of target sets in all the three corpora (positive and statistically significant WEAT effect size). The fact that the trained word vectors capture these trivial stereotypes makes us confident that the word vectors have learned meaningful associations (see Appendix E for a more detailed discussion of these results). Besides that, we observe a negative and statistically significant SWEAT score for Musical instruments, meaning that Musical instrument words are closer to Unpleasant than Pleasant words in

---

[6]Note that the genre information is missing in 382 (3.5%) and 65 (2.5%) songs in Billboard and Billboard top 10 charts respectively.

[7]Again we omitted female and mixed groups from this analysis.

Figure 4 Fraction of songs containing sexist content. Each row refers to a genre and columns to two subsets of the WASABI dataset. Details as in Fig. 3

the male corpus while being closer to Pleasant than Unpleasant words in the female corpus. This might hint at a deeper difference in the way male and female artists refer to music or the tools (instruments) they use to produce it.

Career vs. Family: We find that Career words are significantly closer to Male than to Female proper names in the male lyrics (SC-WEAT X = 2.76), as well as that Female names are significantly closer to Family than to Career words (SC-WEAT Y = −1.04). This is not significant if we consider nouns and pronouns instead of names and may be explained by song lyrics being more gender-stereotypical when songs address a specific person mentioned by their proper name.

Table 3 Results of the association tests performed on three different lyrics corpora: male solo artists, female solo artists and their union (all). Target sets X or Y resulting in statistically significant SWEAT scores (comparing male with the female corpus) highlighted in bold with their corresponding score (*$p < 0.10$, **$p < 0.05$)

A	B	X	Y	corpus	SC-WEAT X	SC-WEAT Y	WEAT effect size
Pleasant	Unpleasant	Flowers	Insects	all	1.34**	−0.70**	1.73**
				female	1.31**	−0.75**	2.00**
				male	0.90**	−0.78**	1.59**
		Musical instruments* **−0.24**	Weapons	all	0.41	−2.79**	2.97**
				female	0.61**	−1.27**	1.86**
				male	0.23	−1.82**	1.99**
Career	Family	**Male names*** **−0.68**	**Female names*** **−0.92**	all	−0.05	−0.76**	0.89**
				female	0.94	−0.00	0.57
				male	−0.24	−1.04**	0.84**
Male names	Female names	Career	Family	all	2.78**	−0.23	2.16**
				female	0.45	−0.53	0.98
				male	2.76**	−0.60	2.56**
Male nouns & pronouns	Female nouns & pronouns	Career	Family	all	0.64	0.15	0.57
				female	−0.22	−0.34	0.04
				male	1.91	0.35	0.38
		MatSci	Arts	all	0.50	−0.45	0.94
				female	0.06	−1.26**	1.29*
				male	0.39	−0.09	0.37
		Intelligence* **0.18**	Appearance	all	0.29	−0.13	0.39
				female	−0.08	−0.09	0.06
				male	0.42	−0.31	0.71*
		Strength* **0.32**	Weakness	all	1.71**	0.40	1.16**
				female	0.70*	0.24	0.32
				male	1.24**	0.36	1.02**

Interestingly, we do not observe the same bias in female lyrics, where the SC-WEAT score for Female names is close to 0 and we therefore observe a significant difference between the two corpora (SWEAT = −0.92 for Female names). Furthermore, the SC-WEAT score for Career is much smaller (0.45), but this does not translate into a significant numerical difference between male and female corpora. However, a significant difference is also found for Male names (SWEAT = −0.68). Although the corresponding SC-WEAT scores are not significant, we observe a positive value (0.94) in female lyrics and a negative score in male lyrics (−0.24). In other words, Male names are closer to Career than to Family terms in female artist lyrics than in those from male artists.

We observe thus that male solo artists associate Female names closer with Family and female solo artists Male names closer to Career, while the opposite is not the case. There are no biases in female lyrics of female names towards Family terms and in male lyrics of Male names towards Career terms (if anything, it would be in the opposite direction, i.e. a negative SC-WEAT score of −0.24). However, when considering career words alone there is a bias in the male corpus of Career terms towards Male names (SC-WEAT X = 2.76). So Career terms are more likely to be mentioned together with Male names, while if Male names are mentioned there, it is slightly more likely to be in relation to Family words.

MatSci vs. Arts: When analysing a potential gender bias of Mathematics or Science vs. Arts terms we observe that female lyrics show a positive and significant WEAT score (1.29), and a negative and significant single category association for Arts, indi-

cating a higher association of Arts words to Female than to Male nouns and pronouns (SC-WEAT Y = −1.26). None of the other tests is statistically significant for the male and all corpora, nor the comparison of the male and female corpora through SWEAT.

Intelligence vs. Appearance and Strength vs. Weakness: Male solo artists associate Strength words significantly more to Male than Female nouns and pronouns (WEAT = 1.02) with a positive and significant single category association for Strength words (SC-WEAT X = 1.24). Although we also find a similar significant relation for female artists lyrics (SC-WEAT X = 0.70), the direct comparison with the male corpus gives a positive and significant SWEAT score for Strength (0.32) terms. This indicates the association of Strength words to Male terms is stronger in lyrics of male than female solo artists.

Finally, we find a positive and slightly significant WEAT score (0.71) comparing the Intelligence and Appearance words in male lyrics. Although the single category scores are not significant by themselves, their opposite sign shows a signal for male artists to associate Intelligence words with Male terms and Weakness words with Female terms. This is different in female lyrics where the corresponding SC-WEAT scores are close to zero indicating no bias in this regard and translates into a slightly significant result for the SWEAT score for Intelligence (0.18).

To summarise, this analysis shows that songs of male solo artists contain more and often stronger gender biases than those of female solo artists, which are closer to gender neutrality. The only exception of this observation is a bias in female lyrics relating females more closely to art terms. Male artist songs emphasize men as stronger and focused on career at the expense of women depicted as less strong and closer to the family-related terms.

5 Discussion

This section discusses our findings in relation to the research questions and findings of previous works, some potential shortcomings in our study design, and potential paths for future research.

Sexism (RQ1): The first research question investigates the presence of sexist content in song lyrics and whether there is any difference between artist gender, type, and musical genre. We found that almost 25% of song lyrics express sexist content, but this share is not uniformly distributed across artist gender and type. Male solo artists have the highest share of sexist lyrics in all the three subsets of the WASABI dataset (all songs, songs in the Billboards charts, and songs in top 10 position). Interestingly, this share is much lower for groups composed of only male members. This difference is worth mentioning because it might indicate a deeper divergence between songs performed by male solo and group artists that, to our knowledge, has not been reported in the literature previously.

Another observation is that the relative number of lyrics containing sexist passages is higher in Billboard charts than in the whole WASABI dataset independently from the gender and type of the artist, a trend that is stronger going forward in time. Other works found similar observations, even though using different definitions or specific aspects related to sexual content. For example, some studies report an increase in sexual content and objectification of women in popular songs during the last five decades [30, 32]. In a study limited to the year 2009, the top 10 song charts were shown to be more likely to contain sexual content if compared to songs from the same album by the same artist that did not enter the top 10 [50]. According to previous studies, we also found an higher fraction of

songs containing sexist content among Hip hop and R&B and soul songs [33, 51], albeit other genres are not exempted from this trend [52]. Hip hop and R&B and soul were also found to display an increasing trend in associating competence more frequently to men than women [35].

Although we obtained aggregated results in line with findings from previous works, we emphasize that these results rely on a sexism classifier trained on out-of-domain data, which might have led to poor generalization when applied to song lyrics. However, the classifier was trained on a combination of diverse corpora of sexist texts that enforce its ability to generalize on out-of-domain data [19]. We have also validated our model on a dataset of sexist lyrics showing good performances and robust results across different classification thresholds. We believe that this exploratory data analysis can encourage further works aimed at identifying sexism in song lyrics, which may need to pass through extensive manual labeling in order to train a classifier on the specific domain.

Language bias (RQ2): The second research question focuses on the differences of language bias in male and female solo artist song lyrics. Our results extend the ones obtained in [34], where all the WEAT scores are positive in magnitude. We enrich this work in two different ways. First, we measure language biases separately in the subsets of male and female solo artist songs, obtaining all positive WEAT scores as well. Some of these associations (e.g., male with career and female with family) align to the ones observed in a large-scale crawl of the Internet [11], indicating how music reflects societal biases. Second, we enrich the analysis through other association tests, which measure associations against a single target set (SC-WEAT) and between the two corpora (SWEAT). Our results show that the biases affecting these two corpora are different. In particular, male lyrics contain larger gender biases. For instance, Female proper names are more associated to Family than Career words, while Career words are closer to Male than Female names. At the same time, Strength words are more associated to Male than Female terms. The female corpus does not contain these two types of biases. Similarly, other works have found song lyrics of male artists to contain stronger gender biases than female artists, in particular depicting men as more competent than women [35].

In both corpora, we find that Strength words are closer to Male than Female terms but this association is again stronger in male lyrics. A bias that is only present in female lyrics is that female solo artists use more frequently Female terms closer to Art than Mathematics and Science words. Interestingly, in addition to a weak association of Intelligence words with Male terms in the male corpus, there is no significant association between Appearance words and gendered terms. This might be related to the observation that objectification does not target uniquely women, but it is also fairly common for men even though not present to the same extent [33]. Besides these gender biases, our word embeddings can capture expected associations like Pleasant with Flower words and Unpleasant with Insects or Weapons that do not depend on the gender of the artist.

Our results lack of an analysis of gender biases across time and genres as done in Boghrati and Berger [35], and a comparison between solo artists and groups songs. However, these subsets of the dataset would have reduced the training data used to learn word embeddings, thus compromising their quality.

Other limitations concern the dataset. We can not claim any generalizability of our conclusions because there is no guarantee that WASABI contains a representative sample of English songs. Moreover, Billboard charts add an additional US bias in the selection of

popular songs. The choice of the variables used to stratify the results is another limitation. It is worth noticing that we split artists into males and females taking into account only the performer of the song, thus ignoring the gender of the songwriter. In addition, we can consider only two genders, thus neglecting other non-binary gender identities, as only binary gender data is available in the WASABI database. It would be interesting to address these gaps in future works, but this will require additional efforts for accurate data collection. Despite these limitations, the size of WASABI database, together with its open access, makes it a relevant study object by itself, and our contribution may be valuable for future users of this resource.

6 Conclusion

In this work, we have exploited the WASABI database to describe how sexist content in music varies during five decades, from 1960 to 2010, and to what extent word embeddings learned from the song lyrics corpus contain language biases. The former analysis, stratified by artist gender and type, shows that popular song lyrics of male solo artists become more sexist over the years, while this behavior is less noticeable for the other categories of artists. The genres (among the ones analyzed) that have the highest fraction of songs containing sexist content are Hip hop and R&B and soul, independently from the gender of the artist. Regarding language biases, we find the lyrics of male solo artists to contain more gender bias than those of female solo artists. This is true for instance for the stereotype depicting men as stronger and focused on success, and women being closer to family. The former bias is present in female solo artist songs as well, even though the association is weaker than for male solo artists. Our results show different ways to extract meaningful metrics about language usage and bias in song lyrics, as well as how to analyse such an important and heavily consumed expression of popular culture that influences how listeners see the world and reflects how artists perceive it.

Appendix A: Genre popularity

We show here in Fig. 5 the fraction of Rock, Pop, Hip hop and R&B and soul songs across time for the three subsets of the WASABI dataset. The popularity of Rock peaked between 1970 and 1990 and then decreased significantly going forward in time, although it is the most frequent genre in WASABI. Pop songs display a similar decrease during the same period of time in Billboard charts. On the other hand, R&B and soul represents more than
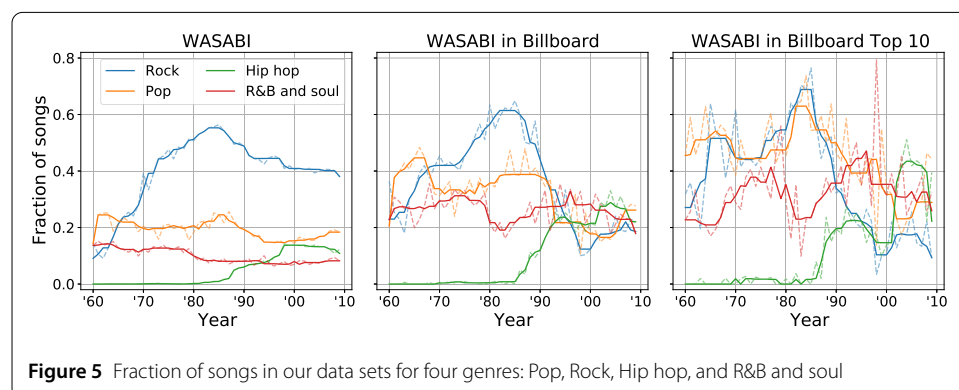
Figure 5 Fraction of songs in our data sets for four genres: Pop, Rock, Hip hop, and R&B and soul

20% of songs on Billboard during the whole time span, while Hip hop starts to flourish during the mid-80s.

Appendix B: Duplicate lyrics detection

The WASABI database was created by collecting the discography of all the artists, including the text of song lyrics when available. This latter information was collected from LyricWiki, an online database of lyrics and encyclopedia curated by users [18]. During this process, the same song lyrics may be collected multiple times (e.g., the same song is present in several albums of the same artist) and it is ideal to reduce as much as possible the presence of duplicates in the dataset. To do that, we used techniques of approximate string matching.

Given a set of song lyrics, we represented each lyrics as the set of 3-grams it is composed of, following a bag-of-words approach. Then, we consider two lyrics to be duplicates if the Jaccard index of their 3-grams representation is higher than 0.80. This threshold allowed us to detect groups of songs whose lyrics are identical up to slight variations. We then consider the song with the earliest publication date as the original song and the other songs of the group as its duplicate.

In case two duplicate songs were performed by two distinct artists, we refer to the song published later as cover song. We removed from the dataset only the duplicate songs but we kept cover versions. We identified 82,531 duplicate songs and 7524 cover songs.

Appendix C: Matching songs from different datasets

Due to the lack of standardized metadata of artist names and song titles, it is not trivial to match entries corresponding to the same artist or song from two datasets. In our case, we needed to match the entries of the WASABI dataset with the entries of the Billboard charts and the sexist lyrics dataset. We used hand crafted logics of approximate string matching to first match the artist name, and then the title of the song. We used Python script from the Million Song Dataset repository,[8] which we only modified slightly.

Appendix D: Sexism classifier on song lyrics

We trained the classifier using the scripts provided by the authors of [19] and stored in the official repository of the paper.[9] The pre-trained model is the base version of the uncased BERT,[10] fine tuned for 3 epochs with batch size equal to 32, learning rate at 2×10^{-5} and 10% of warmup steps.

Since the classifier we used to detect sexist lyrics was trained on short texts, we adapted the representation of the input to classify song lyrics. In detail, lyrics were divided into batches composed of groups of four lines, each of them sharing two lines with the previous and following group. A batch is classified as sexist if the model outputs a probability higher than a certain threshold. Then, lyrics are classified as sexist whenever the model predicts
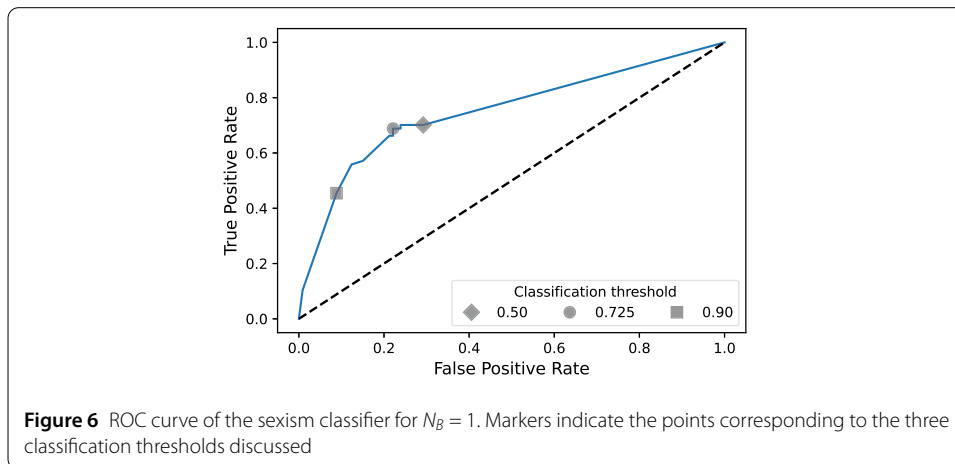
---

[8]https://github.com/tbertinmahieux/MSongsDB/blob/master/NameNormalizer/normalizer.py

[9]https://github.com/gesiscss/theory-driven-sexism-detection

[10]https://tfhub.dev/google/bert_uncased_L-12_H-768_A-12/1

Table 4 Performance of the sexism classifier on the external dataset for three classification thresholds and $N_B = 1$. Metrics for both classes (Sexist and Not sexist) and the corresponding macro average are shown. Right column shows the performance of a naive baseline that always predicts the sexist class

Metric	Class	Classif. threshold			Baseline
		0.50	0.725	0.90	
Precision	Sexist	0.62	0.68	0.78	0.41
	Non-sexist	0.78	0.79	0.71	0.00
	Macro avg.	0.70	0.73	0.74	0.20
Recall	Sexist	0.70	0.69	0.45	1.00
	Non-sexist	0.71	0.78	0.91	0.00
	Macro avg.	0.70	0.73	0.68	0.50
F1-score	Sexist	0.66	0.68	0.57	0.58
	Non-sexist	0.74	0.78	0.80	0.00
	Macro avg.	0.70	0.73	0.69	0.29

Figure 6 ROC curve of the sexism classifier for $N_B = 1$. Markers indicate the points corresponding to the three classification thresholds discussed
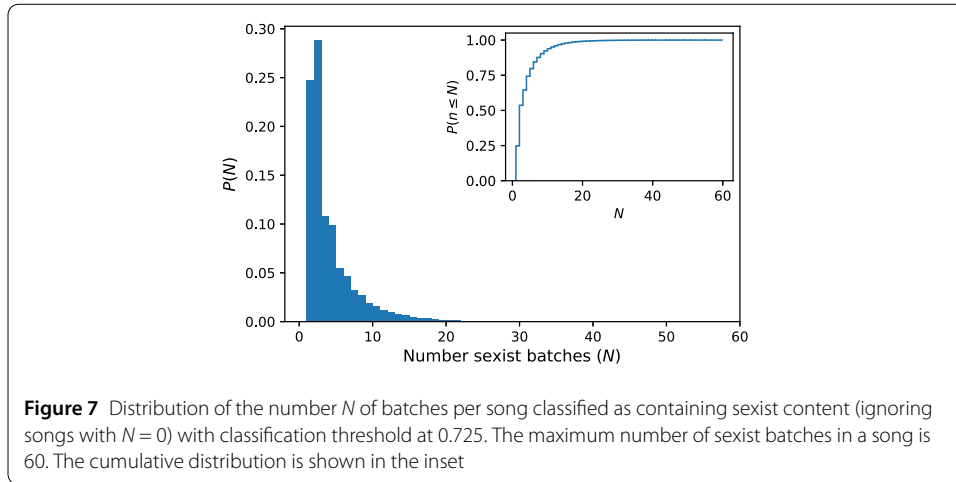
that at least N_B batches contain sexist content. In the main text, we considered $N_B = 1$, meaning that one batch of lines labeled as sexist is enough to propagate the label to the whole song. The optimal classification threshold is chosen as the threshold maximizing the F1-score on the external dataset [42] of sexist songs in order to balance both precision and recall. This corresponds to a threshold equal to 0.725.

Table 4 shows the classification scores on the external dataset and the performance of a naive baseline model that predicts all lyrics as sexist. Besides the optimal classification threshold, we also considered the results for classification thresholds at 0.50 and 0.90, where the latter favors precision on the sexist class over recall, and minimises false positives. Indeed, with the 0.90 threshold the precision for the sexist class reaches 0.78 while the macro F1-score is 0.69 and the recall drops to 0.45. At the same time, the performance on the non-sexist class reaches a recall of 0.91 with a precision of 0.71.

The corresponding ROC curve is shown in Fig. 6. We also note that the F1-score on the sexist class is stable for other choice of N_B ($N_B = 2$: F1-score at 0.69 for an optimal threshold of 0.575; $N_B = 3$: F1-score at 0.65 for an optimal threshold of 0.525).

Now we come back to discuss the results of the sexism classifier on the WASABI dataset. Figure 7 shows the distribution of sexist batches for songs classified as sexist according to

Figure 7 Distribution of the number N of batches per song classified as containing sexist content (ignoring songs with $N = 0$) with classification threshold at 0.725. The maximum number of sexist batches in a song is 60. The cumulative distribution is shown in the inset

Table 5 Percentage of sexist songs identified for each artist type and gender. Results for three different classification thresholds (0.50, 0.725 and 0.90) and $N_B = 1$ are reported, with results for the optimal threshold in bold. Percentages correspond to the fraction of sexist lyrics within the artist type and gender

Artist type	Gender	Classif. thresh.	Songs	Billboard	Billboard (top 10)
Solo artist	Male	0.50	36.7%	50.0%	54.1%
		0.725	**30.0%**	**43.4%**	**48.4%**
		0.90	11.9%	22.0%	28.0%
	Female	0.50	25.0%	35.7%	36.9%
		0.725	**19.6%**	**29.8%**	**30.9%**
		0.90	7.8%	14.6%	16.0%
Group	Male only	0.50	23.2%	39.1%	39.7%
		0.725	**18.0%**	**32.0%**	**33.2%**
		0.90	6.3%	13.6%	14.8%
	Female only	0.50	25.2%	42.5%	50.0%
		0.725	**20.0%**	**34.4%**	**42.9%**
		0.90	8.3%	14.6%	21.4%
	Mixed	0.50	21.5%	32.8%	37.2%
		0.725	**16.2%**	**26.4%**	**30.1%**
		0.90	5.2%	13.3%	16.0%

the optimal classification threshold and $N_B = 1$. Almost half of the lyrics have more than 2 sexist batches, while 25% and 29% of song lyrics have 1 and 2 sexist batches respectively.

To verify the independence of the conclusions we extracted from the optimal classification threshold, we report the fraction of sexist lyrics for all the three thresholds on Table 5. Here, we can observe for all the thresholds the same trend showing that (i) male solo artists performs more sexist songs compared to the other groups, and (ii) Billboard charts contain a higher fraction of sexist songs than the whole WASABI dataset, regardless of the gender and type of the performer. Table 6 shows that these considerations hold for $N_B = 2$ and $N_B = 3$ as well.

Table 6 Percentage of sexist songs identified for each artist type and gender. We report how these percentages change for different values of N_B (i.e. the minimum number of 4 line batches classified as sexist to consider the whole song to contain sexist content). The classification threshold = 0.725 and values in bold are used in the main text. Percentages correspond to the fraction of sexist lyrics within the artist type and gender

Artist type	Gender	N_B	Songs	Billboard	Billboard (top 10)
Solo artist	Male	**1**	**30.0%**	**43.4%**	**48.4%**
		2	23.1%	36.7%	42.1%
		3	14.9%	25.5%	32.8%
	Female	**1**	**19.6%**	**29.8%**	**30.9%**
		2	14.7%	23.8%	25.5%
		3	8.6%	14.9%	16.9%
Group	Male only	**1**	**18.0%**	**32.0%**	**33.2%**
		2	12.9%	24.6%	26.2%
		3	7.4%	14.2%	16.7%
	Female only	**1**	**20.0%**	**34.4%**	**42.9%**
		2	15.0%	26.4%	37.5%
		3	8.4%	14.2%	17.9%
	Mixed	**1**	**16.2%**	**26.4%**	**30.1%**
		2	11.2%	20.8%	25.6%
		3	6.3%	13.8%	17.9%

Appendix E: Word embedding association tests

Table 7 shows the lists of words corresponding to each set of target and attribute words, and we report the references from which each word set was borrowed on the right column. We applied slight modifications to some of those sets to take into account rare or missing words. First, we removed words if they occur less than 5 times in one of the three corpora. Then, whenever pairs of attribute (target) sets contain a different amount of words, we remove the least frequent words from the larger until the two sets have the same size. In doing this, we define the frequency of a word as the minimum of the frequencies among the three corpora under analysis. We used a different procedure to select proper names. We downloaded the yearly counts of names of newborn babies from 1879 until 2020,[11] and searched for them in song lyrics. The male and female name word sets are thus composed of the most frequent names in the lyrics corpora.

Learning word vectors: We used the Gensim implementation of Word2Vec [45] to learn word embeddings with window length 5, embedding dimension 300, and 40 training steps. Words occurring less than 5 times were discarded.

Detailed results on Pleasant vs. Unpleasant words: Results are shown in Table 3. The coupled association measured by the WEAT returns positive and statistically significant scores for all the three corpora and both pairs of target sets, namely Flowers/Insects and Musical instruments/Weapons. The single category associations (SC-WEAT) of the former are significant as well, indicating that, if kept separately, Flower words are more associated to Pleasant than Unpleasant words and vice versa for Insect words. For the latter, the only female corpus shows a significant single category association between Musical instrument and Pleasant words. This reflects in the comparison between the male and female corpus through the SWEAT that returns a negative and statistically significant score, i.e., Musical instrument words are closer to Unpleasant than Pleasant words in the male

---

[11]https://www.ssa.gov/oact/babynames/limits.html (accessed 12/01/2022)

Table 7 Word sets used for the word embedding association tests. The right column reports the references from which the sets were borrowed

Word set name	Words	Ref.
Pleasant	family, honest, gift, wonderful, vacation, miracle, loyal, pleasure, gentle, rainbow, love, peace, lucky, honor, freedom, happy, health, friend, laughter, cheer, joy, heaven, diploma, paradise, diamond, caress, sunrise	[11]
Unpleasant	cancer, agony, stink, sickness, poverty, accident, failure, rotten, hatred, terrible, disaster, tragedy, grief, jail, abuse, awful, prison, ugly, nasty, murder, bomb, poison, evil, crash, death, war, kill	[11]
Career	corporation, professional, career, office, business	[11, 14]
Family	family, marriage, wedding, children, home	[11, 14]
Female attributes	girl, hers, her, aunt, daughter, sister, female, mother, she, grandmother, woman	[11, 14]
Male attributes	brother, grandfather, his, son, father, man, male, uncle, he, him, boy	[11, 14]
Female names	kim, rose, mary, eve, kelly, jane, lisa, juliet, jean, annie, trina, sarah, sally, betty, lucy, taylor, bonnie, marie, jenny, dolly, julia, anna, jill, angie	*
Male names	john, jack, joe, johnny, james, david, paul, billy, jimmy, simon, mark, romeo, bill, peter, bob, lee, jim, bobby, tom, jackson, sam, michael, charlie, adam	*
Flowers	lilac, bluebell, violet, crocus, buttercup, iris, rose, tulip, daisy, marigold, daffodil, orchid, carnation, magnolia, lily, poppy, clover	[11]
Insects	cockroach, maggot, locust, roach, centipede, caterpillar, weevil, beetle, flea, dragonfly, mosquito, ant, cricket, moth, spider, bee, fly	[11]
Musical instruments	banjo, mandolin, trombone, cello, fiddle, tuba, harmonica, harp, violin, piano, trumpet, clarinet, oboe, guitar, lute, saxophone, horn, bongo, flute, bell, viola, drum	[11]
Weapons	harpoon, mace, hatchet, grenade, missile, spear, axe, rifle, cannon, dagger, pistol, shotgun, dynamite, tank, blade, sword, arrow, whip, bomb, knife, club, gun	[11]
MatSci	nasa, addition, einstein, technology, experiment, math, chemistry, science	[11, 14]
Arts words	poetry, novel, symphony, art, dance, shakespeare, sculpture, drama	[11, 14]
Intelligence	brilliant, logical, apt, smart, thoughtful, wise, precocious, genius, intelligent, shrewd, clever	[14]
Appearance	gorgeous, slim, healthy, handsome, ugly, fat, thin, weak, beautiful, pretty, strong	[14]
Strength	triumph, confident, potent, loud, winner, shout, succeed, strong, bold, leader, dynamic, command, power	[14]
Weakness	timid, withdraw, yield, failure, fragile, weakness, loser, shy, surrender, weak, afraid, follow, lose	[14]

* Male and female names were chosen from the most frequent names found in the solo artists songs of the WASABI dataset.

corpus while being closer to Pleasant than Unpleasant words in the female corpus. The fact that the learned word embeddings encode these associations makes us confident of the quality of the learned word vectors.

Abbreviations
NLP, Natural Language Processing; RQ, Research question; WEAT, Word Embedding Association Test; SC-WEAT, Single Category Word Embedding Association Test; SWEAT, Sliced Word Embedding Association Test; BERT, Bidirectional Encoder Representations from Transformers.

Availability of data and materials
Instructions to download the dataset analysed during the current study and code produced for the analysis are available in the author's GitHub repository: https://github.com/Loreb92/sexism_and_bias_in_song_lyrics.

Declarations

Author contributions
LB run the analysis. AK conceptualized the problem and collected the data. All authors contributed to interpret the results and to the writing of the manuscript. All authors read and approved the final manuscript.

Author details
[1]ISI Foundation, Via Chisola 5, 10126 Turin, Italy. [2]Department of Network and Data Science, Central European University, Quellenstraße 51-55, 1100 Vienna, Austria. [3]CENTAI, Corso Inghilterra 3, 10138 Turin, Italy. [4]Sapienza University, Piazzale Aldo Moro 5, 00185 Rome, Italy. [5]Universitat Pompeu Fabra, Tanger 122, 08018 Barcelona, Catalonia, Spain.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Ransom PF (2015) Message in the music: do lyrics influence well-being? Master's thesis, University of Pennsylvania
2. Cobb MD, Boettcher WA III (2007) Ambivalent sexism and misogynistic rap music: does exposure to eminem increase sexism? J Appl Soc Psychol 37(12):3025–3042
3. Treat TA, Farris CA, Viken RJ, Smith JR (2015) Influence of sexually degrading music on men's perceptions of women's dating-relevant cues. Appl Cogn Psychol 29(1):135–141
4. Adams TM, Fuller DB (2006) The words have changed but the ideology remains the same: misogynistic lyrics in rap music. J Black Stud 36(6):938–957
5. Davis S (1985) Pop lyrics: a mirror and a molder of society. ETC Rev Gen Semant 42(2):167–169
6. Miranda ER, Yeung R, Pearson A, Meichanetzidis K, Coecke B (2021) A quantum natural language processing approach to musical intelligence. arXiv:2111.06741
7. Hovy D, Prabhumoye S (2021) Five sources of bias in natural language processing. Lang Linguist Compass 15(8):12432. https://doi.org/10.1111/lnc3.12432
8. Abid A, Farooqi M, Zou J (2021) Large language models associate muslims with violence. Nat Mach Intell 3(6):461–463. https://doi.org/10.1038/s42256-021-00359-2
9. Shah DS, Schwartz HA, Hovy D (2020) Predictive biases in natural language processing models: a conceptual framework and overview. In: Proceedings of the 58th annual meeting of the association for computational linguistics. Association for Computational Linguistics, Cedarville, pp 5248–5264. https://aclanthology.org/2020.acl-main.468. https://doi.org/10.18653/v1/2020.acl-main.468
10. Bengio Y, Ducharme R, Vincent P (2000) A neural probabilistic language model. In: Leen T, Dietterich T, Tresp V (eds) Advances in neural information processing systems, vol 13. MIT Press, Cambridge. https://proceedings.neurips.cc/paper_files/paper/2000/file/728f206c2a01bf572b5940d7d9a8fa4c-Paper.pdf
11. Caliskan A, Bryson JJ, Narayanan A (2017) Semantics derived automatically from language corpora contain human-like biases. Science 356(6334):183–186
12. Bolukbasi T, Chang K-W, Zou JY, Saligrama V, Kalai AT (2016) Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. In: Advances in neural information processing systems, vol 29
13. Garg N, Schiebinger L, Jurafsky D, Zou J (2018) Word embeddings quantify 100 years of gender and ethnic stereotypes. Proc Natl Acad Sci 115(16):3635–3644. https://doi.org/10.1073/pnas.1720347115
14. Chaloner K, Maldonado A (2019) Measuring gender bias in word embeddings across domains and discovering new gender bias word categories. In: Proceedings of the first workshop on gender bias in natural language processing. Association for Computational Linguistics, Florence, pp 25–32. https://aclanthology.org/W19-3804. https://doi.org/10.18653/v1/W19-3804
15. Babaeianjelodar M, Lorenz S, Gordon J, Matthews J, Freitag E (2020) Quantifying gender bias in different corpora. In: Companion proceedings of the web conference 2020. WWW '20. Association for Computing Machinery, New York, pp 752–759. https://doi.org/10.1145/3366424.3383559
16. Shushkevich E, Cardiff J, Shushkevich E, Cardiff J (2019) Automatic misogyny detection in social media: a survey. Comput Sist 23(4):1159–1164. https://doi.org/10.13053/cys-23-4-3299
17. Jahan MS, Oussalah M (2021) A systematic review of hate speech automatic detection using natural language processing. arXiv:2106.00742
18. Meseguer-Brocal G, Peeters G, Pellerin G, Buffa M, Cabrio E, Faron Zucker C, Giboin A, Mirbel I, Hennequin R, Moussallam M, Piccoli F, Fillon T (2017) WASABI: a two million song database project with audio and cultural metadata plus WebAudio enhanced client applications. In: Web audio conference 2017—collaborative audio #WAC2017, London, United Kingdom. Queen Mary University of London. https://hal.univ-cotedazur.fr/hal-01589250
19. Samory M, Sen I, Kohne J, Flöck F, Wagner C (2021) "call me sexist, but…": revisiting sexism detection using psychological scales and adversarial samples. In: Proceedings of the international AAAI conference on web and social media, vol 15, pp 573–584
20. Charlesworth TES, Yang V, Mann TC, Kurdi B, Banaji MR (2021) Gender stereotypes in natural language: word embeddings show robust consistency across child and adult language corpora of more than 65 million words. Psychol Sci 32(2):218–240. PMID: 33400629. https://doi.org/10.1177/0956797620963619
21. Bianchi F, Marelli M, Nicoli P, Palmonari M (2021) SWEAT: scoring polarization of topics across different corpora. In: Proceedings of the 2021 conference on empirical methods in natural language processing, pp 10065–10072. Association for Computational Linguistics, Punta Cana. https://aclanthology.org/2021.emnlp-main.788. https://doi.org/10.18653/v1/2021.emnlp-main.788

22. Stanczak K, Augenstein I (2021) A survey on gender bias in natural language processing. arXiv:2112.14168
23. Sun T, Gaut A, Tang S, Huang Y, ElSherief M, Zhao J, Mirza D, Belding E, Chang K-W, Wang WY (2019) Mitigating gender bias in natural language processing: literature review. In: Proceedings of the 57th annual meeting of the association for computational linguistics, pp 1630–1640
24. Ethayarajh K, Duvenaud D, Hirst G (2019) Understanding undesirable word embedding associations. In: Proceedings of the 57th annual meeting of the association for computational linguistics. Association for Computational Linguistics, Florence, pp 1696–1705. https://aclanthology.org/P19-1166. https://doi.org/10.18653/v1/P19-1166
25. Nadeem M, Bethke A, Reddy S (2021) StereoSet: measuring stereotypical bias in pretrained language models. In: Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing (volume 1: long papers). Association for Computational Linguistics, Cedarville, pp 5356–5371. https://aclanthology.org/2021.acl-long.416. https://doi.org/10.18653/v1/2021.acl-long.416
26. Manne K (2017) Down girl: the logic of misogyny. Oxford University Press, London
27. Jha A, Mamidi R (2017) When does a compliment become sexist? Analysis and classification of ambivalent sexism using Twitter data. In: Proceedings of the second workshop on NLP and computational social science, pp 7–16
28. Samghabadi NS, Patwa P, Pykl S, Mukherjee P, Das A, Solorio T (2020) Aggression and misogyny detection using bert: a multi-task approach. In: Proceedings of the second workshop on trolling, aggression and cyberbullying, pp 126–131
29. Pamungkas EW, Basile V, Patti V (2020) Misogyny detection in Twitter: a multilingual and cross-domain study. Inf Process Manag 57(6):102360
30. Madanikia Y, Bartholomew K (2014) Themes of lust and love in popular music lyrics from 1971 to 2011. SAGE Open 4(3):2158244014547179. https://doi.org/10.1177/2158244014547179
31. Hall PC, West JH, Hill S (2012) Sexualization in lyrics of popular music from 1959 to 2009: implications for sexuality educators. Sex Cult 16(2):103–117
32. Smiler AP, Shewmaker JW, Hearon B (2017) From "I want to hold your hand" to "promiscuous": sexual stereotypes in popular music lyrics, 1960–2008. Sex Cult 21(4):1083–1105
33. Flynn MA, Craig CM, Anderson CN, Holody KJ (2016) Objectification in popular music lyrics: an examination of gender and genre differences. Sex Roles 75(3):164–176
34. Barman MP, Awekar A, Kothari S (2019) Decoding the style and bias of song lyrics. In: Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval, pp 1165–1168
35. Boghrati R, Berger J (2022) Quantifying gender bias in consumer culture. arXiv:2201.03173
36. Hu X, Downie JS, Ehmann AF (2009) Lyric text mining in music mood classification. In: 10th international society for music information retrieval conference, ISMIR 2009, pp 411–416
37. Xia Y, Wang L, Wong K-F (2008) Sentiment vector space model for lyric-based song sentiment classification. Int J Comput Proces Lang 21(04):309–330
38. Mayer R, Rauber A (2011) Musical genre classification by ensembles of audio and lyrics features. In: Proceedings of international conference on music information retrieval, pp 675–680
39. Martin-Gutierrez D, Peñaloza GH, Belmonte-Hernandez A, García FÁ (2020) A multimodal end-to-end deep learning architecture for music popularity prediction. IEEE Access 8:39361–39374
40. Barman MP, Dahekar K, Anshuman A, Awekar A (2019) It's only words and words are all I have. In: European conference on information retrieval. Springer, Berlin, pp 30–36
41. Billboard Hot weekly charts. https://data.world/kcmillersean/billboard-hot-100-1958-2017. Accessed 18 Nov 2020
42. Slim K, Parmentier A, Piccardi T Feminism vs. sexism: a portrait of women in recent music. https://github.com/axnyang/CS401. Accessed 18 Nov 2020
43. Mikolov T, Chen K, Corrado G, Dean J (2013) Efficient estimation of word representations in vector space. arXiv:1301.3781
44. Cohen J (2013) Statistical power analysis for the behavioral sciences. Academic Press, San Diego
45. Rehurek R, Sojka P (2011) Gensim—Python framework for vector space modelling. NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic 3(2)
46. Lafrance M, Worcester L, Burns L (2011) Gender and the billboard top 40 charts between 1997 and 2007. Pop Music Soc 34(5):557–570. https://doi.org/10.1080/03007766.2010.522827
47. Anglada-Tort M, Krause AE, North AC (2021) Popular music lyrics and musicians' gender over time: a computational approach. Psychol Music 49(3):426–444. https://doi.org/10.1177/0305735619871602
48. Guo W, Caliskan A (2021) Detecting emergent intersectional biases: contextualized word embeddings contain a distribution of human-like biases. In: Proceedings of the 2021 AAAI/ACM conference on AI, ethics, and society. AIES '21. Association for Computing Machinery, New York, pp 122–133. https://doi.org/10.1145/3461702.3462536
49. Greenwald AG, McGhee DE, Schwartz JL (1998) Measuring individual differences in implicit cognition: the implicit association test. J Pers Soc Psychol 74(6):1464–1480
50. Hobbs DR, Gallup GG Jr (2011) Songs as a medium for embedded reproductive messages. Evol Psychol 9(3):147470491100900309
51. Hart CB, Day G (2020) A linguistic analysis of sexual content and emotive language in contemporary music genres. Sex Cult 24(3):516–531. https://doi.org/10.1007/s12119-019-09645-z
52. Neff S (2014) Sexism across musical genres: a comparison. Honors thesis, Western Michigan University