

Rösch, Jürgen; Rauch, Maxi-Josephine

Article — Published Version

'Do songs tell stories?' An empirical analysis of the effect of emotional arcs on success in a national song contest

Journal of Cultural Economics

Provided in Cooperation with:

Springer Nature

Suggested Citation: Rösch, Jürgen; Rauch, Maxi-Josephine (2025) : 'Do songs tell stories?' An empirical analysis of the effect of emotional arcs on success in a national song contest, Journal of Cultural Economics, ISSN 1573-6997, Springer US, New York, NY, Vol. 49, Iss. 3, pp. 603-637, <https://doi.org/10.1007/s10824-025-09532-6>

This Version is available at:

<https://hdl.handle.net/10419/330635>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>



‘Do songs tell stories?’ An empirical analysis of the effect of emotional arcs on success in a national song contest

Jürgen Rösch¹ · Maxi-Josephine Rauch¹

Received: 23 April 2024 / Accepted: 7 January 2025 / Published online: 18 February 2025
© The Author(s) 2025

Abstract

This article explores whether the lyrics of songs influence their success in a national song contest. We constructed emotional arcs from over 6.763 song lyrics and examined their impact on contest performance. Our findings reveal that these emotional arcs echo established behavioral phenomena such as the ‘paradox of negative emotion’ and the ‘peak-end rule’. Typically, songs characterized by sorrowful plots and endings receive higher scores than those with more joyful emotional themes. Furthermore, such songs are more likely to be awarded points initially. Despite this, contestants often choose songs with positive endings for their performances. To isolate the effects of lyrics from other song characteristics, we utilized Spotify metrics, including danceability, energy, and tempo. This methodology allows us to differentiate the influence of musical and lyrical elements on listener decision-making. This study contributes to the growing literature on the behavioral aspects of music and its impact on decision-making.

Keywords Music contest · Cultural economics · Econometric analysis · Success factors · Narrative structure · Spotify data

JEL classifications Z10 · Z11 · D12 · D70

1 Introduction

Do songs tell stories? If so, does it matter? Throughout history, music has served as a medium for expressing emotions and performing religious rituals. The word ‘Bhagavadgita,’ the name of a significant Hindu scripture, translates to ‘song of god.’ Operas and musicals are renowned for employing music to narrate compelling stories. In 2016, Bob Dylan became the first musician to be awarded the Nobel Prize in Literature ‘for creating new poetic expressions within the great American

✉ Jürgen Rösch
Juergen.roesch@uni-weimar.de

¹ Faculty Media, Bauhaus-Universität Weimar, Albrecht-Dürer-Str. 2, 99425 Weimar, Germany

song tradition.’¹ Contemporary genres such as rap are often at the center of debates regarding explicit content, leading to parental guidance for listening, but these songs also serve as narratives that reflect societal issues. Similarly, children’s songs use catchy lyrics to entertain and teach important life lessons, such as washing your hands or ensuring safety when crossing the street.

Despite these examples, the narrative capacity of songs remains a subject of debate. Pannese et al. (2016) saw the problem in terms of the multiple dimensions that influence the reception of music, ranging from physical perception to ‘cognitive and contextual contributions.’ Nattiez (1990) opposes the view that music is also a narrative and calls ‘any description of its [music] formal structures in terms of narrativity is nothing but a superfluous metaphor’ (Nattiez, 1990, p. 257). Hamilton and Pearce (2024) found that popular melodies have become less complex and more note dense, especially since the year 2000. They find a pattern that suggests a trend toward simpler, more repetitive melodies, potentially driven by changes in music production technology and shifting cultural preferences. Parada-Cabaleiro et al. (2024) found that lyrics have become simpler, with decreased lexical and structural complexity and increased repetitiveness. Sentiment analysis reveals that lyrics have grown more negative and personal over time. The lyrics reception is also genre-specific: Rock fans prefer lyrics from older songs, while country fans show more interest in newer song lyrics. The authors also claim that ‘the role of lyrics has been understudied’ (Parada-Cabaleiro et al., 2024, p. 11) and call for further research regarding the interplay of cultural artifacts and society.

This paper contributes to this debate and is, to our knowledge, the first to investigate whether song lyrics can influence listeners’ decision-making and behavior. To quantify the narrative nature of songs, we use a method introduced by Reagan et al. (2016). The authors used a methodology based on sentiment analysis to measure the narrative arcs of books. The method was subsequently adapted for the analysis of movies (Del Vecchio et al., 2021). The method enables categorizing texts based on their emotional evolution over their length.

Consequently, it provides an approach for analyzing and categorizing songs beyond traditional measures and sentiment analysis. Additionally, the tool enables the identification of emotional peaks during a song and categorizes its ending as happy or sad. We complement this analysis with more classical sentiment approaches to show the robustness of the results.

Moreover, additional data sources from companies like Spotify have become available, enriching the toolkit available to researchers for evaluating various features of songs and their impact on decision-making. Spotify provides a range of audio features such as *danceability*, *energy*, or *instrumentalness*. These features enable the objective assessment of songs, thereby facilitating their comparison and clustering to unveil patterns in music preferences. These objective measurements allow us to disentangle the effect of features related to the music (as measured by the Spotify algorithm) and the narrative structure of the lyrics.

¹ <https://www.nobelprize.org/prizes/literature/2016/press-release/>.

To investigate the influence of the two new above-mentioned data sources, we analyzed decision-making and music, which has recently gained popularity (Anglada-Tort et al., 2023). We take the Bundesvision Song Contest (BSC) as we need a setting where it can be reasonably assumed that lyrics are understood right away: The BSC was the German version of the Eurovision Song Contest (ESC) and took place from 2005 to 2015 to promote German music and German talent. Every German state (e.g., Bavaria, Berlin, Saxony, etc.) sent one song each year to represent the federal state. After the performance, the winner was determined by votes from the audience by telephone or short message services (SMS). Like in the ESC, the points for each song were presented separately for each state on a scale of zero to twelve. The show was hosted by the TV station ProSieben and supported by local radio stations in each German state.

The BSC is an ideal framework to test the influence of our two new datasets. First, existing research provides a benchmark for our results and enables us to compare our findings with previous ones. Second, we can build on the existing methodology to compare the results. Third, for our study, viewers and listeners must understand the lyrics and their emotional connotations. As popular songs are often in English, the meaning of the lyrics might not be directly accessible to every listener.

The idea of the BSC is to present songs in which the lyrics can be understood and where the lyrics play a major role. Hence, our goal is to separate the effect of the music with the Spotify audio features and the impact of the lyrics with emotional arcs (Reagan et al., 2016) and test their influence on the audience's decision regarding the awarded points. This also contributes to the literature on narrative economics and how narratives can influence decision-making or the behavior of economic actors (Shiller, 2020).

This paper aims to synthesize research on decision-making in the context of music (Anglada-Tort et al., 2023; Budzinski et al., 2021; Oxoby, 2009), thereby combining it with an analysis of the Spotify audio features (e.g., Sciandra & Spera, 2022) as well as the emotional arcs previously applied to books (Reagan et al., 2016) and movies (Del Vecchio et al., 2021). The primary objective is to test if the lyrics of songs contribute to explaining decision-making in song contests. Therefore, we also shed light on the behavioral aspects of music, like the peak-end rule or the reaction to different emotional plot archetypes. However, it should be noted that the study of song success factors is also viewed critically, as the challenges in analyzing musical emotionality lie in the ambiguity of its interpretation and the unpredictability of song success (Pannese et al., 2016).

The remainder of the paper is structured in the following manner: First, we provide an overview of the related literature, particularly research on song contests, the utilization of Spotify's audio feature data, emotional arcs, general behavioral aspects of music, and narrative economics. Second, we introduce two new datasets that enhance the existing literature—Spotify's audio features and the derived emotional archetypes of song lyrics. Third, we present our results, employing a methodology introduced by Budzinski et al. (2021). After this, we perform robustness checks to show the impact of other variables and approaches on our results. Lastly, we discuss the results and limitations before concluding with final remarks.

2 Related literature

Initial discussions regarding the effect of music on emotions can be traced as far back as ancient religious texts in India or Persia. Both Plato and Aristotle mirror the ancient Greek belief that music positively influenced health and could ‘promote the development of virtues such as calm, balance, harmony, and order’ (Pannese et al., 2016, p. 63). Plato and Aristotle even gave advice on how music is best used in education (Pannese et al., 2016). In modern times, scholars highlight the general challenges of analyzing musical emotions; these challenges lie in the ambiguity of the interpretation of these musical emotions and the unpredictability of the success of songs (Berger & Packard, 2018; Pannese et al., 2016). Research has demonstrated that sad or melancholy lyrics exert a more pronounced effect on the emotional perception of listeners than their positive counterparts (Ali & Peynircioğlu, 2006; Baradas & Sakka, 2022; Jamdar et al., 2015). This notion is also represented by the ‘paradox of negative emotion’: Music with strong negative emotions like sadness or grief is perceived as enjoyable (Pannese et al., 2016). In more detail, Pannese et al. (2016) revealed that music has been found to influence mood and physical aspects such as cardiorespiratory activity, psychomotor performance, and sexual arousal (Mitchell et al., 1998; Pignatiello et al., 1986; Vaitl et al., 1993; Witvliet & Vrana, 2007). Music has also been found to influence prosocial behavior, such as the willingness to help others (Fried & Berkowitz, 1979; North et al., 2004) or participation in social events (Wood et al., 1990). From the perspective of economics, studies reveal that music can affect the intention to purchase (Bruner, 1990), the decision time, how people judge distance (Kenealy, 1988), and how they evaluate the likelihood of success or physical attractiveness (May & Hamilton, 1980). However, certain types of music can exacerbate negative emotional states in individuals with depression or emotional instability, and adolescents experiencing strong negative emotional reactions to music may exhibit increased risk-taking behaviors (Roberts et al., 1998; Sakka & Juslin, 2018). Additionally, using music for emotion regulation can sometimes reduce well-being when paired with suppressive strategies (Chin & Rickard, 2014), and certain types of music, such as high activation/negative valence music in high-stress contexts like driving, may sustain stress-related responses, while low activation music can help mitigate these effects (Fairclough et al., 2014).

Furthermore, violent lyrics in heavy metal and rap have been shown to increase aggression and hostile cognitions (Warburton et al., 2024). Genres like punk and gangster rap have been linked to societal divides and contributed to violence (Lozon & Bensimon, 2014). Furthermore, music has been observed to reflect and amplify preexisting negative emotions or stereotypes (Fried, 2003). However, these effects are often mediated by individual predispositions and the specific context of music consumption.

In addition to the various effects of music on users’ emotional perception and behavior, emotions and contextual dimensions of users also impact music preferences (Assuncao et al., 2022; De-Lucio & Palomeque, 2023; Liu et al., 2018; Palomeque & De-Lucio, 2024). Users’ music preferences are affectively and seasonally dependent (Heggli et al., 2021; Park et al., 2019) and fluctuate daily in

several phases (Heggli et al., 2021). Personal well-being influences the choice of music and a person's economic and social status or cultural and geographical factors (De-Lucio & Palomeque, 2023; Liu et al., 2018). Broader societal dimensions such as pandemics, inflation, or recession also impact music choices; listeners prefer more positive music (De-Lucio & Palomeque, 2023; Palomeque & De-Lucio, 2024) during such times. The most recent literature on the analysis of song lyrics shows that they change over time. Parada-Cabaleiro et al. (2024) showed that previous literature also found a decrease in positive sentiment (Dodds & Danforth, 2010; Napier & Shamir, 2018) and an increase in negative sentiment (Brand et al., 2019; Napier & Shamir, 2018) and self-focused language (DeWall et al., 2011).

Several handbooks indicate that the intersection of music and economics is a well-researched area in cultural economics (e.g., Towse & Hernández, 2020; Waldfogel, 2019). Music has gained attention in the field of economics, particularly with the rise of digitization and services such as Napster, which have stimulated research on file sharing (Liebowitz, 2006; Oberholzer-Gee & Strumpf, 2010; Waldfogel, 2012). However, more general aspects—such as the economics of music-related decision-making—have recently emerged as a distinct field of research (Anglada-Tort et al., 2023). In addition, economic researchers also aim to understand the factors that influence the outcomes of song contests (Budzinski et al., 2021; Ginsburgh & Van Ours, 2003).

For our analysis, five streams of literature are particularly important: empirical studies of song contests like the ESC and its national equivalents like the German BSC or the Queen Elisabeth musical competition; studies that utilize audio feature data provided by Spotify; behavioral aspects of music such as the peak-end rule, literature on the narrative arcs of movies, and narrative economics.

First, empirical studies have examined song competitions such as the ESC and the BSC, and identified factors that influence voting behavior (Budzinski & Pannicke, 2017a, 2017b; Ginsburgh & Noury, 2008; Ginsburgh & Van Ours, 2003; Pannicke, 2016). Several ESC studies have revealed that *political, geographical, historical, linguistic, and cultural biases* influence voting behavior (Ginsburgh & Noury, 2008; Mantzaris et al., 2018; Yair, 1995). At the level of national competition, distortions in voting behavior can be demonstrated by *geographical, historical, and cultural affinities*. In addition, Budzinski and Pannicke (2022) also demonstrated the effects of an *artist's popularity* on BSC success. Unlike in the ESC, factors, such as *performance order or starting positions*, do not influence voting behavior in the BSC (Budzinski & Pannicke, 2017a; Pannicke, 2016) or contests like The Queen Elisabeth musical competition (Flôres & Ginsburgh, 1996; Ginsburgh & Van Ours, 2003). In addition, a study has identified patterns of correlation between the behavior of jury members and their ratings in the ESC with regard to proximity and has examined the impact of bias in terms of *geographical, cultural and personal affinities* such as age (Budzinski et al., 2023). The influence of judges' *personal affinities* has also been found in other competitive environments, such as a cooking show, where dissimilarity, such as gender or hair color, was decisive for a good rating (Schüller et al., 2014). Moreover, the ESC and BSC have also shown the effects of *geographical proximity* of the audience, the proximity of neighboring countries had

an impact, but the home bias in particular was significant (Budzinski & Pannicke, 2017a; Pannicke, 2016).

Budzinski et al. (2021) approach is closely related to ours; they measure the influence of objectively and subjectively measurable quality dimensions on the outcome of the BSC. Objective features, such as *speed changes*, *short insertions*, and *specific song structures*, significantly influence success as opposed to *changes in harmony*, *tempo*, or *instrumentation*. In contrast, subjective parameters of the song performance—such as *charisma*, *authentic performance*, and *triviality*—positively influence success, unlike, for example, a *confident voice* and *pleasantness* (Budzinski et al., 2021).

Second, a stream of literature that has recently emerged investigates the general influence of song characteristics on success. Several studies used Spotify's freely accessible audio features. For example, Berger (2017) analyzed over 30,000 trending songs on Spotify across various genres to identify the most common audio characteristics. Al-Beitawi et al. (2020) clustered data on the top 100 trending songs of 2017 and 2018 on Spotify. Sciandra and Spera (2022) study focused on the popularity of 273 songs by only one Italian artist. Thus far, the results have produced ambiguous findings. Sciandra and Spera (2022) found that *energy* has a negative impact on success. Meanwhile, Berger (2017), and Al-Beitawi et al. (2020) found that *energy* and *danceability* are critical factors in explaining success. Berger (2017), Sciandra and Spera (2022) and Al-Beitawi et al. (2020) found a small or negligible effect of the *valence*. Sciandra and Spera (2022) also found a positive effect of *speechiness*.

Third, behavioral economics provides a tool to investigate the behavioral aspects of music, as pioneered by the above-mentioned studies (Fried & Berkowitz, 1979; North et al., 2004; Pannese et al., 2016; Wood et al., 1990). Anglada-Tort et al. (2023) provided a comprehensive overview of studies that utilize behavioral science tools such as economic experiments, field data, and theoretical models. Their literature review focuses on heuristics and biases, like judgment heuristics. The reaction of participants also depended on the framing (framing-effect); participants who were told that a professional musician performed the song rated the song higher than those told the song was performed by a less skilled musician (Anglada-Tort & Müllensiefen, 2017). More drastically, the framing as 'suicide-inducing' or 'life-affirming' affected the perception of participants (North & Hargreaves, 2005). Moreover, the peak-end rule, for example, was found to hold in music, where experiment participants judged the music according to the most intense point and the end of the song but not the entire song (Rozin et al., 2004; Schäfer et al., 2014). In addition, music was also judged on aspects regarding social preferences, such as the profitability of pay-what-you-want schemes (El Harbi et al., 2014; Regner & Barria, 2009), online piracy (Hashim et al., 2014), peer effects like teenagers influencing each other's music tastes, and reinforcing the superstar effect or the desire for conformity in music taste (Berlin et al., 2015; Berns et al., 2010). The aspects that are important for our study are the general effects of music on decision-making and the peak-end rule, which we utilize subsequently in the empirical analysis.

Fourth, another stream of literature analyzes the influence of the narrative structure on success. Reagan et al. (2016) analyzed the narrative structures of books and found six emotional archetypes. Del Vecchio et al. (2021) found similar archetypes

for the movies using the subtitles of 6.427 movies to create emotional arcs. Thus far, the lyrics of songs have received only little attention. A handbook article on superstars Schulze (2020) claims that the content of the lyrics is, among other things, like charm or sex appeal, a rather important but hard-to-measure factor for the success of superstars. Henard and Rossetti (2014) analyzed the most influential words of popular songs over half a century and identified, for example, words such as love, quantity, baby, or girl. Meanwhile, Berger and Packard (2018) concluded that songs with distinct lyrics are more successful.

And fifth, narrative economics (Shiller, 2017, 2020), which proposes that narratives in the media influence market participants and the economy, provides a way to analyze large amounts of text and test its influence on economic variables. Narrative economics is often used in financial economics (Arratia et al., 2021; Dierckx et al., 2021) to analyze the content of social media platforms (Omuya et al., 2023; Rodríguez-Ibáñez et al., 2023) and also in media economics to measure and analyze media bias or influence of media coverage on elections or, for example, immigration (Berlemann & Thomas, 2019; Bernhardt et al., 2023; Dewenter et al., 2019), but also in areas like innovation management (Dang & Xu, 2018; Saura et al., 2023). While those examples illustrate the wide application, given the amount of text produced each day, it can reasonably be assumed that the importance of narrative economics will further increase. This aligns with the current discussion about the limits of stated preferences (Bille, 2024).

2.1 BSC data

During the 11-year period between 2005 and 2015, 16 songs representing the 16 federal states of Germany participated in the BSC each year, resulting in a total of 176 songs featured in the BSC. We collected data on the BSC itself and the bands, as well as Spotify data and the lyrics of the songs. The lyrics were then used together with a larger dataset of song lyrics to derive emotional archetypes. We will describe the Spotify data before showing how the archetypes were derived.

2.2 Spotify data and narrative structure

In our examination of factors influencing the evaluation of music within the context of the BSC, we opted to substitute objectively and subjectively measured data, as presented by Budzinski et al. (2021), with objectively generated information from Spotify. The publicly accessible Spotify API provides a wealth of data; for our analysis, we focused on the following dimensions: *acousticness*, *danceability*, *energy*, *instrumentalness*, *liveness*, *speechiness*, and *valence*.² Each of these variables is measured on a scale ranging from 0 to 100%; the only exceptions are loudness and tempo, which are measured in decibels and beats per minute. Table 1 provides a

² A description of all variables can be found here: <https://developer.spotify.com/documentation/web-api/reference/get-audio-features>.

Table 1 Spotify audio features as described in the developer portal (Spotify API from Kamehkhosh et al. (2020))

Category	Description from Spotify developer portal
Acousticness	Absence of electrical modifications in a track suitability of a track for dancing, based on various information including the best strength tempo, and the stability of the rhythm
Energy	Intensity released throughout a track, based on various information including the loudness and segment durations
Instrumentalness	Absence of vocal content in a track
Liveness	Presence of an audience in the recording
Loudness	Overall loudness of a track in decibels (dB)
Speechiness	Presence of spoken words in a track
Tempo	Speed of a track estimated in beats per minute (BPM)
Popularity	Popularity of a track, based on its total number of plays and the recency of those plays
Valence	Musical positiveness conveyed by a track

Table 2 Summary statistics of Spotify audio features

Variable	Obs	Mean	SD	Min	Max
Acousticness	175	0.1418	0.2121	0.0001	0.9090
Danceability	175	0.6194	0.1288	0.2330	0.9400
Energy	175	0.7258	0.1578	0.3140	0.9900
Instrumentalness	175	0.0081	0.0393	0.0000	0.3160
Liveness	175	0.1820	0.1367	0.0398	0.9620
Loudness	175	- 5.9568	2.0204	- 12.6620	- 1.5950
Speechiness	175	0.0781	0.0804	0.0237	0.4800
Tempo	175	126.5326	30.0315	66.9770	200.8690
Popularity	175	20.23837	16.98455	0.0000	67.0000
Valence	175	0.5820	0.2255	0.1250	0.9770

comprehensive overview of the seven categories employed in our study. Additionally, Spotify offers a measure of popularity, which indicates how popular a song is.

Table 2 provides an overview of the summary statistics derived from these audio features for 175 of the 176 songs³ that competed in the 11 years of the BSC. The data reveals noteworthy variations among the variables; for example, *acousticness* exhibits a wide spread across the 0 to 1 range, whereas *instrumentalness* attains a maximum of 0.3160.

How will these factors influence decision-making in the BSC? Given the uniqueness of the BSC, the findings of Budzinski et al. (2021), and the results generated with the Spotify audio feature data, there is no clear expectation of how each factor

³ Spotify does not provide any data for the song 'Popsong' by 'Die Raketen.'

will affect the results. Budzinski et al. (2021) revealed that winning songs were slower than other songs. This can be seen in conjunction with the ‘paradox of negative emotion’ (Pannese et al., 2016) and Sciandra and Spera (2022), finding that *energy* has a negative impact on success. However, Berger (2017) and Al-Beitawi et al. (2020) reveal that more positive aspects like *energy* and *danceability* drive success. Therefore, we cannot a priori expect more positive and powerful songs to be more successful than slower and more romantic songs, or vice versa.

2.3 Emotional arc analysis

Examining emotional arcs within song lyrics and deriving emotional archetypes is one of our contributions to the literature on song competition, music quality, and decision-making in music. While previous analyses using emotional arcs have focused on texts, such as books or movie subtitles (Del Vecchio et al., 2021; Reagan et al., 2016), the introduction to song lyrics is a new field of application. Despite their brevity, songs often convey messages, undergo developmental shifts, and narrate stories (Negus & Astor, 2015; Pannese et al., 2016). In the context of national contests, in which the majority of songs are presented in the host country’s language and promoting local music is a claimed goal, lyrics may significantly influence judges’ decisions.

Before empirically testing this hypothesis, we provide an overview of the data and describe the characteristics that result from our emotional arc analysis. This involves outlining the methodology to generate emotional arcs and subsequently clustering them according to six archetypes identified in the existing literature.

2.3.1 Emotional arcs

To create emotional arcs, we closely follow the approach of Del Vecchio et al. (2021) and use the R package *syuzhet*.⁴ As the lyrics are provided in text lines, we treat every line as a separate sentence and conduct the analysis exactly as described in Del Vecchio et al. (2021). Compared to Del Vecchio et al. (2021), our sample of 176 songs is relatively small. Therefore, we enlarged the original sample with a broader and more general set of songs to deduct clusters and to use typical archetypes in popular songs independently of the BSC. We obtained a sample of song lyrics of the weekly Top50 single charts from the USA from November 1977 to October 2022, containing 84,098 songs. The raw data has been cleaned and adopted to prepare for our analysis. This led to 6,587 songs, as we only used songs in English (or predominately in English⁵), excluded songs with no or very short lyrics, and excluded duplets. Mainly the exclusion of songs that appeared more than once in the sample reduced the number dramatically, as most songs stayed in the charts

⁴ <https://cran.r-project.org/web/packages/syuzhet/vignettes/syuzhet-vignette.html>.

⁵ We excluded songs which contained more than 20 percent of non-English words.

for longer than one week, and songs also stay in the charts for several weeks.⁶ The remaining 6.587 songs were used to deduct emotional arcs as described below. Only after this were the results of the emotional arc analysis of the German songs from the BSC added. Like this, the song lyrics from the BSC are only a very small part of the sample. This means that the songs from the BSC are placed within the broader clusters rather than the clusters built on this small sample.

Regarding the BSC songs, 176 participated in the BSC over 11 years; the majority (168) of the songs were in German, six in English, one in Dutch, and one in Turkish-German. For these, we used a German dictionary; moreover, dialects (like Bavarian or Frisian) were translated into standard German for the analysis. The English dictionary was used for English songs, and songs in other languages (like Dutch or Turkish) were translated using DeepL Translate.

The analytical approach adopted begins with a traditional sentiment analysis, in which the valence for each sentence (or line) is computed by assigning a sentiment value (negative, neutral, and positive) to individual words within the line. Following the methodology outlined by Del Vecchio et al. (2021), the resultant sentiment trajectory undergoes transformation using the discrete cosine transform (DCT). Subsequently, the outcomes are further standardized to create a uniform scale, ranging from 0% (the beginning of the song) to 100% (the end of the song). This normalization process ensures comparability among songs, rendering them independent of their actual duration. In effect, each song is normalized, thereby enabling a meaningful and standardized comparison across the emotional trajectories of diverse compositions (Del Vecchio et al., 2021).

2.3.2 Arc archetypes

Reagan et al. (2016) and Del Vecchio et al. (2021) established six archetypes for books and movies. In our initial analysis, we employed k-means clustering with six clusters to compare the emotional trajectories of the 6.587 chart songs plus 176 BSC songs against these established archetypes. Figure 1 illustrates the alignment between our clusters and those identified by Del Vecchio et al. (2021). Since we do not have access to the original data, a manual matching process was undertaken by two independent researchers.

Figure 1 provides an overview that compares the results of our study with those of Del Vecchio et al. (2021), providing both the original archetype, graph, and examples from Del Vecchio et al. (2021) and the graph from our complete dataset with representative examples of songs from the BSC. The original archetypes from Del Vecchio et al. (2021) are on the left side, with three corresponding movies next to it. The right side of Fig. 1 shows the plots based on the emotional

⁶ Further data cleaning took place to account for lyrics that mention, for example, [chorus] but do not show the chorus lyrics again or parts where one line is repeated two or three times or more often. For our purpose, it is important to include each of those repetitions separately. For this, the dataset was enhanced by using genius.com to replace songs where these problems occurred. In case this did not solve the problem, song lyrics were researched by hand.

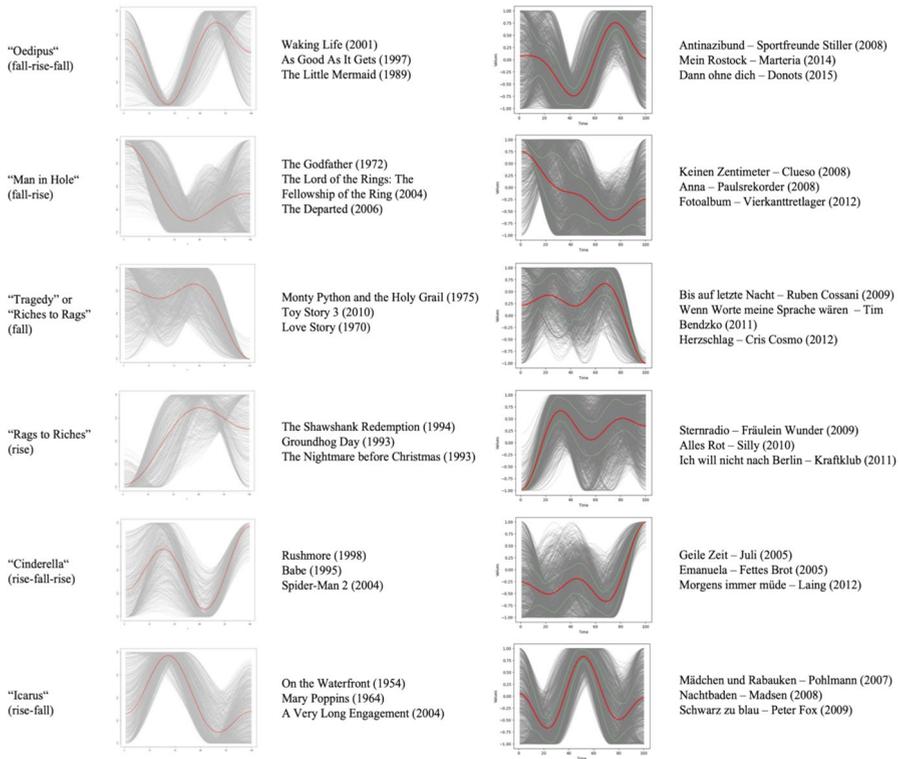


Fig. 1 Arc archetypes for songs (right-hand side) compared to movies from Del Vecchio et al. (2021) on the left-hand side

arcs derived from the song lyrics. The red line shows the average, while the gray lines are 100 randomly chosen song lyrics for each archetype; the examples of songs given on the right side are selected from the BSC dataset. The overall correspondence between our clusters and those identified by Del Vecchio et al. (2021) is close, with occasional divergence.

‘Oedipus’, the first archetype, corresponds very well with the original graph. There is an apparent fall before the lines peak and fall again. Even though the deepest point of the red average is not as deep as in Del Vecchio et al. (2021) and the fall, in the beginning, is less steep, in general, the shape of the lines is similar. The same holds true for the second archetype, ‘Man in Hole’ in Fig. 1; while the lowest point is further to the left and the fall line is less smooth, the shape is still close. ‘Tragedy/Riches to Rags’ is the third archetype, which follows almost the exact same pattern, especially with the steep fall in the end. However, the line is on a relatively high level during most of the course of the songs. The fourth archetype, ‘Rags to Riches’, is exactly the opposite pathway, starting very low and then increasing. The songs, however, show more movements and another small down in the middle of the songs. The fifth archetype, ‘Cinderella’, shows,

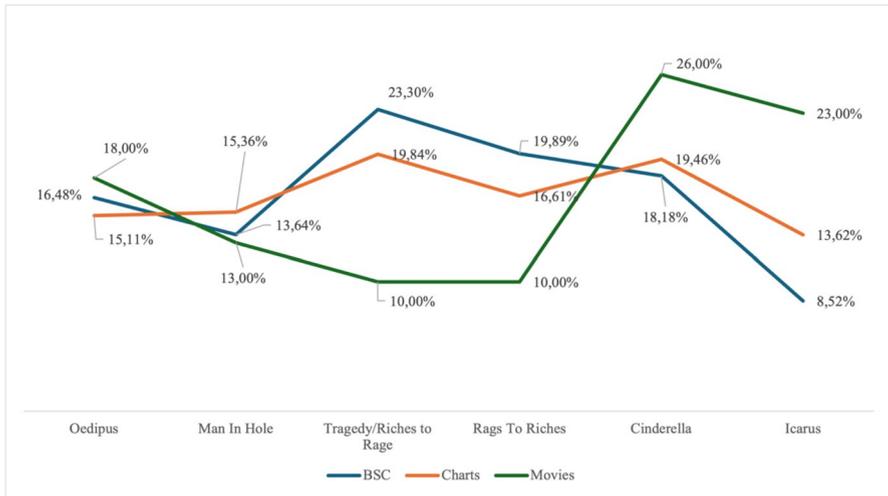


Fig. 2 Distribution of songs in clusters in the whole dataset and in the BSC dataset compared to movies from Del Vecchio (2021)

in our case, a less pronounced rise but a fall before the characteristic rise to the very top in the end. Regarding the ending, ‘Rags to Riches’ and ‘Cinderella’ are very similar, but with some deviation in the middle. The most difficult to plot is the ‘Icarus’ plot, characterized by a rise and fall. In our case, there is a very pronounced peak in the middle. Both the movie and song versions show a similar end, but the peak is further in the middle for the songs compared to the movies on the left side.

Further, the distribution of songs across clusters is not uniform (see Fig. 2). While ‘Oedipus’ and ‘Man in the Hole’ are rather close for the BSC, the chart lyrics in general and for movies. The biggest deviations can be found in ‘Icarus’, where BSC and chart lyrics are 8.52% and 13.62%; the share of movies in this archetype is 23% higher. Exactly the opposite can be found for ‘Tragedy/Riches to Rags’, which represents only 10% in case of movies, but 23.30% and 19.84% for BSC and chart lyrics. The ‘Cinderella’ archetype has the highest share, with 26% for movies, but only about 20% for BSC and chart song lyrics.

2.3.3 Peak analysis

The peak-end rule—which describes that people typically do not remember every aspect of, for example, a vacation, but only the peak and the end—was also found to be relevant in music (Anglada-Tort et al., 2023). We can leverage the analysis of emotional arcs to pinpoint each song’s emotional peaks. To identify these peaks, we determined each emotional arc’s maximum and considered the surrounding area. Specifically, we aggregated *valence* values from the previous and next five positions around the peak, thereby creating a variable labeled *peak*.

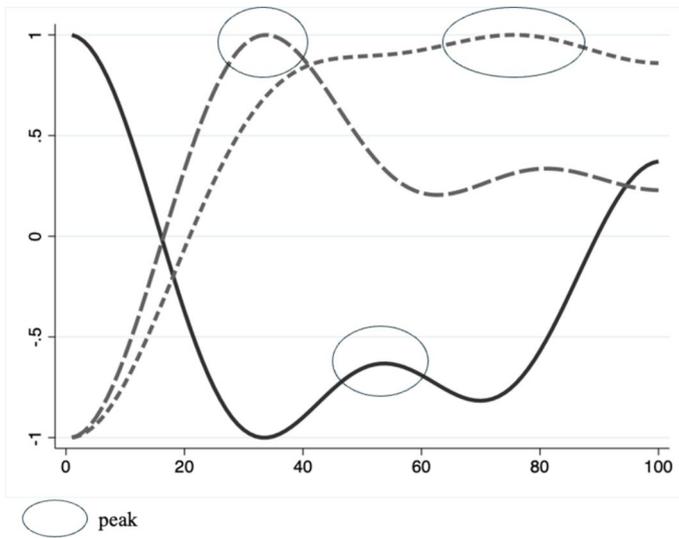


Fig. 3 Peak identification for three exemplary songs

The definition of the *peak* variable excludes peaks at the beginning or the end. The impact of the ending is captured by our happy-end analysis. The *peak* variable enables us to focus exclusively on the emotional peaks occurring during the song. The results of the approach are illustrated in Fig. 3. The three songs illustrated in Fig. 3 showcase that the *peak* variable measures both the peak's size and width. The song represented by the solid line has a relatively short peak that falls below zero. The long-dashed line corresponds to a song with an even more pointed peak, attaining the maximum value of one. Finally, the short-dashed line, with the same highest point, demonstrates that the *peak* variable effectively captures the flatter character of this peak.

Table 3 provides summary statistics for the variable *peak*. This continuous variable quantifies the emotional impact of a song's highest phases. More specifically, the variable represents the sum of the highest point within each emotional arc, thereby incorporating the *valence* values from the surrounding 10 positions. Using the *peak* variables, our objective is to assess the influence of the song's high and/or low phases on the decision-making process.

2.3.4 Happy-end analysis

The second facet of the peak-end rule is the emotional ending of the song. The emotional impact of a song's ending, whether characterized by happy or sad sentiments, may influence the decision of voters. Our analysis employed k-means clustering to categorize the 176 songs into two clusters, focusing exclusively on the final 20% of each song. Figure 4 depicts the resulting clusters, distinguishing songs with happy endings from those with sad endings. This segmentation enables us to explore how

Table 3 Peak variable summary statistics

Variable	# Observations	Mean	SD	Min	Max
<i>peak</i>	176	6.5502	3.1291	- 2.3394	8.9676

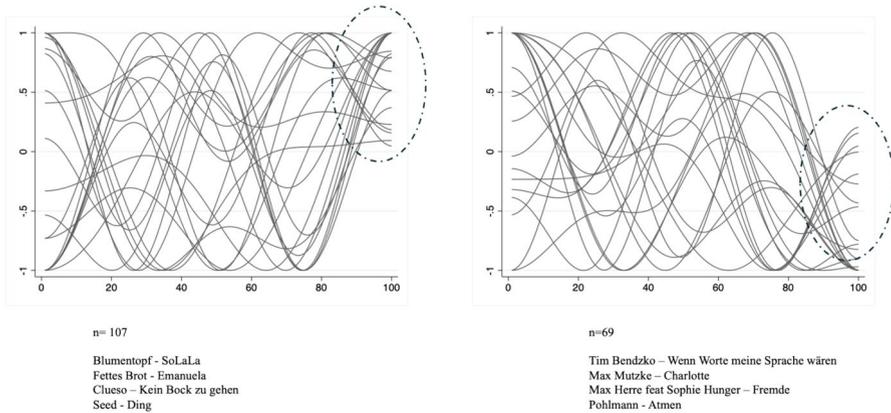


Fig. 4 Happy and sad ending—data clustered with k-means

Table 4 Happy and sad endings summary statistics

	Average points	SD	# Songs
Sad end	4.0550	3.7900	69
Happy end	3.3480	3.8720	107

a song’s emotional culmination may influence a judge’s judgment in a song contest setting.

The analysis reveals an imbalance in the two clusters, with a greater emphasis on songs characterized by happy endings. Table 4 presents the summary statistics for the variable *happy end*, which is either one for a happy ending or zero for a sad ending. According to our definition, only 69 songs have a sad ending, but the average

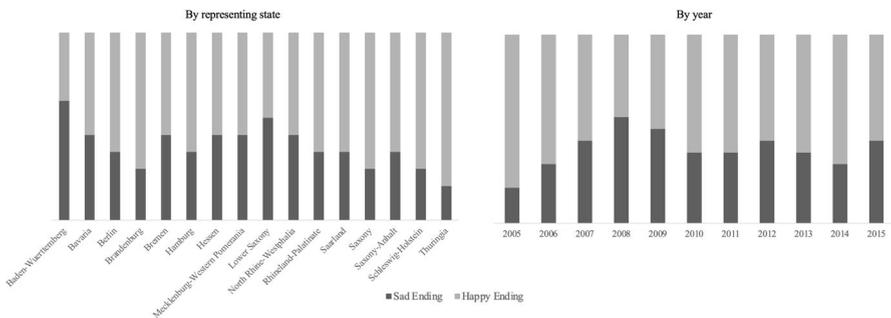


Fig. 5 Happy and sad endings statistics by representing state and per year

Table 5 P values of pairwise Wilcoxon rank-sum tests

	Oedipus	Man in Hole	Tragedy	Rags to Riches	Cinderella	Icarus
Oedipus		0.3341	0.0012	0.0000	0.0000	0.1287
Man in Hole	0.3341		0.0056	0.0001	0.0001	0.0178
Tragedy/Riches to Rags	0.0012	0.0056		0.7886	0.2925	0.0000
Rags to Riches	0.0000	0.0001	0.7886		0.3458	0.0000
Cinderella	0.0000	0.0001	0.2925	0.3458		0.000
Icarus	0.1287	0.0178	0.0000	0.0000	0.000	

sum of points is 4.055, higher than 3.348 for happy endings. A two-sided t-test and a Wilcoxon rank-sum test reveal that the two groups (happy and sad endings) are statistically different from zero. Meanwhile, the non-happy-ending group tends to attract more points on average. Meanwhile, happy song endings are more frequently selected to represent certain states such as Brandenburg, Schleswig–Holstein, or Thuringia (see Fig. 5).

Over the years, there was a variation in the ratio between happy and non-happy endings. Only in 2008, there were more songs with sad endings than happy endings; in 2009, the number of songs with happy and sad endings (according to our definition) was equal. The right-hand side of Fig. 5 adds a temporal dimension to the analysis and reveals how the prevalence of emotional endings in the BSC has evolved over the years (Table 5).

2.3.5 Distribution of points according to emotional archetypes

The emotional archetypes differ regarding the points they have achieved over the 11 years in our dataset, see Fig. 6. The ‘Tragedy/Riches to Rags’ and the ‘Cinderella’ archetypes, characterized by a very negative and a very positive ending (see Fig. 1), attract the fewest points. As can be seen in Fig. 2, those two archetypes represent 23% and 18% of all songs (41% in total). ‘Tragedy/Riches to Rags’ is the most frequently found archetype and ‘Cinderella’ is on position three regarding the frequency. While the ‘Oedipus’ archetype, which is characterized by a fall-rise-fall structure, attracts the most points, it is only the fourth most frequent plot. Number two regarding points is the ‘Icarus’ archetype, which is also characterized by a rise and a fall, is the least frequent archetype in the dataset. Places three and four are two archetypes that end with a rise: ‘Rags to Riches’ representing a constant rise and the second most frequent archetype, and ‘Man in Hole’, which describes a deep fall from a very high starting point and a surge toward the end of the plot.

To assess the similarity across clusters regarding the points allocated to songs, we conducted pairwise Wilcoxon rank-sum tests for the entire BSC sample. Considering the right-hand side of Fig. 6, the numbers are intuitive: ‘Oedipus’ is not significantly different from ‘Man in Hole’ and ‘Icarus’. ‘Man in Hole’ differs significantly from all other archetypes except ‘Oedipus’. The ‘Tragedy/Riches to Rags’ archetype does not differ from ‘Rags to Riches’ and ‘Cinderella’. The same holds for ‘Rags to

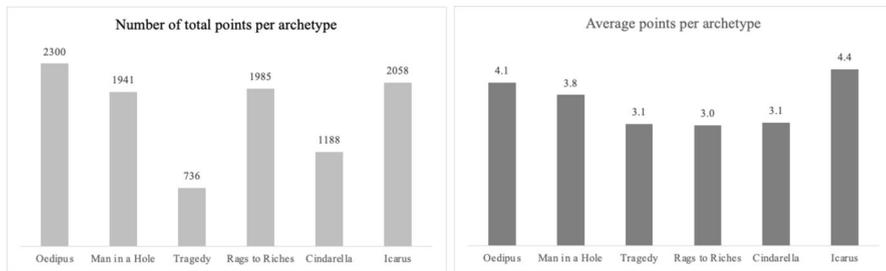


Fig. 6 Total and average points during the BSC per archetype

Riches’ and ‘Cinderella’; all three archetypes differ significantly from ‘Oedipus’ and ‘Icarus’, but not among each other. Finally, ‘Icarus’ differs significantly from ‘Oedipus’ and all other archetypes.

2.4 Empirical analysis

The BSC dataset gives us a unique environment to test the potential effect of the emotional archetypes on the decision if and how many are awarded to a particular song. Compared to international song contests, for a national song contest with predominately songs in the country’s language, it is reasonable to assume that listeners will be able to understand the lyrics and the meaning of the words used in the songs correctly. Language is a major topic in the Eurovision Song Contest (ESC). During two times of the history of the ESC, songs were only allowed to be performed in the official language of the represented country. Since 1999, the language can be freely chosen, whereby the most used languages are the respective language of the country and English.⁷ Even songs performed in the mother tongue hold the potential to be misunderstood. However, the own language makes it much more likely to understand the meaning and the emotional connotation.⁸ Therefore, we argue that while the BSC is no longer an ongoing contest, it is still a perfect setting to investigate emotional archetypes because of most German lyrics. Additionally, we can build on and compare our results to existing research (e.g., Budzinski et al., 2021).

To isolate the correlation of a single aspect with the actual points given by each state, we utilize an estimation strategy that draws from the analysis conducted by Budzinski et al. (2021), albeit with certain modifications. First, we adopt the panel structure, with the state as the panel variable, using fixed effects for the 16 federal states. In the context of the BSC, each of the 16 German states can assign points to every song, with the option to abstain from awarding any points. The points allocated by each state for each song constitute our dependent

⁷ The Eurovision Song Contest and Wikipedia list 57 languages plus several languages and dialects that have been used for song lyrics during the history of the ESC. URL: <https://eurovision.tv/history/origins-of-eurovision>; https://en.wikipedia.org/wiki/List_of_languages_in_the_Eurovision_Song_Contest.

⁸ We discuss this assumption further in our Limitations sections after the discussion.

variable. Given the possibility of states assigning zero points, resulting in a high density around zero, we follow Budzinski et al. (2021) in employing various models to address this issue. These models include the state fixed-effect panel estimation and the Poisson fixed-effects model. We replace the zero-inflated Poisson model utilized by Budzinski et al. (2021) with a logit model to investigate whether our independent variables can explain why a song received more than zero points by a certain state, again with fixed effects on state level. Therefore, we estimate the likelihood of receiving points with a panel logit model. In this case, our dependent variable *vote* determines whether a state's vote is zero when no points are given and one when the vote exceeds zero. This supplementary analysis aims to provide insights into the decision process, as the initial determination of a state involves the choice of whether to award points. Only in cases where points are granted does the number of points become relevant. Therefore, we explore whether the characteristics of the song influence the decision to allocate points. The utilization of a panel structure in which the voting behavior of each state is observed over the 11 years enables the exclusion of time-invariant effects that are specific to each state and might impact voting behavior. While our estimation strategy aligns with previous papers, our independent variables present significant differences. First, we use Spotify's audio features to capture the dimensions related to the song's music. Second, we use different variables to capture the effect of the emotional arcs.

- *Data*: We use the complete dataset of the BSC for the 11 years between 2005 and 2015. Each year, 16 German states sent one representative song to the contest ($16 \times 11 = 176$). Each of the 16 states also voted on all 16 songs, including their own, whereas the maximum points possible are twelve, and the minimum points given are zero. Therefore, we have a total of 2.816 observations, with each state awarding points to each song each year ($16 \times 16 \times 11$).
- *Dependent variables*: Our dependent variable for models 1, 2, 4, and 6, is, therefore, the points awarded by each state to each song. As Spotify does not provide data for one song, our dataset is limited to 2.800 observations when we include Spotify data. For models 3 and 6, we construct a new variable, *vote*; this variable takes the value of one if the points are higher than zero and zero if a particular state awarded no points to a certain song for a specific year.
- *Spotify audio features*: Here, we do not include popularity as we cannot eliminate that the popularity measured by Spotify results from the success (or lack of success) of certain songs in the BSC. If a song won the contest or became popular after reaching a bigger audience through the contest, the popularity measure would be endogenous.
- *Emotional archetypes*: To test the effect of emotional archetypes on the points awarded by each state, we first only consider the *happy-end* and *peak* effects, including the dummy variable *happy end* and the continuous variable *peak*. As *happy end* and *peak* are deducted from the archetype, we substitute these two variables with categorical variables for each of the six archetypes except the first, 'Oedipus'. This is because *happy end* and *peak* capture similar effects as those of the archetypes. Finally, we run the same models for each archetype compared to

all others, again without *happy end* and *peak*. Songs that contained slang or other languages were categorized as slang or English.

- *Other factors*: We also include a few hard factors in the analysis, such as the duration of the songs, which had a significant impact on other studies. The variable *home* is one aspect that comes from the state and is also represented by the same state. Additionally, we include the starting position, which has been found to have no impact thus far. Additionally, we employ the quadratic term to capture a possible nonlinear effect of the starting position. We also control for the genre of the song as different genres might attract more points, which could be captured by other variables otherwise.⁹ To account for correlation issues, we utilize the standardized value for the starting position.

After presenting the results, we perform a series of robustness checks to test the stability of the results.

3 Results

The results presented in Table 6 provide an overview of the findings. The results can only be interpreted as correlations but do not imply a causal relationship. We present the results once for the happy-end and peak analyses (left-hand side, models 1–3) and then include the clusters (right-hand side, models 4–6). In this manner, we separate the *happy-end* and *peak* analyses from the actual emotional arcs, as including both the emotional arcs would include happy endings and peaks.

This overall pattern aligns with Budzinski et al. (2021) where not all variables exhibit significant correlations with the points given to each song. The negative correlation observed with speechiness is interesting, as it indicates that songs with more spoken words tend to receive fewer points. Furthermore, home bias is evident, with states showing a preference for their songs by awarding them higher points. This aligns with typical home bias observed in studies of the ESC and BSC (Budzinski & Pannicke, 2017a; Pannicke, 2016) and in other competitions as sports (Fischer & Haucap, 2021; Jones, 2013; Pollard, 1986, 2006). More interestingly, contrary to Budzinski and Pannicke (2017a), the starting position has a nonlinear influence on the votes.

Further, English songs perform lower and are less likely to receive points higher than zero. Interestingly, slang negatively affects the points given but positively affects the likelihood of receiving greater than zero votes. This effect may be due to the perception that slang adds an innovative or original element to the song but does not resonate for extra points.

A happy ending for a song reveals a significant negative effect on the points given to a song and the likelihood of the song receiving points. Moreover, our *peak*

⁹ Genres have been researched and categorized, which led to the genre alternative/indie, blues/soul, hip-hop, Pop, reggae, rock/metal, techno/dance, and folk music. Pop captures most tracks, with folk music being the smallest category.

variable, which measures the intensity of the highest peak during the song, negatively influences the votes and the likelihood of obtaining points. This is in line with the peak-end-effect theory—the peaks have an influence, but a very positive emotional peak measured by the lyrics translates into lower points given. This result reflects that the peaks are remembered; however, as with the results of the happy-end analysis, more positive peaks are less advantageous for receiving points.

Our results also in contrast previous analyses: Budzinski et al. (2021) revealed that winning songs were slower than losing songs, even when the tempo was not a significant factor. In our case, tempo is also negative and significantly different from zero, although the economic impact is low. Compared to other studies of song popularity using Spotify audio data, we also find similar and divergent results. In our case, *energy* has no influence (only in one case and only on the 10 percent level), while in Sciandra and Spera (2022) it has a negative impact on success. This result is also different from those of Berger (2017) and Al-Beitawi et al. (2020), who find that *energy* and *danceability* are key to success. *Danceability* is weakly negatively correlated with the number of points given but highly with the likelihood of receiving points. This aligns with our finding that a happy ending does not appear to resonate within the BSC. Another finding adding to this is valence's negative and significant impact, which measures how positively the music is perceived (not the lyrics). Here, too, the more positive the music, the fewer points are given and the less likely the song will obtain points higher than zero. This is in contrast to the results of Berger (2017), Sciandra and Spera (2022), and Al-Beitawi et al. (2020), all of which find a minor or negligible effect of the valence variable. In addition, *speechiness* is negatively correlated with points, whereas in Sciandra and Spera (2022) it has a positive correlation.

Substituting *happy ends* and *peak* with the actual clusters as categorical variables yields qualitatively similar results for the other variables (models 4–6). In comparison with the first archetype, 'Oedipus', it is mainly the likelihood of receiving more than zero points that are affected (model 6). 'Tragedy/Riches to Rags', 'Rags to Riches', and 'Cinderella' are less likely to receive points than 'Oedipus', and 'Man in Hole' is more likely. Regarding points given, 'Tragedy/Riches to Rags' and 'Rags to Riches' perform worse than 'Oedipus'. Only 'Icarus' shows no significant results, possibly due to the relative closeness of the two archetypes in our case (see Fig. 1). The results adapt when we change the category that is left out,¹⁰ therefore, with this approach, we can only evaluate the relative effect of the archetype compared to another archetype.

To gain insights into how each emotional archetype performs against all others, we conducted six-panel regressions, each with a dummy variable representing only one archetype in each regression. The results for each emotional arc are presented in Table 7.

The results align with early findings. Notably, the 'Man in Hole' and 'Icarus' archetypes perform significantly better than all the other archetypes, while 'Oedipus'

¹⁰ Choosing 'Man in Hole' as the relative category, for example, shows that 'Tragedy/Riches to Rags,' 'Rags to Riches,' and 'Cinderella' perform significantly worse with regard to points given and the likelihood of receiving points, while 'Oedipus' and 'Icarus' perform only significantly worse regarding the likelihood of receiving points.

Table 6 Overview of results: The left-hand side shows results with happy end and the peak. Right-hand side shows results with categorical dummies for each clustered as defined by Del Vecchio (2021)

Variables	(1)	(2)	(3)	Variables	(4)	(5)	(6)
	Panel regression	Panel Poisson	Panel logit		Panel regression	Panel Poisson	Panel logit
Duration	0.20*** (0.064)	0.055*** (0.016)	-0.13*** (0.042)	Duration	0.22*** (0.070)	0.062*** (0.018)	-0.12*** (0.043)
Liveness	0.35 (0.48)	0.029 (0.12)	1.21*** (0.30)	Liveness	-0.21 (0.48)	-0.10 (0.12)	0.85*** (0.32)
Instrumentalness	1.23 (1.27)	1.03*** (0.35)	0.80 (0.80)	Instrumentalness	2.30* (1.28)	1.25*** (0.35)	1.57** (0.76)
Danceability	-1.18* (0.61)	-0.24 (0.16)	-1.78*** (0.38)	Danceability	-0.62 (0.69)	-0.13 (0.18)	-1.12*** (0.38)
Energy	-0.083 (0.44)	0.031 (0.12)	-0.47* (0.26)	Energy	0.0066 (0.52)	0.065 (0.14)	-0.12 (0.28)
Tempo	-0.0084*** (0.0021)	-0.0020*** (0.00057)	-0.0090*** (0.0016)	Tempo	-0.0073*** (0.0021)	-0.0018*** (0.00058)	-0.0084*** (0.0015)
Speechiness	-2.54*** (0.77)	-0.71*** (0.22)	-2.48*** (0.58)	Speechiness	-2.31*** (0.77)	-0.59*** (0.23)	-2.37*** (0.63)
Valence	-1.50*** (0.38)	-0.43*** (0.10)	-1.03*** (0.22)	Valence	-1.32*** (0.42)	-0.36*** (0.11)	-1.03*** (0.23)
Happy end	-0.57*** (0.11)	-0.17*** (0.029)	-0.52*** (0.059)	Man in Hole	0.32 (0.19)	0.075 (0.052)	0.51*** (0.082)
Peak	-0.048*** (0.016)	-0.011** (0.0047)	-0.025** (0.013)	Tragedy	-0.53** (0.22)	-0.20*** (0.069)	-0.26*** (0.10)
English	-1.24*** (0.25)	-0.47*** (0.097)	-0.74*** (0.24)	Rags to riches	-0.85*** (0.21)	-0.27*** (0.059)	-0.13* (0.069)
Slang	-1.07**	-0.28***	0.74*	Cinderella	-0.26	-0.11	-0.40***

Table 6 (continued)

Variables	(1)	(2)	(3)	Variables	(4)	(5)	(6)
	Panel regression	Panel Poisson	Panel logit		Panel regression	Panel Poisson	Panel logit
Home	(0.39) 1.97***	(0.099) 0.45***	(0.43) 0.54***	Icarus	(0.27) 0.38	(0.077) 0.050	(0.11) 0.18
Start_position	(0.29) 0.32***	(0.059) 0.081***	(0.21) 0.21***	English	(0.22) - 1.29***	(0.054) - 0.44***	(0.15) - 0.85***
Start_position ²	(0.018) 0.032***	(0.0059) 0.0060***	(0.012) 0.020***	Slang	(0.27) - 1.03**	(0.10) - 0.27***	(0.25) 1.02**
Constant	(0.0038) 4.26***	(0.0012) (0.0021)	(0.0021)	Home	(0.38) 1.97***	(0.098) 0.45***	(0.45) 0.55***
Observations	(0.31) 2800			Start_position	(0.30) 0.32***	(0.063) 0.082***	(0.21) 0.21***
Genre dummies	Yes	Yes	2800		(0.018)	(0.0058)	(0.012)
Year dummies	Yes	Yes	Yes	Start_position ²	0.032***	0.0057***	0.019***
R ²	0.21		Yes	Constant	(0.0038) 3.08***	(0.0012)	(0.0020)
				Observations	(0.52) 2800		2800
				Genre dummies	Yes	Yes	Yes
				Year dummies	Yes	Yes	Yes
				R ²	0.22		

Level of significance: *p < 0.1, **p < 0.05, ***p < 0.01

Table 7 Performance of each emotional arc compared to all other arcs in separate panel regression

	Oedipus	Man in Hole	Tragedy/ Riches to Rags	Rags to Riches	Cinderella	Icarus
Panel regression	0.25 (0.16)	0.57*** (0.13)	− 0.40 (0.26)	− 0.90*** (0.15)	− 0.095 (0.22)	0.64*** (0.16)
Panel Poisson	0.096** (0.043)	0.17*** (0.038)	− 0.15** (0.076)	− 0.26*** (0.048)	− 0.052 (0.067)	0.14*** (0.038)
Panel logit	− 0.056 (0.055)	0.56*** (0.089)	− 0.21** (0.10)	− 0.20*** (0.043)	− 0.46*** (0.096)	0.27** (0.12)

Level of significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

only performs better in the Poisson model. These archetypes represent a rather negative mood with a positive peak but relatively sad endings. The ‘Tragedy/Riches to Rags’ archetype—where the average line is relatively high throughout the song but has a very steep decline in the end—performs worse than the other archetypes. Interestingly, the archetype associated with more positive themes, such as ‘Rags to Riches’, also performs worse. ‘Cinderella’, associated with a very positive ending, does not affect the points given but is less likely to receive points.

4 Robustness of the results

The results show interesting dynamics regarding the Spotify data and the emotional archetypes derived from the lyrics. To test the robustness of the findings, we first add control variables for the gender of the singer or band. This might influence how the song is perceived, which is not captured by the Spotify valence variable or the archetypes. Second, we use another approach to measure the sentiment of the lyrics.

Information about the gender of the band or singer was researched and coded.¹¹ Including dummies for females and mixed bands reveals no qualitative change regarding the *happy-end* and *peak* analysis (left side Table 8). Females receive fewer points, and mixed-gender bands receive more points in the Poisson model and are more likely to receive points. Concerning the models that include the archetypes, the same gender dynamic can be observed: Females perform worse than men, and mixed bands perform better. However, *energy* becomes more important in this model, as well as the differences between ‘Oedipus’ and ‘Man in Hole’ and ‘Icarus’ (see Table 8 in the appendix). Regarding the models where only one archetype is analyzed compared to all others, the results are very stable in terms of direction and significance level. Only ‘Oedipus’ is less significant in the Poisson model but more significant in the logit model (see Table 9 in the appendix).

¹¹ Code used: 0 = male, 1 = female, 2 = divers, and 3 = mixed gender in a band. However, we did not find any performer in the category divers.

Besides the archetypes, another approach to analyzing the lyrics is a more classical sentiment analysis. For the sentiment analysis of German text files, we utilized the Hugging Face Transformers library¹² a fine-tuned version of BERT designed for sentiment classification in the German language (Guhr et al., 2020). BERT, a transformer-based architecture, can classify text into sentiment categories such as positive, neutral, and negative, coded as 1,0 and -1,0. The model provides a score indicating the confidence of its prediction for each class, where the highest score determines the final sentiment label. We conducted a sentiment analysis for the whole lyrics and to be able to compare it to the happy-end results also for the last 20% of the lyrics.

For the analysis, we use the models (1) to (3). We replace the *happy-end* variable with the results from the new sentiment analysis for the last 20% of each of the lyrics and leave out the *peak* variable based on the original approach. Again, the *happy-end* variable is significantly negative, showing that a *happy end* is associated with lower points and a reduced likelihood of receiving points (see *happyend_hf* in Table 10 in the appendix). Taking the sentiment of the whole song shows that a higher sentiment category leads to fewer points but does not influence the likelihood of receiving points (see *sentiment_hf* in Table 10 in the appendix).

5 Discussion

While the importance of storytelling in books or movies is evident, its significance in popular music is less apparent. Using emotional archetypes helps to analyze the potential impact on listeners. As the BSC is a national event conducted in Germany with predominately German songs, it provides an ideal setting for exploring the influence of various narrative archetypes on the points achieved in the contest. For the lyrics to impact the results, they must be understood, and emotional connotations must be recognized. While we cannot directly control it, the national setting makes this assumption much more plausible than in an analysis of the ESC, where both the language of the performance and the language of the audience differ very clearly. This unique setting enables us to treat the narrative structure as an independent variable and examine its correlation with the results. We also consider Spotify's audio features to account for other factors influencing the decision-making process comprehensively. This dual perspective adds new dimensions to the literature on song contests.

Our analysis of an extensive number of popular song lyrics reveals narrative structures similar to those of movies and books and that those structures correlate with the results of the BSC. Additionally, we uncover evidence indicating that the peak-end rule impacts the points given. We also find that both a happy ending and peaks during the song have a negative impact on the points awarded.

Our study reveals a correlation between more negative emotional archetypes, such as 'Man in Hole' or 'Icarus' and to a certain degree, 'Oedipus' performs

¹² Specifically employing the *oliverguhr/german-sentiment-bert* model.

better, even though they describe a very challenging emotional journey and with a pronounced fall in the end perform better at the BSC. Emotional archetypes that are associated with a more positive mood, like ‘Rags to Riches’ describing a constant rise or ‘Cinderella’ which has a clear rise in the end, or ‘Tragedy/ Riches to Rags’ which sounds negative, but during most of the song the average line of the emotional archetype is very high as in the ‘Rags to Riches’ and only falls steeply in the end tend to perform worse. This might appear counterintuitive, as the format of the contest aims to provide entertainment. However, the results align with previous findings and the so-called ‘paradox of negative emotion’ (Pannese et al., 2016). In contrast, a few of the studies using Spotify data that consider different acoustic factors found that *energy* and *danceability* have a positive impact on success (Al-Beitawi et al., 2020; Berger, 2017); however, Sciandra and Spera (2022) find that *energy* has a negative impact on success, but that the overall *valence* has a negligible effect. Therefore, the influence of the factors appears to depend on the measurement of success. While the top songs on Spotify may include more party songs with a more energetic profile, viewers of song contests may look for a different type of entertainment. Here, a more detailed analysis of different formats, success definitions, or charts might be able to disentangle the effect of varying audio features. The differences found in our study compared to more general studies may be explained by different consumption motives and audiences between general music consumption, as measured by the Spotify data, and a song contest.

Our analysis demonstrates that the emotional archetype of a song can influence its perception. However, the BSC is an older format that is no longer conducted. Therefore, there are no direct implications for the participants of the contest. Nonetheless, the uniqueness of the BSC lies in its local format, where language plays a significant role, and the audience typically speaks German as the primary language. Hence, most people can immediately understand the lyrics of the song and, according to our analysis, react to it. It would be intriguing to explore whether the narrative structure has a similar effect in other contests, such as the ESC, where the songs are performed in different languages and the language of the performance may not always be native for the audience. If the correlation is smaller there, it could serve as a strong indicator for the reception of emotional arcs in the lyrics. This would also lead to more generalizable results.

Further, the impact of music and song lyrics extends beyond entertainment, holding implications for marketing strategies, for example. Marketers can leverage insights to determine how music influences consumer behavior (for example, Bruner, 1990), thereby employing this knowledge to evoke specific emotions to potentially influence consumers’ purchasing decisions. The findings also provide actionable insights for music managers and contest organizers. Emotional archetypes, such as ‘Man in Hole’ and ‘Icarus’, demonstrate strong audience appeal in national song contests. Music managers could prioritize these archetypes when selecting or promoting songs to optimize performance outcomes. Additionally, strategic adjustments to the presentation format—such as emphasizing the narrative elements of performances—could enhance audience engagement and understanding of the emotional dimensions conveyed in songs.

This study further contributes to the understanding of music as a cultural/economic good by showcasing the significance of narrative structures in shaping audience perceptions and decision-making. The paradox of negative emotion highlights that songs with sorrowful or complex emotional themes have a unique capacity to captivate audiences, emphasizing the pivotal role of storytelling in music as a driver of audience engagement. Furthermore, the results provide valuable insights into narrative economics by illustrating how emotional engagement can influence consumption behavior within the cultural domain.

In addition, audience music preferences are influenced by affective and contextual dimensions (Assuncao et al., 2022; De-Lucio & Palomeque, 2023; Liu et al., 2018; Palomeque & De-Lucio, 2024). Users' preferences are affective and depend on personal well-being and societal dimensions (Heggli et al., 2021; Palomeque & De-Lucio, 2024; Park et al., 2019). Therefore, the audience votes and awards are situational and difficult to define but could have been an influencing factor.

6 Limitations

Our study has limitations that also show opportunities for further research. Most importantly, our analysis assumes that the lyrics are perceived and understood. However, we cannot control whether the whole song was listened to and understood. While this is true to a certain degree for other analyses regarding the Song Contests as well, as we use the song's course, it is more important here. Considering the ongoing debate about problems associated with stated preferences in Cultural Economics (Bille, 2024), this limitation might be a starting point for further research. Given the results, it can be attractive to perform laboratory experiments with different emotional archetypes, where more factors, including understanding the lyrics, can be controlled. For example, more control variables like the artist's age could be used to control for more personal aspects that drive the creativity of the singers (Borowiecki, 2017).¹³

Furthermore, meaningful or trivial lyrics could be tested in a controlled setting. This would also allow for causal analysis of narrative archetypes and behavior. Given the growing importance of narrative economics (Shiller, 2020), behavioral aspects of music, and experimental economics (Anglada-Tort et al., 2023), understanding the influence of emotional archetypes could open new approaches and opportunities in cultural economics.

In addition, testing the methodology's application and robustness in analyzing a larger number of songs, perhaps from historical charts, could help to further understand the logic and usefulness of the insights generated.

Second, while Spotify's audio features exhibit good explanatory power, a direct comparison of objective and subjective quality variables from Budzinski et al.

¹³ Unfortunately, not all ages of all singers and bands could be researched; even using news articles and other sources leads to 19 missing ages. However, the average age for the remaining singers was 30.89, with a standard deviation of 6.71.

(2021) Spotify data could offer valuable insights into measuring music quality. Thus, our study serves as an initial exploration of the explanatory power of both narrative structures and available objective music measurements.

7 Conclusion

To the best of our knowledge, this study represents the first analysis incorporating emotional archetypes of lyrics to investigate their impact on song perception. In doing so, we introduced novel aspects to the existing literature on the influence of characteristics and the contest format on the outcomes of song contests. The analysis suggests that emotional archetypes offer insights into voting behavior and the general perception of music.

As mentioned earlier, it is worthwhile to pursue this avenue further by applying the methodology to a more extensive set of songs and lyrics and conducting controlled laboratory experiments. This expanded exploration can help us better understand the methodology's potential and robustness.

While our study primarily encourages further research, we also recognize practical implications for song contests, particularly for those involved in selecting participating songs. For example, even a marginal increase in points received at the ESC, as observed in our study, can be considered advantageous for the host country of our research, Germany.

Our results suggest that songs do tell stories. Songs inherently convey emotional narratives, and our study highlights how these stories are presented and how they can influence a person's perception of music. While our findings serve as initial indications, they also pave the way for further research in the field of music, thereby offering insights not only from a methodological perspective but also for formulating new research questions.

Appendix

See Tables 8, 9, 10, 11.

Funding Open Access funding enabled and organized by Projekt DEAL.

Table 8 Robustness check Gender: Influence of gender

Variables	(1)	(2)	(3)	(4)	(5)	(6)
	Panel regression	Panel Poisson	Panel logit	Panel regression	Panel Poisson	Panel logit
Duration	0.15** (0.065)	0.053*** (0.018)	-0.19*** (0.045)	0.17** (0.070)	0.058*** (0.019)	-0.17*** (0.044)
Liveness	0.028 (0.50)	-0.11 (0.13)	1.15*** (0.32)	-0.60 (0.49)	-0.26** (0.13)	0.80** (0.35)
Instrumentalness	-0.23 (1.27)	0.77** (0.36)	0.22 (0.85)	0.67 (1.27)	0.99*** (0.36)	0.73 (0.81)
Danceability	-0.58 (0.60)	-0.0065 (0.16)	-1.54*** (0.37)	0.040 (0.69)	0.10 (0.18)	-0.80** (0.40)
Energy	-1.22** (0.44)	-0.19* (0.11)	-1.39*** (0.29)	-1.33** (0.53)	-0.25* (0.14)	-1.04*** (0.32)
Tempo	-0.0056** (0.0021)	-0.0013** (0.00057)	-0.0074*** (0.0016)	-0.0037 (0.0022)	-0.00093 (0.00058)	-0.0063*** (0.0016)
Speechiness	-3.03*** (0.75)	-0.75*** (0.22)	-2.71*** (0.59)	-2.87*** (0.77)	-0.68*** (0.23)	-2.69*** (0.66)
Valence	-1.88*** (0.37)	-0.53*** (0.10)	-1.24*** (0.23)	-1.56*** (0.40)	-0.39*** (0.11)	-1.22*** (0.23)
Happy end	-0.66*** (0.11)	-0.20*** (0.030)	-0.58*** (0.060)	0.52** (0.20)	0.12** (0.053)	0.62*** (0.088)
Peak	-0.056*** (0.016)	-0.011** (0.0050)	-0.030** (0.013)	-0.31 (0.22)	-0.13* (0.068)	-0.13 (0.11)
English	-0.87*** (0.25)	-0.41*** (0.10)	-0.58** (0.26)	-0.65*** (0.21)	-0.21*** (0.060)	-0.026 (0.077)
Slang	-0.775* (0.38)	-0.26*** (0.10)	1.03** (0.42)	0.12 (0.28)	-0.0089 (0.081)	-0.19* (0.11)
Home	1.97*** (0.32)	0.45*** (0.068)	0.55** (0.22)	0.77*** (0.23)	0.18*** (0.057)	0.40** (0.16)

Table 8 (continued)

Variables	(1)	(2)	(3)	(4)	(5)	(6)
	Panel regression	Panel Poisson	Panel logit	Panel regression	Panel Poisson	Panel logit
Start_position	0.32*** (0.017)	0.084*** (0.0058)	0.23*** (0.012)	-0.95*** (0.27)	-0.39*** (0.11)	-0.69*** (0.27)
Start_position ²	0.031*** (0.0038)	0.0055*** (0.0012)	0.021*** (0.0021)	-0.75* (0.37)	-0.24** (0.099)	1.24*** (0.42)
Female	-1.92*** (0.12)	-0.56*** (0.043)	-1.07*** (0.095)	1.97*** (0.33)	0.45*** (0.070)	0.56** (0.22)
Mixed	0.11 (0.099)	0.062** (0.032)	0.24*** (0.073)	0.33*** (0.017)	0.086*** (0.0058)	0.23*** (0.011)
Constant	4.96*** (0.31)			0.032*** (0.0038)	0.0052*** (0.0013)	0.021*** (0.0020)
Observations	2800	2800	2800	-1.84*** (0.14)	-0.54*** (0.049)	-1.03*** (0.10)
Genre dummies	Yes	Yes	Yes	0.33***	0.12***	0.20***
Year dummies	Yes	Yes	Yes	(0.090)	(0.027)	(0.078)
R ²	0.24			3.35*** (0.51)		
				2800	2800	2800
				Yes	Yes	Yes
				Yes	Yes	Yes
				0.24		

Level of significance: *p < 0.1, **p < 0.05, ***p < 0.01

Table 9 Robustness check Gender: all emotional archetypes

	Oedipus	Man in Hole	Tragedy/Riches to Rags	Rags to Riches	Cinderella	Icarus
Panel regression	0.0033 (0.17)	0.58*** (0.14)	- 0.43 (0.26)	- 0.92*** (0.15)	0.076 (0.22)	0.85*** (0.16)
Panel Poisson	0.023 (0.045)	0.16*** (0.039)	- 0.15* (0.077)	- 0.26*** (0.048)	- 0.0078 (0.068)	0.22*** (0.040)
Panel logit	- 0.20*** (0.063)	0.57*** (0.091)	- 0.24** (0.11)	- 0.21*** (0.049)	- 0.37*** (0.099)	0.37*** (0.12)

Level of significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 10 Robustness check Sentiment: Hugging face sentiment analysis

Variables	(1)	(2)	(3)	Variables	(4)	(5)	(6)
	Panel regression	Panel Poisson	Panel logit		Panel regression	Panel Poisson	Panel logit
Duration	0.18** (0.064)	0.048*** (0.016)	- 0.13*** (0.042)	Duration	0.20*** (0.062)	0.054*** (0.016)	- 0.14*** (0.062)
Liveness	0.46 (0.49)	0.062 (0.12)	1.13*** (0.30)	Liveness	0.42 (0.50)	0.044 (0.12)	1.08*** (0.40)
Instrumentalness	1.30 (1.29)	1.04*** (0.34)	0.67 (0.78)	Instrumentalness	1.76 (1.27)	1.18*** (0.33)	0.88 (1.26)
Danceability	- 1.10* (0.63)	- 0.23 (0.17)	- 1.71*** (0.39)	Danceability	- 1.09* (0.61)	- 0.23 (0.16)	- 1.63*** (0.47)
Energy	0.066 (0.45)	0.045 (0.12)	- 0.33 (0.27)	Energy	- 0.022 (0.43)	0.011 (0.12)	- 0.39 (0.36)
Tempo	- 0.0082*** (0.0020)	- 0.0020*** (0.00057)	- 0.0086*** (0.0015)	Tempo	- 0.0084*** (0.0021)	- 0.0021*** (0.00057)	- 0.0088*** (0.0017)
Speechiness	- 2.25*** (0.75)	- 0.63*** (0.21)	- 2.11*** (0.62)	Speechiness	- 2.32*** (0.75)	- 0.66*** (0.21)	- 2.05*** (0.80)
Valence	- 1.48*** (0.38)	- 0.40*** (0.11)	- 0.97*** (0.21)	Valence	- 1.37*** (0.38)	- 0.38*** (0.10)	- 0.97*** (0.25)
Happyend_hf	- 0.57*** (0.17)	- 0.18*** (0.061)	- 0.28*** (0.084)	Sentiment_hf	- 0.26** (0.092)	- 0.059*** (0.026)	0.048 (0.080)
English	- 1.01*** (0.26)	- 0.41*** (0.098)	- 0.60** (0.25)	English	- 1.11*** (0.27)	- 0.44*** (0.10)	- 0.62*** (0.28)
Slang	- 0.50 (0.36)	- 0.089 (0.089)	1.11*** (0.42)	Slang	- 0.65* (0.36)	- 0.16* (0.090)	1.09** (0.48)
Home	1.97***	0.45***	0.54**	Home	1.97***	0.45***	0.54***

Table 10 (continued)

Variables	(1) Panel regression	(2) Panel Poisson	(3) Panel logit	Variables	(4) Panel regression	(5) Panel Poisson	(6) Panel logit
Start_position	(0.29) 0.31***	(0.058) 0.081***	(0.21) 0.21***	Start_position	(0.28) 0.32***	(0.056) 0.082***	(0.19) 0.21***
Start_position ²	(0.018) 0.033***	(0.0059) 0.0060***	(0.012) 0.020***	Start_position ²	(0.018) 0.033***	(0.0060) 0.0060***	(0.012) 0.020***
Constant	(0.0039) 3.43***	(0.0012)	(0.0020)	Constant	(0.0039) 3.15***	(0.0012)	(0.0027)
Observations	(0.35) 2800			Observations	(0.37) 2800		2800
Genre dummies	Yes	Yes	Yes	Genre dummies	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Year dummies	Yes	Yes	Yes
R ²	0.21			R ²	0.21		

Level of significance: *p < 0.1, **p < 0.05, ***p < 0.01

Table 11 Correlation between Spotify data and emotional archetypes

	Liveness	Instrumental- ness	Danceability	Energy	Tempo	Speechiness	Valence
Oedipus	- 0.0636	- 0.0931	0.0466	- 0.0955	0.0971	- 0.1005	0.0171
Man in Hole	0.2251	- 0.0006	- 0.0895	0.0245	- 0.0256	0.0031	- 0.0548
Tragedy	- 0.0997	- 0.0505	0.2713	0.0011	- 0.0155	0.0641	0.2466
Rags to riches	- 0.0983	- 0.0146	- 0.0814	0.0058	0.0382	- 0.0088	0.0155
Cinderella	0.0087	0.0721	0.0692	0.1062	- 0.0512	- 0.0217	- 0.1045
Icarus	0.0135	0.0898	- 0.1315	- 0.0267	- 0.0632	0.0864	- 0.0697

Declarations

Conflict of interest The research was conducted without extra funding.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Al-Beitawi, Z., Salehan, M., & Zhang, S. (2020). What makes a song trend? Cluster analysis of musical attributes for Spotify top trending songs. *Journal of Marketing Development and Competitiveness*, 14(3), 79–91.
- Ali, S. O., & Peynircioğlu, Z. F. (2006). Songs and emotions: Are lyrics and melodies equal partners? *Psychology of Music*, 34(4), 511–534. <https://doi.org/10.1177/0305735606067168>
- Anglada-Tort, M., Masters, N., Steffens, J., North, A., & Müllensiefen, D. (2023). The behavioural economics of music: Systematic review and future directions. *Quarterly Journal of Experimental Psychology*, 76(5), 1177–1194.
- Anglada-Tort, M., & Müllensiefen, D. (2017). The repeated recording illusion: The effects of extrinsic and individual difference factors on musical judgments. *Music Perception An Interdisciplinary Journal*, 35(1), 94–117.
- Arratia, A., Avalos, G., Cabaña, A., Duarte-López, A., & Renedo-Mirambell, M. (2021). Sentiment analysis of financial news: Mechanics and statistics. In *Data Science for Economics and Finance: Methodologies and Applications* (pp. 195–216). Springer International Publishing Cham.
- Assuncao, W. G., Piccolo, L. S., & Zaina, L. A. (2022). Considering emotions and contextual factors in music recommendation: A systematic literature review. *Multimedia Tools and Applications*, 81(6), 8367–8407. <https://doi.org/10.1007/s11042-022-12110-z>
- Barradas, G. T., & Sakka, L. S. (2022). When words matter: A cross-cultural perspective on lyrics and their relationship to musical emotions. *Psychology of Music*, 50(2), 650–669. <https://doi.org/10.1177/03057356211013390>
- Berger, J., & Packard, G. (2018). Are atypical things more popular? *Psychological Science*, 29(7), 1178–1184.
- Berger, W. (2017). Why is this Song Popular? (feat. P-i-t-b-u-l-l-Spotify) <https://medium.com/@albert.w.berger/what-makes-a-song-popular-in-a-certain-country-feat-p%CC%B6%CC%B6%CC%B6%CC%B6%CC%B6%CC%B6%CC%B6-spotify-cd705abc59>

- Berlemann, M., & Thomas, T. (2019). The distance bias in natural disaster reporting—empirical evidence for the United States. *Applied Economics Letters*, 26(12), 1026–1032.
- Berlin, N., Bernard, A., & Fürst, G. (2015). Time spent on new songs: Word-of-mouth and price effects on teenager consumption. *Journal of Cultural Economics*, 39, 205–218.
- Bernhardt, L., Dewenter, R., & Thomas, T. (2023). Measuring partisan media bias in US newscasts from 2001 to 2012. *European Journal of Political Economy*, 78, 102360.
- Berns, G. S., Capra, C. M., Moore, S., & Noussair, C. (2010). Neural mechanisms of the influence of popularity on adolescent ratings of music. *NeuroImage*, 49(3), 2687–2696.
- Bille, T. (2024). The values of cultural goods and cultural capital externalities: State of the art and future research prospects. *Journal of Cultural Economics*, 5, 1–19.
- Borowiecki, K. J. (2017). How are you, my dearest Mozart? Well-being and creativity of three famous composers based on their letters. *Review of Economics and Statistics*, 99(4), 591–605.
- Brand, C. O., Acerbi, A., & Mesoudi, A. (2019). Cultural evolution of emotional expression in 50 years of song lyrics. *Evolutionary Human Sciences*, 1, e11.
- Bruner, G. C. (1990). Music, mood, and marketing. *Journal of Marketing*, 54(4), 94–104.
- Budzinski, O., Gaenssle, S., & Weimar, D. (2023). Disentangling individual biases in jury voting: An empirical analysis of voting behavior in the eurovision song contest. *Ilmenau Economics Discussion Papers*, 28(171), 963.
- Budzinski, O., Kohlschreiber, M., Kuchinke, B. A., & Pannicke, J. (2021). Does music quality matter for audience voters in a music contest? *Creative Industries Journal*, 14(2), 109–129.
- Budzinski, O., & Pannicke, J. (2017a). Culturally biased voting in the Eurovision Song Contest: Do national contests differ? *Journal of Cultural Economics*, 41, 343–378.
- Budzinski, O., & Pannicke, J. (2017b). Do preferences for pop music converge across countries?—empirical evidence from the eurovision song contest. *Creative Industries Journal*, 10(2), 168–187.
- Budzinski, O., & Pannicke, J. (2022). Does popularity matter in a TV song competition? Evidence from a national music contest. *Creative Industries Journal*, 7, 1–18.
- Chin, T., & Rickard, N. S. (2014). Emotion regulation strategy mediates both positive and negative relationships between music uses and well-being. *Psychology of Music*, 42(5), 692–713.
- Dang, T. V., & Xu, Z. (2018). Market sentiment and innovation activities. *Journal of Financial and Quantitative Analysis*, 53(3), 1135–1161.
- De-Lucio, J., & Palomeque, M. (2023). Music preferences as an instrument of emotional self-regulation along the business cycle. *Journal of Cultural Economics*, 47(2), 181–204. <https://doi.org/10.1007/s10824-022-09454-7>
- Del Vecchio, M., Kharlamov, A., Parry, G., & Pogrebna, G. (2021). Improving productivity in Hollywood with data science: Using emotional arcs of movies to drive product and service innovation in entertainment industries. *Journal of the Operational Research Society*, 72(5), 1110–1137.
- DeWall, C. N., Pond, R. S., Jr., Campbell, W. K., & Twenge, J. M. (2011). Tuning in to psychological change: Linguistic markers of psychological traits and emotions over time in popular US song lyrics. *Psychology of Aesthetics, Creativity, and the Arts*, 5(3), 200.
- Dewenter, R., Linder, M., & Thomas, T. (2019). Can media drive the electorate? The impact of media coverage on voting intentions. *European Journal of Political Economy*, 58, 245–261.
- Dierckx, T., Davis, J., & Schoutens, W. (2021). Quantifying news narratives to predict movements in market risk. *Data Science for Economics and Finance*, 265, 2589.
- Dodds, P. S., & Danforth, C. M. (2010). Measuring the happiness of large-scale written expression: Songs, blogs, and presidents. *Journal of Happiness Studies*, 11, 441–456.
- El Harbi, S., Grolleau, G., & Bekir, I. (2014). Substituting piracy with a pay-what-you-want option: Does it make sense? *European Journal of Law and Economics*, 37, 277–297.
- Fairclough, S. H., van der Zwaag, M., Spiridon, E., & Westerink, J. (2014). Effects of mood induction via music on cardiovascular measures of negative emotion during simulated driving. *Physiology & Behavior*, 129, 173–180.
- Fischer, K., & Haucap, J. (2021). Does crowd support drive the home advantage in professional football? Evidence from German ghost games during the COVID-19 pandemic. *Journal of Sports Economics*, 22(8), 982–1008.
- Flôres, R. G., & Ginsburgh, V. A. (1996). The Queen Elisabeth musical competition: How fair is the final ranking? *Journal of the Royal Statistical Society: Series D (The Statistician)*, 45(1), 97–104.
- Fried, C. B. (2003). Stereotypes of music fans: Are rap and heavy metal fans a danger to themselves or others. *Journal of Media Psychology*, 8(3), 1–27.

- Fried, R., & Berkowitz, L. (1979). Music hath charms... and can influence helpfulness 1. *Journal of Applied Social Psychology*, 9(3), 199–208.
- Ginsburgh, V., & Noury, A. G. (2008). The Eurovision song contest. Is voting political or cultural? *European Journal of Political Economy*, 24(1), 41–52.
- Ginsburgh, V. A., & Van Ours, J. C. (2003). Expert opinion and compensation: Evidence from a musical competition. *American Economic Review*, 93(1), 289–296.
- Guhr, O., Schumann, A.-K., Bahrmann, F., & Böhme, H. J. (2020). Training a broad-coverage german sentiment classification model for dialog systems. *European Language Resources Association*, 5, 1620–1625.
- Hamilton, M., & Pearce, M. (2024). Trajectories and revolutions in popular melody based on US charts from 1950 to 2023. *Scientific Reports*, 14(1), 14749.
- Hashim, M. J., Kannan, K. N., Maximiano, S., & Ulmer, J. R. (2014). Digital piracy, teens, and the source of advice: An experimental study. *Journal of Management Information Systems*, 31(2), 211–244.
- Heggli, O. A., Stupacher, J., & Vuust, P. (2021). Diurnal fluctuations in musical preference. *Royal Society Open Science*, 8(11), 210885.
- Henard, D. H., & Rossetti, C. L. (2014). All you need is love?: Communication insights from pop music's number-one hits. *Journal of Advertising Research*, 54(2), 178–191.
- Jamdar, A., Abraham, J., Khanna, K., & Dubey, R. (2015). Emotion analysis of songs based on lyrical and audio features. *International Journal of Artificial Intelligence & Applications*, 6(3), 71.
- Jones, M. B. (2013). The home advantage in individual sports: An augmented review. *Psychology of Sport and Exercise*, 14(3), 397–404.
- Kamekhosh, I., Bonnin, G., & Jannach, D. (2020). Effects of recommendations on the playlist creation behavior of users. *User Modeling and User-Adapted Interaction*, 30, 285–322.
- Kenealy, P. (1988). Validation of a music mood induction procedure: Some preliminary findings. *Cognition & Emotion*, 2(1), 41–48.
- Liebowitz, S. J. (2006). File sharing: Creative destruction or just plain destruction? *The Journal of Law and Economics*, 49(1), 1–28.
- Liu, M., Hu, X., & Schedl, M. (2018). The relation of culture, socio-economics, and friendship to music preferences: A large-scale, cross-country study. *PLoS ONE*, 14(2), e0212495.
- Lozon, J., & Bensimon, M. (2014). Music misuse: A review of the personal and collective roles of “problem music.” *Aggression and Violent Behavior*, 19(3), 207–218.
- Mantzaris, A. V., Rein, S. R., & Hopkins, A. D. (2018). Preference and neglect amongst countries in the Eurovision Song Contest. *Journal of Computational Social Science*, 1, 377–390.
- May, J. L., & Hamilton, P. A. (1980). Effects of musically evoked affect on women's interpersonal attraction toward and perceptual judgments of physical attractiveness of men. *Motivation and Emotion*, 4, 217–228.
- Mitchell, W. B., Dibartolo, P. M., Brown, T. A., & Barlow, D. H. (1998). Effects of positive and negative mood on sexual arousal in sexually functional males. *Archives of Sexual Behavior*, 27, 197–207.
- Napier, K., & Shamir, L. (2018). Quantitative sentiment analysis of lyrics in popular music. *Journal of Popular Music Studies*, 30(4), 161–176.
- Nattiez, J.-J. (1990). Can one speak of narrativity in music? *Journal of the Royal Musical Association*, 115(2), 240–257.
- Negus, K., & Astor, P. (2015). Songwriters and song lyrics: Architecture, ambiguity and repetition. *Popular Music*, 34(2), 226–244.
- North, A. C., & Hargreaves, D. J. (2005). Brief report: Labelling effects on the perceived deleterious consequences of pop music listening. *Journal of Adolescence*, 28(3), 433–440.
- North, A. C., Tarrant, M., & Hargreaves, D. J. (2004). The effects of music on helping behavior: A field study. *Environment and Behavior*, 36(2), 266–275.
- Oberholzer-Gee, F., & Strumpf, K. (2010). File sharing and copyright. *Innovation Policy and the Economy*, 10(1), 19–55.
- Omuya, E. O., Okeyo, G., & Kimwele, M. (2023). Sentiment analysis on social media tweets using dimensionality reduction and natural language processing. *Engineering Reports*, 5(3), e12579.
- Oxoby, R. J. (2009). On the efficiency of AC/DC: Bon Scott versus Brian Johnson. *Economic Inquiry*, 47(3), 598–602.
- Palomeque, M., & De-Lucio, J. (2024). The soundtrack of a crisis: More positive music preferences during economic and social adversity. *Journal of Happiness Studies*, 25(5), 1–24.
- Pannese, A., Rappaz, M.-A., & Grandjean, D. (2016). Metaphor and music emotion: Ancient views and future directions. *Consciousness and Cognition*, 44, 61–71.
- Pannicke, J. (2016). Abstimmungsverhalten im Bundesvision Song Contest. *Zeitschrift Für Kulturmanagement*, 2(2), 39–66.

- Parada-Cabaleiro, E., Mayerl, M., Brandl, S., Skowron, M., Schedl, M., Lex, E., & Zangerle, E. (2024). Song lyrics have become simpler and more repetitive over the last five decades. *Scientific Reports*, *14*(1), 5531.
- Park, M., Thom, J., Mennicken, S., Cramer, H., & Macy, M. (2019). Global music streaming data reveal diurnal and seasonal patterns of affective preference. *Nature Human Behaviour*, *3*(3), 230–236. <https://doi.org/10.1038/s41562-018-0508-z>
- Pignatello, M. F., Camp, C. J., & Rasar, L. A. (1986). Musical mood induction: An alternative to the Velten technique. *Journal of Abnormal Psychology*, *95*(3), 295.
- Pollard, R. (1986). Home advantage in soccer: A retrospective analysis. *Journal of Sports Sciences*, *4*(3), 237–248.
- Pollard, R. (2006). Worldwide regional variations in home advantage in association football. *Journal of Sports Sciences*, *24*(3), 231–240.
- Reagan, A. J., Mitchell, L., Kiley, D., Danforth, C. M., & Dodds, P. S. (2016). The emotional arcs of stories are dominated by six basic shapes. *EPJ Data Science*, *5*(1), 1–12.
- Regner, T., & Barria, J. A. (2009). Do consumers pay voluntarily? The case of online music. *Journal of Economic Behavior & Organization*, *71*(2), 395–406.
- Roberts, K. R., Dimsdale, J., East, P., & Friedman, L. (1998). Adolescent emotional response to music and its relationship to risk-taking behaviors. *Journal of Adolescent Health*, *23*(1), 49–54.
- Rodríguez-Ibáñez, M., Casáñez-Ventura, A., Castejón-Mateos, F., & Cuenca-Jiménez, P.-M. (2023). A review on sentiment analysis from social media platforms. *Expert Systems with Applications*, *223*, 119862.
- Rozin, A., Rozin, P., & Goldberg, E. (2004). The feeling of music past: How listeners remember musical affect. *Music Perception*, *22*(1), 15–39.
- Sakka, L. S., & Juslin, P. N. (2018). Emotional reactions to music in depressed individuals. *Psychology of Music*, *46*(6), 862–880.
- Saura, J. R., Palacios-Marqués, D., & Ribeiro-Soriano, D. (2023). Exploring the boundaries of open innovation: Evidence from social media mining. *Technovation*, *119*, 102447.
- Schäfer, T., Zimmermann, D., & Sedlmeier, P. (2014). How we remember the emotional intensity of past musical experiences. *Frontiers in Psychology*, *5*, 911.
- Schüller, D., Tauchmann, H., Upmann, T., & Weimar, D. (2014). Pro-social behavior in the TV show “Come Dine With Me”: An empirical investigation. *Journal of Economic Psychology*, *45*, 44–55. <https://doi.org/10.1016/j.joep.2014.07.005>
- Schulze, G. G. (2020). Superstars. In *Handbook of Cultural Economics, Third Edition* (pp. 485–493). Edward Elgar Publishing.
- Sciandra, M., & Spera, I. C. (2022). A model-based approach to Spotify data analysis: A Beta GLMM. *Journal of Applied Statistics*, *49*(1), 214–229.
- Shiller, R. J. (2017). Narrative economics. *American Economic Review*, *107*(4), 967–1004.
- Shiller, R. J. (2020). *Narrative economics: How stories go viral and drive major economic events*. Princeton University Press.
- Towse, R., & Hernández, T. N. (2020). *Handbook of cultural economics*. Edward Elgar Publishing.
- Vaitl, D., Vehrs, W., & Sternagel, S. (1993). Prompts-leitmotif-emotion: Play it again, Richard Wagner. *The structure of emotion: Psychophysiological, cognitive, and clinical aspects*, 169–189.
- Waldfogel, J. (2012). Music piracy and its effects on demand, supply, and welfare. *Innovation Policy and the Economy*, *12*(1), 91–110.
- Waldfogel, J. (2019). *Digital Renaissance: What data and economics tell us about the future of popular culture*. Princeton University Press.
- Warburton, W. A., Mohi, S., Sweller, N., Tarabay, C., Spencer, L., & Olsen, K. (2024). Violent and prosocial music: Evidence for the impact of lyrics and musical tone on aggressive thoughts, feelings, and behaviors. *Aggressive Behavior*, *50*(3), e22148.
- Witvliet, C. V., & Vrana, S. R. (2007). Play it again Sam: Repeated exposure to emotionally evocative music polarises liking and smiling responses, and influences other affective reports, facial EMG, and heart rate. *Cognition and Emotion*, *21*(1), 3–25.
- Wood, J. V., Saltzberg, J. A., & Goldsamt, L. A. (1990). Does affect induce self-focused attention? *Journal of Personality and Social Psychology*, *58*(5), 899.
- Yair, G. (1995). ‘Unite Unite Europe’ The political and cultural structures of Europe as reflected in the Eurovision Song Contest. *Social Networks*, *17*(2), 147–161.