First-mover advantage in music

Oleg Sobchuk, Mason Youngblood, Olivier Morin

[re-submitted after revisions to *EPJ Data Science*]

Abstract. Why do some songs and musicians become successful while others do not? We show that one of the reasons may be the "first-mover advantage": artists that stand at the foundation of new music genres tend to be more successful than those who join these genres later on. To test this hypothesis, we have analyzed a massive dataset of over 920,000 songs, including 110 music genres: 10 chosen intentionally and preregistered, and 100 chosen randomly. For this, we collected the data from two music services: Spotify, which provides detailed information about songs' success (the precise number of times each song was listened to), and Every Noise at Once, which provides detailed genre tags for musicians. 91 genres, out of 110, show the first-mover advantage – clearly suggesting that it is an important mechanism in music success and evolution.

Keywords: first-mover advantage, innovation, success, music, genre.

Introduction. What is the first-mover advantage – and is it present in music?

Gustav Elijah Åhr died of fentanyl overdose in 2017, before reaching 22 years old. By this time, Åhr, better known as Lil Peep, was one of the most promising hip hop artists in the United States. A year later, his posthumous album *Come Over When You're Sober, Pt. 2* debuted in the top-5 of the US music chart *Billboard 200* – above Lady Gaga, Muse, and Taylor Swift. Lil Peep had no musical education, no connections in the musical industry, and no wealth. His story is not uncommon: in fact, Lil Peep's two contemporaries – XXXTentacion and Juice Wrld – achieved even bigger fame, being ranked at number one in the charts, before 21 years old. How is this possible? How could these young, unprivileged men become so successful so rapidly?

This question is a part of a more general question: **why do some artworks – songs, books, movies, paintings – become successful while others do not?** The emerging "science of culture" offers us two contesting explanations. First explanation: success in the arts results from *luck* and the related effects that amplify it, like the rich-get-richer effect (Barabási, 2018; Liu et al., 2018; Salganik et al., 2006). In this view, success – especially extreme success – is random. Second explanation: success in the arts is largely meritocratic and stems from the *quality* of artworks (Dubourg & Baumard, 2022; MacCallum et al., 2012; Ravignani et al., 2018; Varnum et al., 2021). In this view, success is nothing but random: it is predictable and deserved.

We argue that a hypothesis borrowed from economics – **the first-mover advantage** – reconciles these polar explanations. It suggests that individuals that move to a new market niche early on ("first movers") obtain advantages that may lead to larger success, compared to those who move to this niche later (Kerin et al., 1992; Mueller, 1997). First movers enjoy a temporary near-monopoly: since they enter a niche early, they have little to no competition, and so they can charge larger prices and spend more time building a loyal customer base. Think of the iPhone: in 2007, Apple, Inc. was the first company to enter a new market niche for smartphones, and today the newer versions of iPhone dominate the smartphone market worldwide. Arguably, the first-mover advantage is also present in science, where the early scholars of a new academic sub-discipline enjoy a larger number of citations (Dodds et al., 2017; Newman, 2009; Wang & Barabási, 2021).

The hypothesis of first-mover advantage explains how **luck and quality can co-exist**: the first movers, often by sheer luck, make an innovation – and thus become more "interesting" (i.e., have higher perceived quality) than their competitors, for a limited period of time. Their competitors, making similar cultural products, which aren't necessarily worse (songs in the same genre, books in the same style, etc.), are already disadvantaged – simply because they are not "new". Being new is important in any market, but it may be especially important in the popular arts, which operate in the "marketplace of attention" (Webster, 2014), where novel products usually have an edge over old products, simply by being new. The temporary advantage of the first movers, based on their temporary novelty, can become more permanent due to runaway cultural processes such as the rich-get-richer effect and other forms of frequency-based selection (Newberry & Plotkin, 2022): common forces that work against market diversity.

We suggest that **the first-mover advantage is present in artistic domains**, such as music, and can explain the success of artists. The artists and music producers who recognize the hidden potential of a new artistic technique, genre, or style, have bigger chances of reaching success. Having an artistic innovation that your competitors do not have or cannot quickly acquire may become advantageous on the winner-take-all artistic market (Frank, 2016). In this way, artistic innovations may be regarded as similar to technological innovations (Rossman, 2015). Inventors or early adopters of beneficial innovations thus have an upper hand over their competitors.

Musical innovations can be of different sorts. Some are technological: a new musical instrument or another *technology* that brings the previously unavailable possibilities for making music. The artists that are particularly quick to adopt such new technologies often have an advantage. For example, certain innovative songs get remembered as the first uses (or the first *tasteful* uses) of a particular music technology. Say, the German synthpop band Kraftwerk became an early adopter of a vocoder – a novel tool for changing the voice and making it sound "robotic". Kraftwerk was one of the bands that jumped on the opportunity offered by this new tool. Many technologies had similar effects: from pianoforte and violin to synthesisers and autotune (Burgess, 2014; Campbell et al., 2004). But innovations in music are not only technological. Many have to do not with the invention of new musical tools, but with the creative ways of using the already existing tools. Not technologies but *techniques* of music-making. For example, a vocal technique called "growling" became a novel way of singing in heavy metal music, becoming the key feature of its subgenre – death metal. Within the existing domain of metal music, growling became something new and defined a new subgenre of metal.

Many of these innovations are captured by **music genres**. "Synthpop", "death metal", and other genre labels are created dynamically – by listeners, music critics, or artists themselves – to signal that a particular kind of music is new and different from what was heard (by that population of listeners) before. In other words, it allows capturing musical innovations. These innovations don't have to be novel in absolute terms: they only have to be distinctive enough in a particular community in a particular period of time.

Let's return to our initial example. Lil Peep, as well as XXXTentacion and Juice Wrld, are recognized among the founders of a new genre called "emo-rap", which suddenly became hyper-popular in the mid-2000s. It combined some of the already existing components – personal "emo" lyrics, trap beats, and rapping – into an entirely new blend. Could it be that these young artists became so famous because they were among the first-movers of this genre? Is there the first-mover advantage in music? This is the main question of our study.

Song data

Music streaming apps – Spotify, Apple Music, Deezer, and others – offer an unprecedented opportunity for testing the first-mover advantage hypothesis, for several reasons. First, they have a high-resolution metric of success – precise "playcounts": the total number of times each song was listened to on the platform. Second, they hold exhaustive repertoires of music, both mainstream and underground, containing the vast majority of music released commercially (albeit the record for pre-2000 music seems to be patchy). Third, they provide the means for obtaining detailed genre tags. To test the first-mover advantage in music, we have chosen two databases: Spotify (https://open.spotify.com/), which provides detailed information on *success* of songs, and Every Noise at Once (https://everynoise.com/), which provides detailed *genre tags* of songs. These tags are fairly detailed clusters of songs: much more detailed than the generic labels like "jazz", "rock", or "pop". Instead, Every Noise website contains populations of songs organized into much more specific clusters: "juzz funk", "psychedelic rock", or "Korean electropop". These clusters were generated automatically, based on the behavior of Spotify users as well as the features of songs themselves. Importantly for our study, these specific genre labels are much more likely to reflect some genuine musical innovations, compared to generic genre labels ("rock", "pop", etc.). Spotify offers two **metrics of success**: its own proprietary "popularity" metric (a score between 1 and 100), generated by an undisclosed algorithm, and a more basic, but also more transparent "playcount" metric, which is simply the total number of times a song was listened to on Spotify. We decided to use song playcount as our metric of success, due to its transparency.

Importantly, artist IDs on Every Noise at Once are the same as on Spotify. Thus, they can be used to collect all available data about these artists from the Spotify API, including their number of followers, overall popularity, all genres they are associated with, and all of their releases – with format, release date, etc. (Spotify, n.d.). We used a third-party API based on Librespot (triph, 2020/2022), an open-source client library for Spotify, for collecting the raw playcounts of songs in a variety of genres. In total, we collected three samples of songs.

Three datasets

10-genre dataset. The first one is a dataset of 10 genres, manually picked by ourselves and preregistered on Open Science Framework's website (**link, anonymized for peer-review**). We did this because we had doubts about the quality of the genre tags on Every Noise: they might include artists and songs that do not belong to these genres, by mistake. Thus, we selected ten genres that one of the authors has firsthand experience with, from over a decade of DJing and participating in community radio, and manually checked all the artists belonging to these genres. Another feature of this dataset is that it includes only the genres that appeared after the year 2000. Why? We did this to overcome the *preservation bias*: more recent cultural data is usually better preserved, and the genres that appeared after the year 2000 would be more complete than genres that appeared, say, in the year 1900. And our goal was to have as complete genre populations as possible. This sample includes the following genres: *blackgaze, chillwave, deconstructed club, drill, emo rap, funk ostentação, future bass, gengetone, gqom,* and *grime* (see **Table S1** for the details on each genre). These 10 genres represent various broad strands of music – like rap, electronic, or metal – and various geographies: Brazil, Kenya, South Africa, United Kingdom, and others. We did not include the songs with very low playcounts (<1000) because this information is not provided by Spotify. Overall, we collected 166,364 songs belonging to these 10 genres. The historical dynamics of these genres is visualized on **Figure 1C**.

100-genre dataset. We have also collected a dataset of 100 randomly sampled genres – to look at the bigger picture, despite the possible problems with patchy Spotify coverage or imperfect Every Noise genre tags. Since the data on Spotify from the early 20th century is extremely sparse, we included only the songs released after 1950. (And, to match our 10-genre dataset, we did not include songs released after 2021.) The yearly number of songs released in each genre is shown on **Figure 1B**. Most of the genre labels in this dataset are rather specific and correspond to actual historical – and sometimes local – kinds of music: *cold wave*, *funktronica*, *tech trance*, *Ukrainian black metal*, etc. Others may be less meaningful – representing either very broad types of sound (e.g., *polyphony*) or particular musical instruments in the context of some genre (e.g., *jazz clarinet* or *metal guitar*). Similarly to the 10-genre sample, we excluded all the songs with tiny playcounts (<1000), as these exact numbers are unknown. We ended up having 525,647 songs belonging to 100 genres. The complete list of genres and descriptive statistics about them can be found in **Table S2**.

Mixed-genre dataset. We also collected a third dataset, which is entirely unstructured. We randomly drew, without replacement, 232,057 songs from Spotify, belonging to any period or genre imaginable. Importantly, in this sample each artist has *one and only one song*. We needed this dataset to account for biases in the data, as will be explained further.

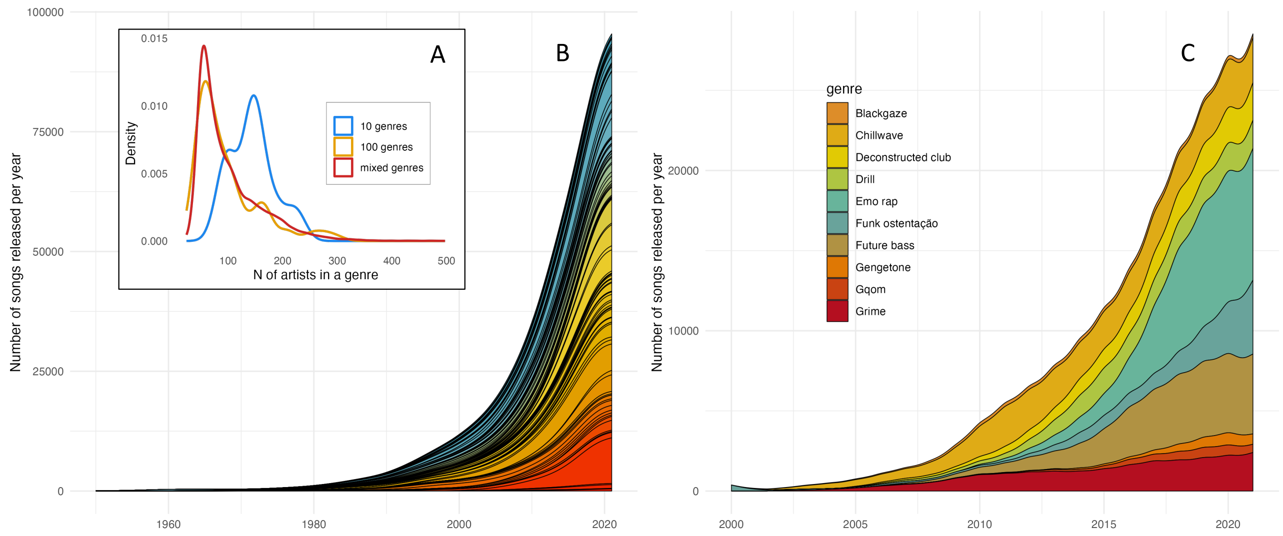

Figure 1. A: Density curves showing the sizes of genre populations in three samples used in our analysis. While the median population size of our 10 preregistered genres is somewhat larger (144 artists per genre), the median population of genre sizes in the random sample of 100 genres (77 artists) is almost the same as the median across all the genres on Every Noise at Once (80 artists). **B:** Adoption curves of 100 random music genres. **C:** Adoption curves of 10 preregistered music genres.

Possible data biases

We must account for several biases that are likely to hide in the datasets. On Spotify, more recent songs have much larger cumulative playcounts (**Fig. S1**). Of course, this is not because more recent songs are objectively more successful than older songs. Instead, the larger playcounts in recent years reflects the growth of Spotify's audience. In the past, music listeners used other means for accessing music: radio, vinyl records, cassette and CD players, or early music sharing services like Napster. Therefore, the success of earlier songs is heavily underestimated. Besides this recency bias, there is a different problem: artists who started their careers in the early 2000s were arguably disadvantaged compared to those who started in the 1990s or 2010s. 1990s artists were backed by a thriving recording industry, which underwent a collapse due to the rise of streaming, then progressively regained some stability (Waldfogel, 2018). Our analysis thus has to contend with two problems. First, the absolute cumulative Spotify playcount for a song from 2000 is likely to be lower than that of a song from 2010, even if we assume that both songs were equally successful, because of the recency bias in our data (the song from 2000 was listened to many times, but not on Spotify). Second, historical fluctuations in the music industry's potential make a song's success intrinsically dependent on time, regardless of any bias in our ability to capture it. So, we did two forms of debiasing – by normalizing and by resampling – described below.

Debiasing via standardizing. To account for the recency bias, we standardize the playcount of each song –by considering the song's relative success compared to all the other songs from the same dataset released in the same year. This was done to make sure that we measure the popularity of each song only compared to the songs of the same year, which helps us overcome the issue that earlier songs have more time to be listened to – and thus, to accumulate playcount – than later songs. We transform the playcounts using the "robust (or modified) Z-score" (Iglewicz & Hoaglin, 1993), which is a metric similar to the commonly used Z-score, but it uses median instead of mean, and median absolute deviation instead of the standard deviation:

$$Modified\ Z\ score_{song} = \frac{0.6745 \cdot (\log(playcount_{song_{year}}) - \log(\widetilde{playcount}_{year}))}{\mathrm{median}_{song}\left\{ |\log(playcount_{song_{year}}) - \log(\widetilde{playcount}_{year})| \right\}}$$

The reason for using median-based Z-score is that the median values in our data, not the mean values, show more extreme recency bias (see **Fig. S2**). Prior to standardizing, playcount values were log-transformed to account for the highly uneven distribution of playcounts in each year: a small number of songs are "superhits", their playcounts are hundreds of times larger than average. Note that standardizing the playcounts within years does not remove time trends in success from the data: it allows us to capture time trends in *relative* success.

Debiasing via resampling. To tackle the issue of historical fluctuations, we have collected an additional "mixed-genre dataset" of 232,057 randomly chosen songs. From this mixed-genre dataset, we sequentially sample song corpora that are similar to the actual genre populations in our 10-genre and 100-genre samples: they have the exact same size and the same yearly distribution of releases. However, these sampled corpora are not genres, they are "pseudo-genres": songs in them do not belong to any single genre (they are a mixture of an unknown number of random genres), and thus, according to our hypothesis, should not exhibit the first-mover advantage. This is because, in any genre population of songs, early songs came out early in the history of their genre while other songs came late; but in a pseudo-genre sampled from the mixed-genre dataset, some early songs may be late relative to the history of their own genre, and vice-versa. In a mixed-genre pseudo-genre, we do not predict any first-mover advantage. As an example of our use of these pseudo-genres, consider **Figure 2**: blue dots represent all songs in one genre – *chillwave* (A and C); dots of random colors (B and D) represent all songs in a sampled mixed-genre dataset. If, for each of the two datasets, we regress a song's total Spotify playcounts against the date the song was released (see **Methods**), we observe that the relationship between time and playcount for the *chillwave* genre is more negative than the same relationship in the mixed-genre sample. We do not expect any general relationship between a song's release date and its success: it could be negative, positive, or neutral, depending on historical fluctuations in the music market. However, inside a given genre, the first-mover advantage should favor early songs compared to late ones, all else being equal.

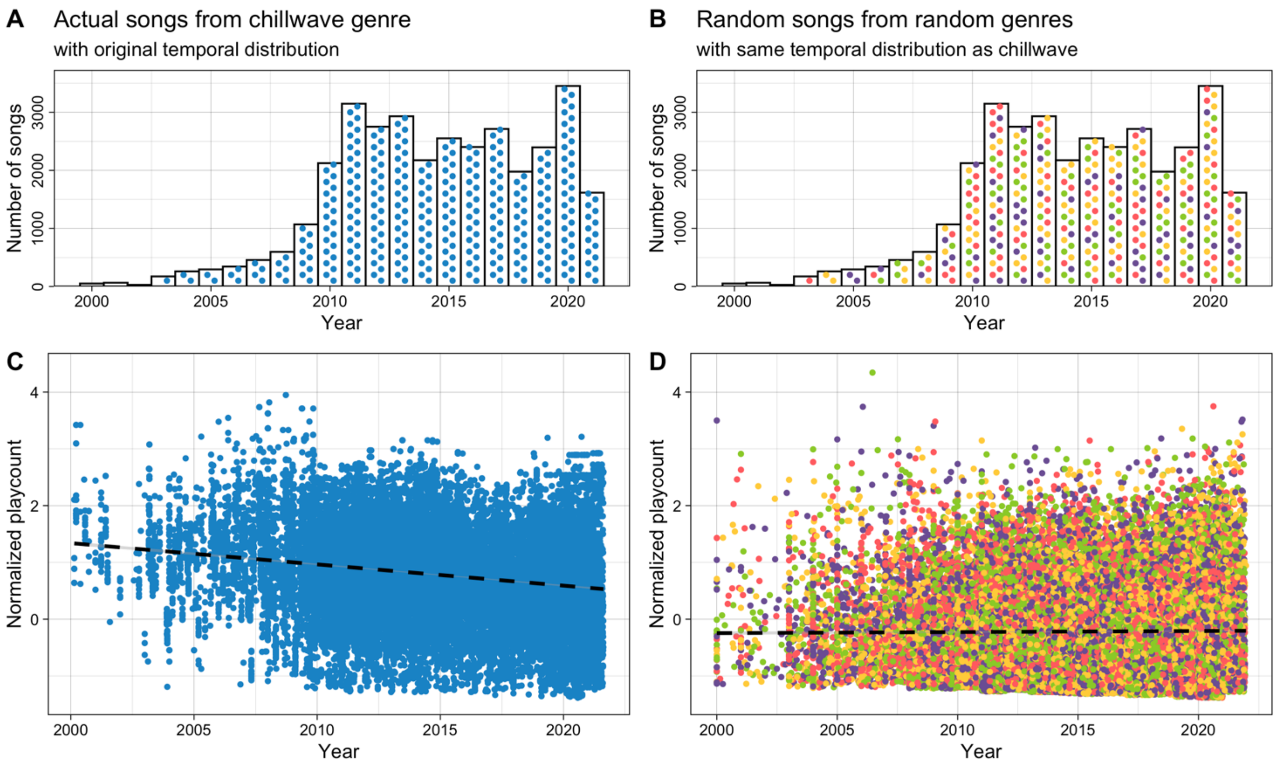

Figure 2. An illustration of our approach to creating "pseudo-genres", sampled from the mixed-genre dataset. **A:** Yearly distribution of songs in one music genre – *chillwave*. **B:** The exact same yearly distribution of songs, drawn from the mixed-genre dataset. To stress that these songs belong to different genres, they are painted in random colors. **C:** The number of times each song was listened to (i.e., its playcount), normalized by the median playcount of all songs from the same year in the respective dataset, across all genres. Dashed line indicates the regression line. **D:** Normalized playcounts of songs from the particular mixed-genre sample (various colors exemplify the multitude of genres included in this sample). The relationship between each song's release date (x axis) and playcount in the chillwave genre (**C**) is less positive than in the mixed-genre sample (**D**) – as predicted. This procedure is repeated for each genre 1000 times.

Hypothesis

Given our data, the biases that are likely to plague it, and our means of removing them, we formulated the following hypothesis. The relationship between a song's *playcount* (log-transformed and normalized) and its *year of release* should be **more negative** in the genre corpora than in the equivalent pseudo-genres, sampled from the mixed-genre dataset. This is because the actual genres should exhibit the first-mover advantage, while the pseudo-genres should not. We preregistered this prediction on the Open Science Framework's website (**link, anonymized for peer-review**).

Linear models

To test this hypothesis, for both the 10-genre dataset and the 100-genre dataset we built a linear mixed-effects model predicting the *playcount* of each *song* by its *year of release* (as a fixed effect) and *artist* (as a random effect). We added *artist* as random effect because we want to exclude the possibility that the results are driven by a small number of highly prolific artists. For each genre, a separate model was run – this one:

$$playcount_{song} = \alpha_{artist[song]} + \beta \cdot year\ of\ release_{song} + \epsilon_{song}$$

In the mixed-genre dataset, each artist is represented by one and only one song – that is, we have already controlled for artists' productivity: no musician is more prolific than any other musician. And so, for this sample we no longer need a random effect of artist. For each mixed-genre sample we used a simple linear model predicting the *playcount* of each *song* by its *year of release*. For each pseudo-genre, a separate model was run, which looks like this:

$$playcount_{song} = \alpha + \beta \cdot year\ of\ release_{song} + \epsilon_{song}$$

We have pre-registered these models, as well as our predictions and data, on the Open Science Framework's website. Both linear models were run in R using the package *lme4* (Bates et al., 2015).

For each genre, we sample 1000 pseudo-genres from the mixed-genre dataset and then compare the relationship between *year of release* and *playcount* in the actual genre with the distribution of estimates of the relationship between *year of release* and *playcount* in 1000 mixed-genre samples.

Results and discussion

The results of analysis are shown on **Figure 3**. Image on the left (A) shows the results for 10-genre sample. The blue dots are effects of song release on the success of songs in this genre. The yellow distributions are effects of 1000 sampled pseudo-genres drawn from the mixed-genre dataset, as described above. We predicted that those effects in the actual genres should be more negative than in pseudo-genres, because actual genres should exhibit the first-mover advantage. In the 10-genre sample (panel B), 8 genres support our hypothesis. Image on the right (C) shows the results for the 100-genre sample. Similarly, blue dots show the effects in the actual genres, while yellow dots with whiskers – effects for 1000 pseudo-genres, drawn from the mixed-genre dataset. Here, our prediction is again supported for the vast majority of genres: 83 genres, out of 100, show the first-mover advantage.

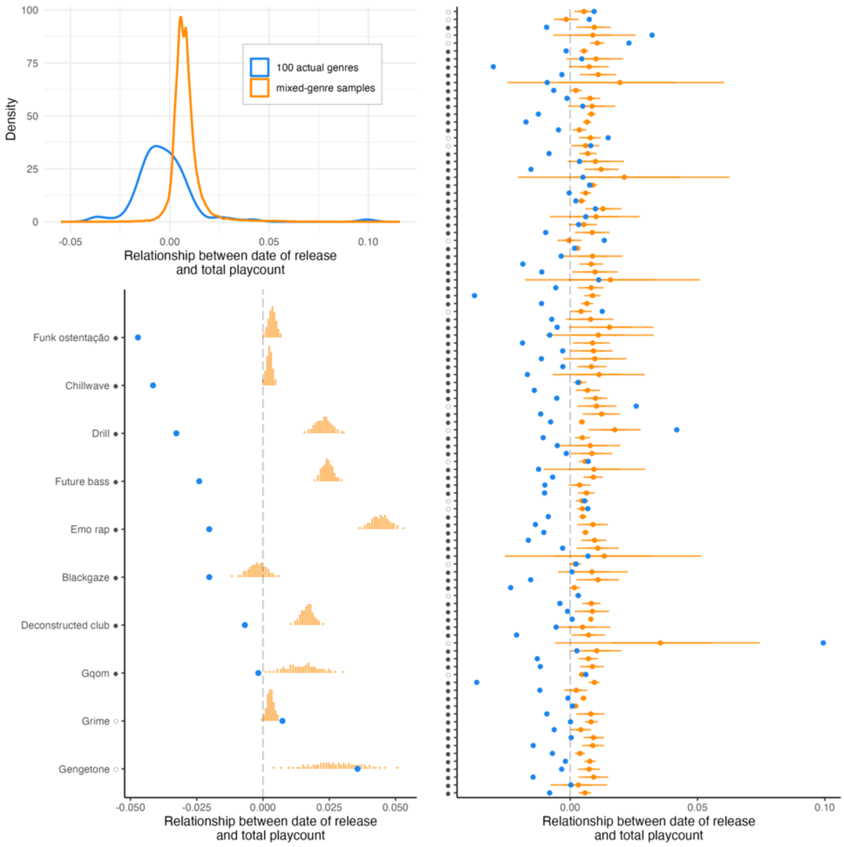

Figure 3. A: Density curves of the relationship between year of release and songs' success in 100 random genres and in the samples from the mixed-genre dataset. **B:** Blue dots: the relationship between time and total playcount for each of the 10 preregistered genres. Yellow distributions: the same relationships, but for the samples drawn from the mixed-genre corpus of songs (1000 samples per each actual genre). Panel **C** shows the same, but for 100 randomly picked genres. On panels **B** and **C**, full circles (●) on y-axis mean that the first-mover hypothesis is confirmed, empty circles (○) – that it is not.

In our mixed-genre dataset, the overall relationship between date of songs' release and their success is somewhat positive (yellow curve on **Figure 3A**). It reflects the already described recency bias: recent music is consumed more often than older music. The first mover effect shows up despite this bias, pushing earlier songs "up".

19 genres, out of 110, do not show the first-mover advantage (2 in our 10-genre dataset, and 17 in the 100-genre dataset). There are a variety of reasons why this could be. As an example, take *grime* – a peculiar combination of hip hop and electronic music that rose from the suburbs of London in the early 2000s. There are many good – or rather bad – reasons why the first grime artists did not make a big breakthrough: grime artists were predominantly black, and in the eyes of the police their music was associated with violence and crime. And so, London police was actively cancelling grime concerts and closing the clubs playing grime. Grime music could only spread through pirate radio stations and mixtapes. The artists were often making their beats using simple software on video game consoles (such as *Music 2000* on PlayStation), instead of professional instruments, and self-releasing their albums on the Internet, instead of relying on professional studios (Fatsis, 2019; Hancox, 2019). This situation lasted until the early 2010s, when the genre was finally "discovered" by the mainstream music producers and consumers. However, for many of the very first movers of the genre, it was already too late: who knows what levels of success they could have reached if their music hadn't been actively suppressed. A closer look at other genres might show similar stories – or other reasons why this or that genre did not support our hypothesis.

First-mover advantage in other arts. Should we expect the first-mover advantage in other artistic domains? There is anecdotal evidence that the inventors of new genres in literary fiction and in movies enjoy extreme success too. For example, Agatha Christie – one of the recognized founders of "classical" crime fiction – is also one of the best-selling authors ever (Knight, 2010; Sobchuk, 2018). Similarly, William Gibson's novel *Neuromancer* – a canonical work in the genre of cyberpunk – is also one of the earliest books in this strand of science fiction (Brouillette, 2002). In films, the cult classic *The Blair Witch Project* is the first recognized member of the highly successful genre of found-footage horror fiction (Heller-Nicholas, 2014). The analytical approach taken in our study can be used for studying similar datasets of books, films, and other cultural artifacts. However, doing so would require high-resolution information about genres, dates of publication, and popularity – something that exists for songs, thanks to Spotify and Every Noise at Once, but may be hard to find for books or movies. Testing the first-mover advantage in other arts remains a challenge.

One might ask: **what is the role of algorithms in the first-mover advantage?** We have used data that comes from a platform – Spotify – that uses non-transparent algorithms. Could it be that the effect we have found is simply an artifact of "machine culture" (Brinkmann et al., 2023) that we are surrounded by rather than "organic" listening behavior? It has been shown that Spotify's recommendation algorithm decreases the diversity of one's exposure to music (Anderson et al., 2020). It is indeed possible that Spotify's recommendation system alters the extent of the first-mover advantage, making it stronger or weaker than it would have been without the algorithm. Still, we think that algorithms are unlikely to cause the first-mover advantage where there is none: algorithmic recommendation systems are built based on human preferences and are trying to suggest music that will be pleasant to humans. But the extent of algorithmic impact on the first-mover advantage is yet to be understood.

Are genre labels the best measure of musical innovation? Probably not. We have used the fine-grained genre labels provided by Every Noise at Once, but an even "finer" approach is possible: using lyrical and acoustic features of songs themselves to measure song novelty, as was done in a number of studies. In this case, we would be able to also address the well-known, but sometimes disputed, claim that the form of novelty with the highest chances to be preferred may be "optimal" novelty: e.g., songs must be different from the already existing songs, but not *too* different; they must hit a sweet spot of new-but-not-too-new (Askin & Mauskapf, 2017; O'Toole & Horvát, 2023); but also see (Berger & Packard, 2018; Yu et al., 2023).

Recently, there has been growing interest in empirically studying the regularities in the success of music and other arts: preferential attachment (Fraiberger et al., 2018), conformity (Youngblood, 2019), content-based selection (Morin & Sobchuk, 2021), intergroup contact (Granito et al., 2019), and others. They allow seeing the history of arts as the history of *mechanisms*: the "gears" driving the evolution of some aspects of culture in repeated, predictable ways. The first-mover advantage, which has not yet been studied in the arts, should be added to this list of mechanisms.

Abbreviations

API: Application Programming Interface.

Authors' contributions: OS and OM designed the study. MY collected the data. OS ran the analysis. OS and MY designed the visuals. OS wrote the manuscript, later reviewed and edited by OM and MY.

Availability of data and materials: Full data, code, and the research diary showing the progress of this study are available at the Open Science Framework's website, where we had also preregistered our analysis. https://osf.io/fqjnz/?view_only=d3792904063f4303beb8fd15b294e794

Competing interests: The authors declare no competing interests.

Funding: This study was funded by the Max Planck Society.

Acknowledgements: [Hidden for peer review.]

References

Anderson, A., Maystre, L., Anderson, I., Mehrotra, R., & Lalmas, M. (2020). Algorithmic Effects on the Diversity of Consumption on Spotify. *Proceedings of The Web Conference 2020*, 2155–2165. https://doi.org/10.1145/3366423.3380281

Askin, N., & Mauskapf, M. (2017). What Makes Popular Culture Popular? Product Features and Optimal Differentiation in Music. *American Sociological Review*, *82*(5), 910–944. https://doi.org/10.1177/0003122417728662

Barabási, A.-L. (2018). *The Formula: The Universal Laws of Success*. Little, Brown.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Berger, J., & Packard, G. (2018). Are Atypical Things More Popular? *Psychological Science*, *29*(7), 1178–1184. https://doi.org/10.1177/0956797618759465

Brinkmann, L., Baumann, F., Bonnefon, J.-F., Derex, M., Müller, T. F., Nussberger, A.-M., Czaplicka, A., Acerbi, A., Griffiths, T. L., Henrich, J., Leibo, J. Z., McElreath, R., Oudeyer, P.-Y., Stray, J., & Rahwan, I. (2023). Machine culture. *Nature Human Behaviour*, *7*(11), Article 11. https://doi.org/10.1038/s41562-023-01742-2

Brouillette, S. (2002). Corporate Publishing and Canonization: "Neuromancer" and Science-Fiction Publishing in the 1970s and Early 1980s. *Book History*, *5*, 187–208.

Burgess, R. J. (2014). *The History of Music Production*. Oxford University Press.

Campbell, M., Greated, C. A., & Myers, A. (2004). *Musical Instruments: History, Technology, and Performance of Instruments of Western Music*. Oxford University Press.

Dodds, P. S., Dewhurst, D. R., Hazlehurst, F. F., Van Oort, C. M., Mitchell, L., Reagan, A. J., Williams, J. R., & Danforth, C. M. (2017). Simon's fundamental rich-get-richer model entails a dominant first-mover advantage. *Physical Review E*, *95*(5), 052301. https://doi.org/10.1103/PhysRevE.95.052301

Dubourg, E., & Baumard, N. (2022). Why imaginary worlds? The psychological foundations and cultural evolution of fictions with imaginary worlds. *Behavioral and Brain Sciences*, *45*, e276. https://doi.org/10.1017/S0140525X21000923

Fatsis, L. (2019). Grime: Criminal subculture or public counterculture? A critical investigation into the criminalization of Black musical subcultures in the UK. *Crime, Media, Culture*, *15*(3), 447–461. https://doi.org/10.1177/1741659018784111

Fraiberger, S. P., Sinatra, R., Resch, M., Riedl, C., & Barabási, A.-L. (2018). Quantifying reputation and success in art. *Science*, *362*(6416), 825–829. https://doi.org/10.1126/science.aau7224

Frank, R. (2016). *Success and Luck: Good Fortune and the Myth of Meritocracy*. Princeton University Press.

Granito, C., Tehrani, J., Kendal, J., & Scott-Phillips, T. (2019). Style of pictorial representation is shaped by intergroup contact. *Evolutionary Human Sciences*, *1*, e8. https://doi.org/10.1017/ehs.2019.8

Hancox, D. (2019). *Inner City Pressure: The Story of Grime* (Illustrated edition). HarperCollins UK.

Heller-Nicholas, A. (2014). *Found Footage Horror Films: Fear and the Appearance of Reality*. McFarland.

Iglewicz, B., & Hoaglin, D. C. (1993). *How to Detect and Handle Outliers*. ASQC Quality Press.

Kerin, R. A., Varadarajan, P. R., & Peterson, R. A. (1992). First-Mover Advantage: A Synthesis, Conceptual Framework, and Research Propositions. *Journal of Marketing*, *56*(4), 33–52. https://doi.org/10.2307/1251985

Knight, S. (2010). *Crime Fiction since 1800: Detection, Death, Diversity* (2nd edition). Red Globe Press.

Liu, L., Wang, Y., Sinatra, R., Giles, C. L., Song, C., & Wang, D. (2018). Hot streaks in artistic, cultural, and scientific careers. *Nature*, *559*(7714), Article 7714. https://doi.org/10.1038/s41586-018-0315-8

MacCallum, R. M., Mauch, M., Burt, A., & Leroi, A. M. (2012). Evolution of music by public choice. *Proceedings of the National Academy of Sciences*, *109*(30), 12081–12086. https://doi.org/10.1073/pnas.1203182109

Morin, O., & Sobchuk, O. (2021). The Shortlist Effect: Nestedness contributions as a tool to explain cultural success. *Evolutionary Human Sciences*, 1–27. https://doi.org/10.1017/ehs.2021.48

Mueller, D. C. (1997). First-mover advantages and path dependence. *International Journal of Industrial Organization*, *15*(6), 827–850. https://doi.org/10.1016/S0167-7187(97)00013-1

Newberry, M. G., & Plotkin, J. B. (2022). Measuring frequency-dependent selection in culture. *Nature Human Behaviour*, *6*(8), 1048–1055. https://doi.org/10.1038/s41562-022-01342-6

Newman, M. E. J. (2009). The first-mover advantage in scientific publication. *EPL (Europhysics Letters)*, *86*(6), 68001. https://doi.org/10.1209/0295-5075/86/68001

O'Toole, K., & Horvát, E.-Á. (2023). Novelty and cultural evolution in modern popular music. *EPJ Data Science*, *12*(1), Article 1. https://doi.org/10.1140/epjds/s13688-023-00377-7

Ravignani, A., Thompson, B., Grossi, T., Delgado, T., & Kirby, S. (2018). Evolving building blocks of rhythm: How human cognition creates music via cultural transmission. *Annals of the New York Academy of Sciences*, *1423*(1), 176–187. https://doi.org/10.1111/nyas.13610

Rossman, G. (2015). *Climbing the Charts: What Radio Airplay Tells Us about the Diffusion of Innovation*. Princeton University Press.

Salganik, M. J., Dodds, P. S., & Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, *311*, 854–856. https://doi.org/10.1126/science.1121066

Sobchuk, O. (2018). *Charting artistic evolution: An essay in theory* [Thesis, University of Tartu]. https://dspace.ut.ee/handle/10062/62406

Spotify. (n.d.). *Spotify Web API (v.1.0.0)*. Retrieved July 12, 2022, from https://developer.spotify.com/documentation/web-api/reference/

triph. (2022). *Sp-playcount-librespot* [Java]. https://github.com/entriphy/sp-playcount-librespot (Original work published 2020)

Varnum, M. E. W., Krems, J. A., Morris, C., Wormley, A., & Grossmann, I. (2021). Why are song lyrics becoming simpler? A time series analysis of lyrical complexity in six decades of American popular music. *PLOS ONE*, *16*(1), e0244576. https://doi.org/10.1371/journal.pone.0244576

Waldfogel, J. (2018). *Digital Renaissance: What Data and Economics Tell Us about the Future of Popular Culture*. Princeton University Press.

Wang, D., & Barabási, A.-L. (2021). *The Science of Science*. Cambridge University Press. https://doi.org/10.1017/9781108610834

Webster, J. G. (2014). *The Marketplace of Attention: How Audiences Take Shape in a Digital Age*. The MIT Press. https://doi.org/10.7551/mitpress/9892.001.0001

Youngblood, M. (2019). Conformity bias in the cultural transmission of music sampling traditions. *Royal Society Open Science*, *6*(9), 191149. https://doi.org/10.1098/rsos.191149

Yu, Y., Cheung, P. Y., Ahn, Y.-Y., & Dhillon, P. S. (2023). Unique in What Sense? Heterogeneous Relationships between Multiple Types of Uniqueness and Popularity in Music. *Proceedings of the International AAAI Conference on Web and Social Media*, *17*, 914–925. https://doi.org/10.1609/icwsm.v17i1.22199