Where'd You Get that Idea? Determinants of Creativity and Impact in Popular Music

Bernardo Mueller
Department of Economics
Universidade de Brasília

September 03, 2018

Abstract
What are the determinants of creativity, innovation and impact? In this paper I explore this question through an analysis of data from the Song Explorer podcast, where composers describe how they created a specific song. I mine their accounts to classify their processes into seven different, but not mutually exclusive, theories of the creative process. The result of this exercise suggests that the recombination of existing songs is a major process for the creation of new successful songs. The second step considers what kind of recombinations are associated with high impact. For each song in the sample I have one or more other songs which were explicitly indicated as an influence or inspiration. I use the music genre classification system *Every Noise at Once,* that provides a map of over 1,800 genres and millions of songs to create a set of descriptive statistics of the similarity of each song to their inspiration-songs. These statistics are then used as explanatory variables in a regression that seeks to explain impact (YouTube views per day since the songs` video release), while controlling for other determinants of song impact, such as the artists' established level of popularity. The results confirm the optimal differentiation hypothesis that the simultaneous presence of conventionality together with novelty, and not just one or the other, is a major determinant of creativity and impact.

Keywords: popular music, creativity, impact, innovation.
JEL codes: O3, Z1, D16, D7

ORCID ID: 0000-0002-0268-8377

Address:
Department of Economics
Universidade de Brasília
Brasilia, DF 71665-185
Email: bmueller@unb.br
Home-page: http://bpmmueller.wixsite.com/bernardo-mueller

Acknowledgements: For comments and suggestions to earlier versions of this paper I recognize and thank Felipe Carneiro, Noah Askin, and Leonardo Monastério. Grants from CNPq (302757/2015-5) and CAPES (BEX 6028/14-4) are recognized.

1. Introduction

Where do good ideas come from? One hypothesis was memorialized in the film *Amadeus*, by Milos Forman (1984), in a scene where a scheming and jealous Salieri offers to annotate the music as a bed-ridden Mozart composes *Confutatis Maledictis* of the Requiem. Salieri wants to understand how this garish and vulgar young man can create such sublime and touching music that so elude his own efforts. As Mozart dictates the notes, Salieri starts to feel that he can see where the piece is going. But all the sudden Mozart takes an unexpected turn leading the composition in a wholly unanticipated direction. Exasperated, Salieri protests that this cannot be done, that it won't work. But as Mozart carries on and reveals additional chords Salieri has an epiphany at the unprecedented beauty that was just magically materialized. Now that it was fully there before his eyes, so precise and perfect, the piece seemed so obvious and unavoidable that he wondered how he himself had not seen it before.[1]

A different hypothesis is shown in another memorable scene of a different movie. In *24 Hour Party People* (2002), a semi-fictionalized account of the rise of the Manchester music scene in the late 1970s. The scene is set in a concert hall where the Sex Pistols, a punk band from London, are about to go on stage. It is early days in the punk movement, which had not yet reached Manchester. The venue is desolate with just 42 people dispersed among rows of empty seats. As the band unleashes its wall of sound, the camera swivels from the stage to the audience and we see the horror and disbelief in their faces. They had never seen or heard anything like this before. The narrator goes on to describe how, unwittingly inspired by that moment, each one of them would go on to form their own bands or become game-changing producers or label owners (Buzzcocks, Joy Division, New Order, Simply Red, Martin Hannett, Tony Wilson), turning Manchester into a worldwide hub of the new punk and subsequent movements.

In the first hypothesis, creativity and innovation is something performed by genius and talent, pulling new ideas out of thin air. In the second inspiration comes from exposure and recombination of already existing ideas. This paper explores these and other hypotheses described below. I test for creativity and impact in popular music using data from a podcast (*Song Exploder*) where artists deconstruct a single specific song in detail, describing how it was conceived and composed.[2] Like Salieri drafting for Mozart, these accounts allow a window into

[1] For information on the film Amadeus see https://www.imdb.com/title/tt0086879/?ref_=fn_al_tt_1.
[2] http://songexploder.net/.

the creative process of a sample of successful artists. In the next section I single out seven, not necessarily mutually exclusive, theories of creativity that might explain where good songs come from. The main result from this exercise is a measure of the extent and the pattern through which successful songs recombine existing ideas to produce beauty and impact.

The pattern I uncover confirms the findings from several different papers in different areas that have also investigated where novelty and good ideas come from; Askin and Mauskapf (2017) for popular musci; Uzzi, Mukherjee, Stringer, and Jones (2013) and Wang, Veugelers, and Stephan (2017) for scientific publications; Youn, Strumsky, Bettencourt, and Lobo (2015) for patents and inventions; and Barron, Huang, Spang, and DeDeo (2018) for innovation in speech patterns (in the French Revolution). All these papers found a similar signature to successful creative processes, where the existing ideas that were used to produce new combinations were neither too typical or conventional, nor too extreme or radical. Successful ideas in all these areas seem to have some elements of the recognized and established, that give the recipient a foothold and familiarity, as well as simultaneously, novelty and surprise, making it stand out of the mass competing for attention. It balances a trade-off between exploration and exploitation that is common in evolving complex systems (Holland, 1992; March, 1991)

The sample of songs here is smaller than the massive data sets used in these other studies, as it is limited by the number of *Song Exploder* podcasts episodes (115 at the time of the writing of this paper). It does, however, provide something unique: the explicit declaration by the composer of which other songs and artists inspired that particular song in my sample. It is this that enables the specific exercise performed below.[3] As with all the studies cited above, I need a 'map' of the entire set of possible combinations so as to compare those that I actually observe. The map I use is a resource created by Echo Nest that provides an online 'music intelligence platform' that uses Spotify data from more than 30 million songs and more than 3 million artists to classify artists into genres (i.e. rock, funk, indie, etc.) and relate genres among themselves in a way that provides a measure of artists similarity.[4] There are more than 1,800 genres of music,

3 Beside the smaller sample size, there are other limitations to this study. The first is that there are no guarantees that the narrative given by the creator is faithful to how the song was actually created, or whether they are ex-post rationalizations, wishful thinking or some other form of cognitive dissonance. The second limitation lies in the fact that the narratives are assessed and coded by the investigator, so that even using objective criteria the final classification of the evidence extracted from each narrative can also have some subjectivity. This is a common limitation of many measures of creativity and can be somewhat mitigated by being explicit on the protocols used when coding the data.

4 Available at http://everynoise.com/.

from a capella to zydeco, passing through brutal death metal and capoeira as well as crunk, fado and gnawa, among many others. Because each of the songs in my sample has from 1 to 10 genres, this allows me to create a set of four descriptive statistics for each, providing a description of the nature of the recombination employed. These four statistics are (i) the mean of the genre pairs between a song and its inspiration-song; (ii) the dispersion of the genre pairs; (iii) the average of the 10% most dissimilar pairs; and (iv) a bimodality coefficient. I then test whether these statistics (that is the recombination strategy) are able to explain the songs' impact, as measured by the average daily views of the songs' main YouTube video, while controlling for other determinants of success, such as artists' previous achievements in the Billboard 200 charts, their years on the road, whether they are signed to a major label, and a measure of popularity based on the artist's total number of monthly Spotify listens. The results indicate that impact is associated with a recombination strategy based on conventionality together with novelty, that is, inspiration simultaneously close and far from the artist's own genres.

2. Seven theories of where good songs come from

In this section I describe seven theories that seek to explain the determinants of musical creativity. These theories are a compilation and classification of separate explanations encountered in different sources using my own judgement, as there seems to be a lack of consensus in the broad field of Creativity regarding definitions, measurement and determinants.[5] The name given to each theory is therefore simply used as a moniker to represent a set of common ideas, and not an already-established term in the literature. Each theory will be briefly described leaving details to the citations therein. My classification of the Song Exploder podcasts put a check next to each of the theories for that song if the narrative given by the composer suggested elements associated with that theory. A same song can be classified in more than one theory.

It is important to remember throughout this exercise that the songs being used are not randomly selected, but rather a sample of songs that was chosen precisely because they were successful. That is why they were invited to participate in the Song Exploder podcast. The exercise is therefore one of establishing the determinants of relative success and impact of songs, conditional on the songs already being successful in terms of sales, airplay or notoriety. In fact,

[5] For good reviews see Kozbelt, Beghetto, and Runco (2010), Runco and Jaeger (2012), Simonton (2012), and Batey (2012).

as the podcast became more popular, it clearly managed to attract bigger and more famous artists.

Theory 1 and 2 – Conceptual innovator and Experimental Innovator

These two theories are presented together as they are opposite poles of a classification of creative approaches suggested by Galeson (2006, 2009). A conceptual innovator is one that rationally plans and executes an idea well aware and in control of the process. An experimental innovator follows a more open-ended process, without much planning or foresight, allowing the creative process to flow with many detours and repetition. It is often the case that conceptual innovators flourish early in life while experimental innovators do their best work as late bloomers (Galeson, 2006; Ginsburgh & Weyers, 2006). In painting, a canonical example of each type are Picasso, who by the age of 25 had done many of his greatest works, and Cezanne who did his best work as an old master.

These notions often refer to the entire career of the creator, whereas my data refers to a specific song. The theories were marked if the composer`s narrative suggested a creative process that was fast, rational and planned (conceptual) or open-ended, incremental and tentative.

Theory 3 – Diversity/Team

The creative process can be enhanced when it is performed by groups instead of individuals. There is much research showing the impact of diversity and team effort on a wide variety of creative processes (Hong & Page, 2004; Scott E. Page, 2008; Scott E. Page, 2011; Uzzi et al., 2013). Different participants bring different perspectives and different capabilities so that teams often have a performance that is greater than the sum of the parts.

This theory is marked whenever the composer's narrative explicitly cites the participation of someone outside the artist's usual collaborators (band members, old co-authors, usual producers).

Theory 4 – Recombination

Novelty, creativity and impact often do not come from scratch, but are the result of the recombination of known and tested elements, be they songs (Askin & Mauskapf, 2017), scientific papers (Uzzi et al., 2013; Wang et al., 2017), technologies (Arthur, 2009; Mokyr, 1990), phenotypes (Darwin, 1859), economic capabilities (Hidalgo & Hausmann, 2009), ideas (Weitzman, 1998), beliefs (Mueller, 2016); or property rights (Alston & Mueller, 2015). This is the notion that novelty is not pulled out of thin air, but from an adjacent possible, which "is a kind of shadow future, hovering on the edges of the present state of things, a map of all the ways in which the present can reinvent itself" (Johnson, 2010a). The notion of an adjacent possible captures both the idea that there are immense opportunities for creativity through recombination of what is already known, but also the notion that at any given time what can be done is limited (Johnson, 2010b; Kauffman, 1995).[6]

The recombination theory is marked when the composer's narrative makes explicit reference to another artist or song stating that this was a direct influence for the composition of that song.

Theory 5 – The medium is the message

McLuhan (1965) famously argued that 'the medium is the message', that is that the medium through which content is delivered is often more important and impactful that the content itself. In popular music this is important as content delivery has changed from live performance only, to radio, television, vinyl, cassettes, MP3, YouTube, iTunes, Spotify and others.

This theory is marked when the composer explicitly states that the creative process was meaningfully affected by something related to an instrument, studio, recording equipment, or some related situation.

Theory 6 – Serendipity

In some cases, an artist or problem-solver is purposefully looking for something or trying to achieve a given result, but by accident hits upon something else unexpected, that nevertheless turns out to be novel and creative. Drug research by pharmaceutical companies, for example, relies to a great extent on such serendipity (Taleb, 2007). Mokyr (1990, p. i) stresses the role throughout the history of technological change of "luck, serendipity, genius, and the unexplained drive of people to go somewhere where none has gone before."

This theory was marked when the narrative explicitly referred to an accidental or serendipitous role in the composition of the song.

Theory 7 - Adversity

[6] It is important to differentiate naïve recombination theories that see novelty arising from simple additive aggregation of existing elements from more elaborate theories that allow for complex interactions and, recognizing the exceptionality of fruitful combinations, try to find patterns that explain success. See DeDeo (2018) for a critique of recombination theories of creativity.

Inspiration and creativity are often associated with adversity and negative experiences. Thomson and Jaque (2018), for example, found that in a sample of 234 musically related performers, those who reported more adverse childhood experiences exhibited significantly stronger creative experiences. Songs are often about heartbreak, loneliness and despair.

This theory was marked when the composer's narrative made an explicit link of the song to some adverse or negative experience.

Results

The results of the exercise are shown in Table 1. Each of the theories have some support from the composers' narratives, but by far recombination was the most prevalent with 71% of the songs cited as being inspired by one or more specific songs. This is somewhat surprising. In science and academic work, it is standard practice and even a requirement to make explicit citations of previous work. In music, however, there is no such convention. Quite to the contrary, preoccupation with accusations of plagiarism often make authors reluctant to admit being influenced by existing songs. Presumably, in the context of the Song Exploder podcast composers felt at ease to discuss their creative process. In one podcast, for example, the interviewee stated half-jokingly: "That immediately reminded me of a Johnny Marr type of chord progression. That longing, that ache, that a lot of the Smiths records have. So, I was really just trying to rip him off."[7] Several others also cited existing songs and explicitly stated that they wanted to make something similar.

[Table 1 here]

This result suggests that the creativity, at least in music, is often not a process of pulling ideas out of thin air or from a genius's brain or soul, but rather a process of recombination of existing material. This does not denigrate the process, for as we shall see in the next section, it is not any recombination that will do, rather recombination is itself an art. Another interviewee expressed this notion as follows:

> We were messing around with a sample from a gospel record ... The darkness and the blues, it had the thing that I wanted for this song. … I wrote that part, but I don't take credit because again it's a very traditional kind of thing. You will hear something like that in hundreds of quartet gospel records. We wanted the textures to come from different atmospheres, which is a very Hip-Hop kind of thing. It`s sample based traditionally, and every sample is taken from a different record, different time, different genre, and that's part of what makes the soup and the gumbo so beautiful, making a collage out of different

[7] Song Exploder podcast # 99 with Sleigh Bells, http://songexploder.net/sleigh-bells .

> sounds and bringing it all together, recycling it in a kind of way that creates something new even though you are taking it from things that are preexisting. (Ghostface Killah[8])

Second to recombination, the theory which receives the most support is the team theory. Half of the songs in the sample mentioned a crucial role of outside collaborators, producers or other outsiders. In a sense, these collaborations can be thought of as a type of recombination, where the different ideas being combined come from different brains.

The other 5 theories receive some support, especially that which associates creativity with adversity, but much less than recombination or teams.

3 – Testing the determinants of impact

If the recombination of existing ideas in the adjacent possible is one of the preponderant means through which impact is created in music, then what kind of recombination is most effective? Certainly, it is not any mix that yields something novel that has value. The sample used in this paper consists only of songs that were successful in some way or another, which is why they were chosen for the *Song Exploder* podcast. Nevertheless, even in this biased sample, some songs turned out to be more impactful and stimulating than others. In this section I estimate the determinants of success and impact, conditional on the songs already being successful in the first place. The focus of the estimation is to determine which strategy for recombination is associated with most impact, as measured by average daily YouTube views of the songs' videos, while controlling for any inertial success that the artist may carry with them and which would yield views even without quality or value in the specific song being analyzed.

A recombination strategy refers to the different recipes for mixing and matching existing ideas into new ideas. Songs have many dimensions which can be sampled and drawn upon; melody, rhythm, harmony, beat, tone, texture, form, tempo, riffs, rolls, lyrics, vocalizations, etc. Genres can be thought of as subsets of songs that have some fixed constraints along some of these dimensions, allowing for variability only along other dimensions. But even with such constraints, with so many elements to pick from, the combinations are almost infinite. Yet the great majority of these combinations will be either gibberish or will turn out to be songs of little value or beauty. The problem is how to find those few cases where the recombined elements come together symbiotically creating something greater than their sum. The ambition of the exercise in this section is not to uncover the formula for the perfect pop song, but rather to

[8] Song Exploder podcast # 26 with Ghostface Killah http://songexploder.net/ghostface-killah .

determine if there is any pattern in successful sampling. Is it, for example, better to sample from songs that are near or far from what one normally does, that is, one's own genre?

To do this, I need a map that allows me to establish the distance, musically, between the song I am analyzing and the songs that were cited as inspiration. In the context of scientific papers Uzzi et al. (2013) created such a map by generating a randomized network of all the possible pairs of citing and cited papers, covering 15,613 journals and 122 million pairs. This was then used as a benchmark to ascertain the typicality of the observed citation pairs. This allowed them to conclude that "papers that combine high median conventionality and high tail novelty are hits in 9.11 out of 100 papers, a rate nearly double the background rate of 5%" (Uzzi et al., 2013, p. 470).

Barron et al. (2018) test for novelty in the transcripts of the French Revolution's first parliament covering over 40,000 speeches and a thousand speakers over a two-year period. They use an information theory-based method to measure novelty as surprise in new speech or text patterns, given existing patterns (measured as Kullback-Leibler divergence). By measuring the extent to which new patterns persist into the future they also have a measure of transience. Their results show that novelty and transience are highly positively correlated, so that most new ideas or patterns are quickly forgotten. However, they also show that those ideas which do persist and make a mark, tend to have high novelty. Although these results are not exactly based on the notion of recombination of existing ideas, they reach a similar conclusion that novelty is necessary but not sufficient for impact. Creativity requires novelty but not just any kind of novelty will do.

In the context of popular music, Askin and Mauskapf (2017) use the same Echo Nest data and analysis that is behind the *Every Noise at Once* platform that I use, to create a typicality index based on 10 song features (acousticness, danceability, energy, instrumentalness, key, liveliness, mode, speech, tempo and time signature). This index is then used as an explanatory variable in a regression to explain success as measured by peak position in the Billboard 100 chart (as well as weeks on the chart). They find that:

> … songs must strike a balance between being recognizable and being different. Those that best manage this similarity–differentiation tradeoff will attract more audience attention and experience more success. Stated more formally, we predict an inverted U-shaped relationship between a song's relative typicality and its performance on the *Billboard* Hot 100 charts. Our analysis highlights the opposing pressures of crowding and differentiation by constructing a summary measure of song typicality. (Askin & Mauskapf, 2017, p. 6)

In this paper I also test for what Uzzi et al. (2013) call 'atypical combinations' and Askin and Mauskapf (2017) call 'optimal differentiation'. I have a sample of successful songs for which I have from one to three other songs that were used as direct and explicit inspiration. The *Every Noise at Once* platform gives a classification of the genre of each of the artists. Each artist is classified in up to 10 genres, with those genres at the top of the list the more dominant genres for that artist. Figure 1 provides a snapshot of part of the musical-genre map, which contains more than 1,800 different genres spatially placed according to their similarity based on algorithmically processed data from Spotify listeners' habits.[9] Figure 2 shows the first 29 positions of the list that ranks each genre according to its similarity to 'pop' according to the platform's algorithm. Clicking on any other genre re-orders the list according to similarity to that genre. This allows me to calculate a measure of similarity as illustrated in Table 2 for the case of the song *Andromeda* by the band *Gorillaz*. In the *Song Exploder* podcast, Damon Albarn, who composed the song stated:

> This originated from a conversation between myself and Twilight Zone, the guy who co-produced the record. We were talking about two of the greatest 80s pop songs and we decided that Billy Jean by Michael Jackson and I Can't Go for That by Hall and Oates were two of our favorite tunes, in their tempo and their pop sensibilities, and how could we somehow chemically channel the greatness of those into our own music. " http://songexploder.net/gorillaz minute 2:10.

Table 2 shows that *Gorillaz* are accorded two genres, Michael Jackson two other genres, and Hall and Oates five genres, thus creating 14 genre pairs between the first and the other two songs. The distance between the genres is measured through the rank of each of the inspiration songs' genres in the similarity list for 'alternative hip hop' and then 'art pop'. By using multiple genre-pairs instead of only the main pair I allow for the complexity and multidimensionality of music, artists and genres.[10]

[Figure 1 here]

[Figure 2 here]

[Table 2 here]

The various genre-pairs are then used to produce four different statistics that describe the genre-pair distribution to be used to measure impact. The first two statistics are the mean and the

[9] View the entire map and see more information on how the map is built in http://everynoise.com/ and in its creator's (Glenn MDonald) blog http://www.furia.com/ .

[10] The use of artist-level genre to compare songs, rather than song-level genres, is similar to how Uzzi et al. (2013) and Wang et al. (2017) measured combinations in scientific papers. Their unit of measurement was at the level of journal-pairs and not paper-pairs. In a sense, a journal tells us what is the 'genre' of an author. Like musicians, authors can have more than one genre by publishing in different fields or sub-fields.

median, which capture how far on average the song reached out for inspiration. To assess the magnitudes, consider that the ranking runs from 1 to 1,873. The third statistic is a measure of the most extreme genre-pairs in the distribution. It is the average of the 10% highest genre-pairs. In the example in Table 2 there are 14 pairs, so that the most extreme 10% (rounding down from 1.4 to 1) is simply the furthest pair equal to 932. The fourth measure is the standard deviation of the distribution of genre-pairs, to capture the concentration vs dispersion of the influencing songs' genres. Finally, the fifth statistic is a bimodality coefficient which is used to determine if the inspiration for the song came from mostly a same region in the genre-space, or whether it was taken from more than one place.[11] As a rule of thumb, a uniform distribution has a bimodality coefficient of 0.55, with bimodality for higher coefficients and unimodality for lower. In the example, the 0.522 coefficient indicates that this song sampled narrowly.

In Figure 3 I illustrate the variability of different sets of statistics, that is, different recombination strategies by showing the distribution of genre-pairs for three different songs. The first song sampled from close to its own genres. The mean is low, the extreme values are low, the dispersion is narrow and the distribution is unimodal. The second song has a similar dispersion and bimodality coefficient, but it sampled much further afield. The third song clearly adopted a bimodal recombination strategy, sampling both near and far.

With these statistics we can proceed to estimate the determinants of creativity and impact. The dependent variable is the average number of daily views on YouTube of the song's official (or most viewed) video since the launch of the video.[12] This is intended as a measure of the success and impact of the specific song and not of the artist. I did not use peak position or weeks in the Billboard 100 chart because many of the songs in the sample did not make it to that selective chart.

The explanatory variables of interest are the genre distribution statistics. But it is necessary, in addition, to control for artists' characteristics that may lead to YouTube video views independent of the song's intrinsic quality. If a highly successful band, such as Metallica,

---

[11] The bimodality coefficient is calculated through the following formula: $BC = \frac{m_3^2+1}{m_4+3\left(\frac{(n-1)^2}{(n-2)(n-3)}\right)}$, where m_3 is the skewness of the distribution and m_4 its excess kurtosis. "The BC of a given empirical distribution is then compared to a benchmark value of $BC_{crit} = 5/9 \approx 0.555$ that would be expected for a uniform distribution; higher numbers point toward bimodality whereas lower numbers point toward unimodality." (Pfister, Schwarz, Janczyk, Dale, & Freeman, 2013).

[12] I take the log of the average daily YouTube views to deal with the skewness that arises from having some very big hits in the sample. Results are similar without this transformation.

REM, or U2, in my sample, release a new song video, it will naturally attract a large number of views even if the song and video are mediocre. But this inertial success can only carry a song so far. For real impact, the song must have something special that will take it further than the artist's past achievements. Therefore, to explain a specific song's success, it is necessary to control for name recognition and inertial success. I use four different variables for this purpose. The first is a variable that captures the artist's best placement in the Billboard 200 album charts in its career prior to the song's release. If the artist never had a Billboard 200 placement, the value of 250 is attributed. If the artist never had a Billboard 200 album but had some other minor Billboard chart placement (for example, Hot 100, or Hot Latin Song) I add the peak position in that chart to 200. An Oscar nomination, Grammy or Emmy gives the artist a value of 200 for this variable. The idea of this variable is to capture an artist's received popularity at the time it released the song in question.

The second variable to control for artist stature is a dummy equal to 1 if the song was released by a major label or an independent label. Major labels have more financial resources, personnel, contacts and leverage to promote their artists than independent labels, which should translate to more views for any given level of song quality. The third variable is the number of years the artist has been active, which may capture experience, learning and recognition, but can also be a sign of tedium and sameness. The fourth variable is the artist's number of monthly views on Spotify. This variable is an artist-level measure as opposed to YouTube views, which is a song-based measure.[13]

Another set of controls are dummies for some of the major genres: Pop, Rap, Rock, Folk, Indie, Soundtrack, Electronic, Metal, and Soul. Assigning the artists to a major genre is somewhat arbitrary as the system usually classifies them under several sub-genres such as indie-folk, downtempo, art pop, and chillwave. But the idea of the dummies is to control for the fact that certain types of music might have fans that are more adept and used to consuming music through YouTube. Descriptive statistics for all variables are shown in Table 3. There are 104 observations, as 10 of the cited influences were not actual artists, but a generic beat, a cartoon, a

[13] If an artist is famous primarily due to the song I am analyzing, then this variable would be endogenous, that is, the Spotify-listens variables would be determined by the YouTube-views variable. But in most cases, the Spotify-listens variable is composed of views of a large number of songs, so that the specific song in question has very little individual impact. For example, U2 had already released 12 albums with more than 100 songs by the time it released Cedarwood Road, which is in my sample, and which was not one of its major hits. Therefore, it is safe to assume that this song makes up only a small part of the U2 Spotify-listens variable.

Christmas song, or other such inspiration that could not be found in the *Every Noise at Once* database. Of the 104 observations, 71 recombined existing songs as part of their creative process and the remaining 33 did not. I run the regressions with all 104 observations setting the mean, 10% tail, standard deviation and bimodal variables equal to zero for those that do not cite having received inspiration from other songs. I also run the model with only the 71 observations that used a recombination strategy to ascertain that the sample choice does not determine the results.

Table 4 presents the ordinary least square regressions where the dependent variable is the natural log of the number of average daily YouTube videos for each song. In column 1 the explanatory variables are those that measure the artists' popularity and stature at the time of the release of the song in question. The variable that measures the artists' previous Billboard success is not significant, but has the expected negative sign, indicating that artists that climbed higher (have lower values, closer to #1) in the album charts have more views. The major-label dummy is also not statistically significant, perhaps because this is a sample of only successful songs. It might also be an indication that in the current disrupted music market major labels are not as powerful as they once were. The variable that measures how long the artists have been active is statistically significant and positive, which means that more recent artists have more views than those that have been around for longer, *ceteris paribus*. This result indicates that for this sample of successful artists, novelty trumps experience and familiarity.[14] Keeping all other explanatory variables fixed, a one year more recent release increases the number of daily views by 8.6%.[15] Also highly significant is the artists' monthly Spotify-listens, a measure of artist popularity. The variable is in units of one-million listens, so the estimated coefficient of 0.51 indicates that an additional 1 million listens to the artist on Spotify is associated with 66.5% additional YouTube daily views of the song in question. To interpret these impacts in terms of number of views, consider that if all variables are set at their mean values (major label dummy set equal to 1), the average number of YouTube views per day would be 620. If the same artist had an additional 1 million Spotify listens, then the number of view would increase to 1,033. Similarly, increasing the year since first active from 2000 to 2010 would yield an increase in views to 1,419. Naturally, this is a statistical manipulation with 'average' artist characteristics. The regression's

[14] The average starting year in the sample is 2000, the minimum is1960, and the maximum 2015.

[15] Because the estimation is done in log-linear form the interpretation of the coefficients requires that they be transformed by the formula $\%\ \Delta dep.variable = \left(e^{\beta} - 1\right) * 100$ and interpreted as percent change in the dependent variable (where β is the estimated coefficient shown in table 4).

adjusted R-squared of 0.26 indicates that around three-quarters of the variation in views is not explained by this specification, so for any specific artist there are many other factors influencing success and impact.

In column 2 I use only the genre-pairs distribution statistics, which measure the recombination strategy used by the songs' composers. All the variables are statistically significant except for the standard deviation of the distribution. The higher the mean of the genre-pair distribution, that is, the further away from the artist's own genres this song's inspiration was drawn from, the greater its impact, *ceteris paribus*. However, the more extreme the 10% furthest genre-pairs, the lower will be that impact. These are opposing forces, so that on average it pays to sample distantly, but not to go too far afield. This result is similar to the optimal differentiation found by Askin and Mauskapf (2017) for popular music, and by Uzzi et al. (2013) for academic publications. In addition, a bimodal recombination strategy is associated with more views than one which is centered around a same portion of the genre-space. Remember that as the bimodality coefficient increases the distribution of genre-pairs goes from more unimodal at low values, to normally distributed at around 0.55, to bimodal at higher levels.

The impacts of these variables are not only statistically significant, but they also have large effects on the number of daily YouTube views. With all variables set at their mean value, if the mean of the genre-pairs moves 100 rank points further down the list of 1,875 genres, the number of daily YouTube-views increases from 588 to 876 views. But an increase in the mean would probably also increase the 10% most extreme genre-pair values, which would, all other variables fixed, yield 432 YouTube views. So, the final effect of sampling further afield depends on the relative sizes of the changes in the mean and the average 10% tail. Increasing both the mean and the 10% tail average by 100 positions would increase the number of daily YouTube views from 588 to 643. In addition, any change in these two parameters of the distribution would probably also affect its bimodality coefficient. The estimation shows that an increase in the bimodality coefficient from a uniform distribution, 0.55, to a somewhat more bimodal distribution of genre pairs of, say, 0.65, would be associated with a shift from 954 to 1292 daily YouTube views. This bimodality effect is, once again, in the same spirit of the optimal differentiation and the atypical combinations hypotheses of creativity and impact.

In column 3 both sets of variables are included simultaneously. The results from columns 1 and 2 remain very much the same in column 3, indicating that the two sets of variables are highly orthogonal to each other. This is confirmed by the fact that the adjusted R-squared increases to 0.35, which is close to the sum of the adjusted R-squared of the separate regressions. Together, these variables explain more than one third of the variability of the songs' average daily YouTube views.

Perhaps the best way to understand these results is in terms of the genre-pair histograms presented in Figure 3. The first histogram in the figure (St. Vincent's song New York) has the same standard deviation and a similar bimodality coefficient to the second histogram (Phoenix's song Ti Amo). It has, however, a much lower mean (304 vs. 989) which, according to the estimated coefficients, implies a lower level of YouTube daily views. However, it has a lower 10% tail (741 vs. 1,356), which increases the number of views. So, the final effect of the recombination strategy depends on which effect, mean or 10% tail, is stronger. Similarly, the third histogram (Dirty Projector's song Up in Hudson) has a high mean of 920, but also a high 10% tail of 1,649, which partially outweigh each other. But it also has a high bimodality coefficient of 0.68, which increases the number of YouTube-views.

The coefficients for the four recombination strategy variables should be interpreted with care. By definition, they show the percent change in the dependent variable due to a unit increase in a given explanatory variable keeping all other variables fixed. But the four variables which describe the genre-pair distributions are necessarily linked. If you increase the mean, then the upper tail, the standard deviation and the bimodality coefficient are likely to also change. Thus, the results are best understood by considering how changes in an entire distribution, composed of the four statistics simultaneously, affect the predicted number of YouTube views.

With all variables set at the mean value of the sample (and the Major Label dummy set at 1), the predicted number of daily YouTube views using the results in column 3 would be 613. Keeping the four artist level variables at their means and putting in the recombination strategy for each of the three artists above, the predicted number of views would be 553 for St. Vincent's New York, 1,301 for Phoenix's Ti Amo, and 355 for Dirty Projector's Up in Hudson. Whereas the first sampled close to home the second did so further afield but still in a concentrated fashion. The third, however, reached far out and sampled in a bimodal fashion from separate areas of the genre space. Although the high bimodality favored more views, the high tail average and large standard deviation reduced the number of predicted views.

These results represent an average for the entire sample. For any given song, however, there may be several other idiosyncratic elements that affect its impact that are captured only in the regression`s error term. The adjusted R-squared indicates that about two thirds of the variability in daily YouTube views remains unexplained by the regression. This could be due to some characteristic specific to each song, to the artist at that point in time, to the *gestalt* of the times, or even to luck, that just somehow made the song special. This is perhaps just as important a result as the pattern that has been uncovered. Although there does seem to be some method to creativity and novelty, a large part just cannot be explained.

Column 5 adds the main genre dummies to the previous specification, but these do not add any explanatory power.[16] Finally, column 6 repeats column 4 but uses only the 71 songs that stated their inspirations and for which there is a genre-pair distribution. The results remain basically the same in terms of statistical significance and estimated coefficient magnitude.

4 - Conclusion

Salganik, Dodds, and Watts (2006) devised an ingenious test of the determinants of the success of popular music by creating an artificial music market in which participants could download previously unknown songs for free. By manipulating the information about how many times each song had been downloaded by other participants, they showed that success is to a large extent determined by peoples' perceptions of what others like, leading to a few and unpredictable songs capturing disproportionate amounts of attention. Importantly, the test also showed that quality, though not the determining factor, also mattered, as good songs rarely did poorly and bad songs rarely did well.

The sample of songs used in this paper had already passed the popularity test, as they were chosen for the *Song Exploder* podcast precisely because they were already successful by some measure or another. The test in this paper sought to explain relative impact among a sample of already successful songs. The first result was that a large majority of the songs were composed with explicit inspiration from already existing songs. The common view of talent or genius pulling new ideas out of thin air seems to be much less prevalent than recombination of existing material as a strategy for creativity. This does not mean, however, that talent and genius do not matter. The second result was that not any recombination will do, but rather some

[16] The excluded category is *Pop*.

strategies seem to work better than others, so that talent, genius or luck may be needed to identify the best ways of mixing and sampling. I have shown that the basic signature for how to recombine from the vast space of existing songs and genres involves reaching far from your own genre and style, but not too far. Sampling from different areas of the genre-space simultaneously was also associated with higher impact. These results confirmed similar conclusions regarding the effect on impact of optimal differentiation and atypical combinations (Askin & Mauskapf, 2017; Barron et al., 2018; Uzzi et al., 2013).

References

Alston, L., & Mueller, B. (2015). Towards a More Evolutionary Theory of Property Rights. *Iowa Law Review, 100*, 2255-2274.

Arthur, B. (2009). *The Nature of Technology: What It Is and How It Evolves*: Simon and Schuster.

Askin, N., & Mauskapf, M. (2017). What Makes Popular Culture Popular? Product Features and Optimal Differentiation in Music. *American Sociological Review, 82*(5), 910-944.

Barron, A. T. J., Huang, J., Spang, R. L., & DeDeo, S. (2018). Individuals, institutions, and innovation in the debates of the French Revolution. *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.1717729115

Batey, M. (2012). The Measurement of Creativity: From Definitional Consensus to the Introduction of a New Heuristic Framework. *Creativity Research Journal, 24*(1), 55-65.

Darwin, C. (1859). *On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life*: London : John Murray, 1859.

DeDeo, S. (2018). The Data Science of Culture: from the French Revolution to String Theory. Presentation at the University of California Merced, April 29, 2018. YouTube video: https://www.youtube.com/watch?v=S7MYdqzOcy4.

Galeson, D. W. (2006). *Old Masters and Young Geniuses: Two Life Cycles of Artistic Creativity*. Princeton: Princeton University Press.

Galeson, D. W. (2009). *Conceptual Revolutions in Teentieth-Century Art*. Cambridge: Cambridge University Press.

Ginsburgh, V., & Weyers, S. (2006). Creativity and life cycles of artists. *Journal of Cultural Economics, 30*(2), 91-107.

Hidalgo, C., & Hausmann, R. (2009). The Building Blocks of Economic Complexity. *Proceedings of the National Academy of Sciences of the United States, 106*(26), 10570-10575.

Hong, L., & Page, S. E. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences of the United States of America, 101*(46), 16385-16389.

Holland, J. (1992). *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence*. Cambridge: MIT Press.

Johnson, S. (2010a, July 16, 2010). The Genius of the Tinkerer. *The Wall Street Journal*.

Johnson, S. (2010b). *Where Good Ideas Come From: The Natural History of Innovation*. New York: Penguin Group.

Kauffman, S. A. (1995). *At home in the universe : the search for laws of self-organization and complexity*. New York: Oxford University Press.

Kozbelt, A., Beghetto, R. A., & Runco, M. A. (2010). Theories of Creativity. In J. C. Kaufman & R. J. Sternberg (Eds.), *Cambridge Handbook of Creativity* (pp. 20-47). New York: Cambridge University Press.

MacLuhan, M. (1964). *Understanding media: The Extentions of Man*: MIT Press.

March, J. G. (1991). Exploration and Exploitation in Organizational Learning. *Organization Science, 2*(1), 71-87.

Mokyr, J. (1990). *Lever of the Riches: Technological Creativity and Economic Progress*. New York: Oxford University Press.

Mueller, B. (2016). Beliefs, Institutions and Development on Complex Landscapes. *Economic Analysis of Law Review, 7*(2), 474-495.

Page, S. E. (2008). *The Difference How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*. Princeton: Princeton University Press.

Page, S. E. (2011). *Diversity and Complexity*: Princeton University Press.

Pfister, R., Schwarz, K. A., Janczyk, M., Dale, R., & Freeman, J. B. (2013). Good things peak in pairs: a note on the bimodality coefficient. *Frontiers in Psychology, 4*, 700.

Runco, M. A., & Jaeger, G. J. (2012). The Standard Definition of Creativity. *Creativity Research Journal, 24*(1), 92-96.

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.

Simonton, D. K. (2012). Quantifying creativity: can measures span the spectrum? *Dialogues in Clinical Neuroscience, 14*(1), 100-104.

Taleb, N. N. (2007). *The black swan : the impact of the highly improbable*: First edition. New York : Random House.

Thomson, P., & Jaque, S. V. (2018). Childhood Adversity and the Creative Experience in Adult Professional Performing Artists. *Frontiers in Psychology, 9*(111). doi:10.3389/fpsyg.2018.00111

Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013). Atypical Combinations and Scientific Impact. *Science, 342*, 468-472.

Wang, J., Veugelers, R., & Stephan, P. (2017). Bias against novelty in science: A cautionary tale for users of bibliometric indicators. *Research Policy, 46*(8), 1416-1436.

Weitzman, M. L. (1998). Recombinant Growth. *Quarterly Journal of Economics, 113*(2), 331-360.

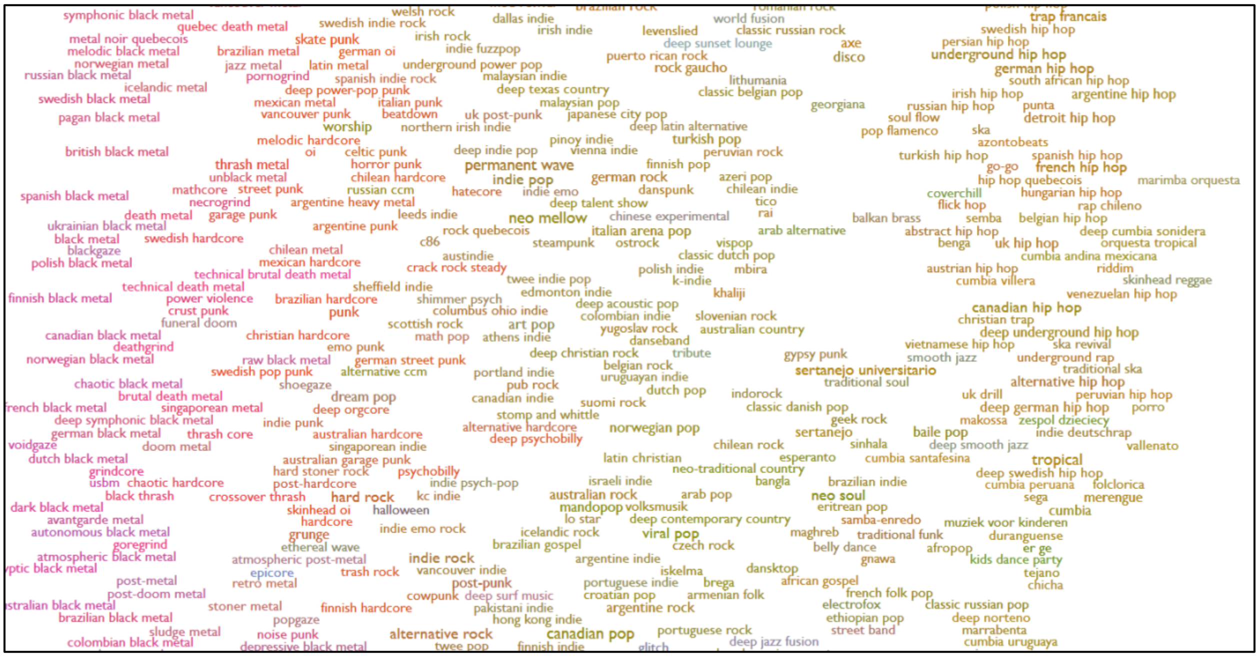
Youn, H., Strumsky, D., Bettencourt, L. M. A., & Lobo, J. (2015). Invention as a combinatorial process: evidence from US patents. *Journal of The Royal Society Interface, 12*(106). doi:10.1098/rsif.2015.0272

Table 1 – Testing theories of creativity

N = 114	Conceptual Innovator	Experimental Innovator	Diversity / Teams	Recombi-nation	The medium is the message	Serendipity	Adversity
Total	15	27	57	81	24	19	36
%	13%	24%	50%	71%	21%	17%	32%

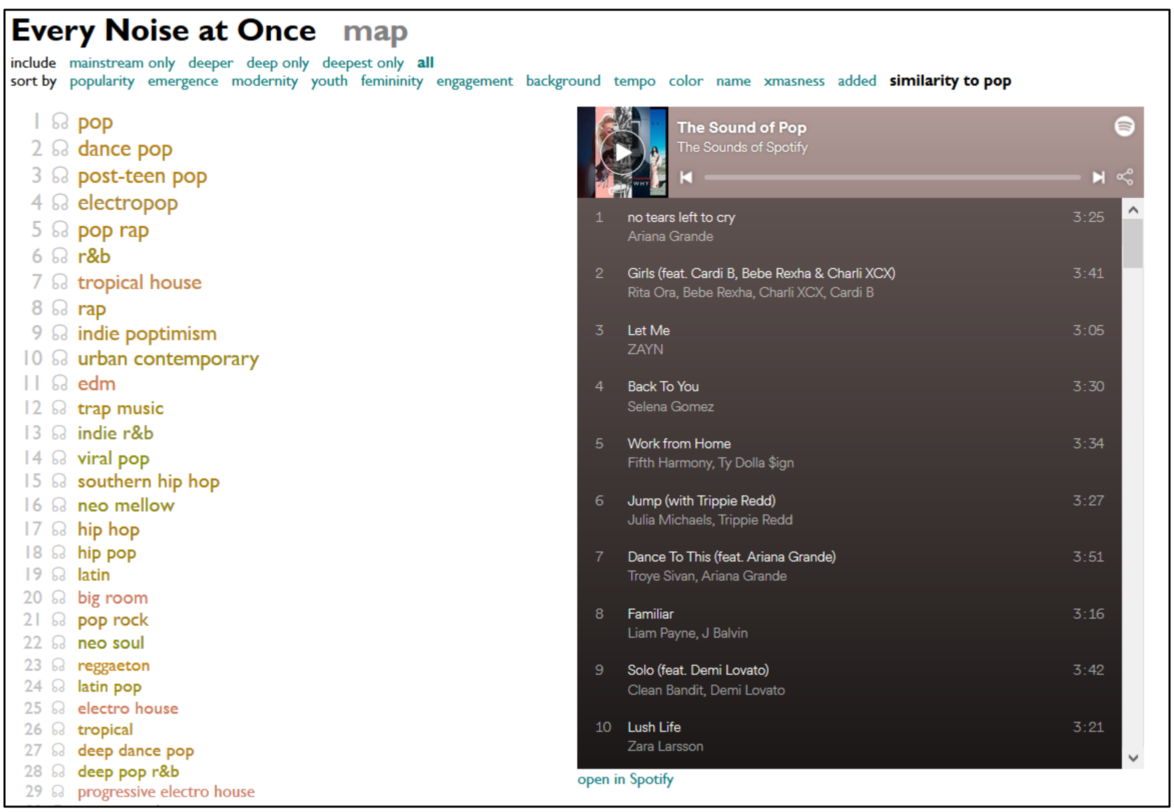
Source: Calculated from Song Exploder podcasts http://songexploder.net/.

Figure 1 – Every Noise at Once Genre Map (partial excerpt)



Source: http://everynoise.com/engenremap.html

Figure 2 – Every Noise at Once (partial excerpt)



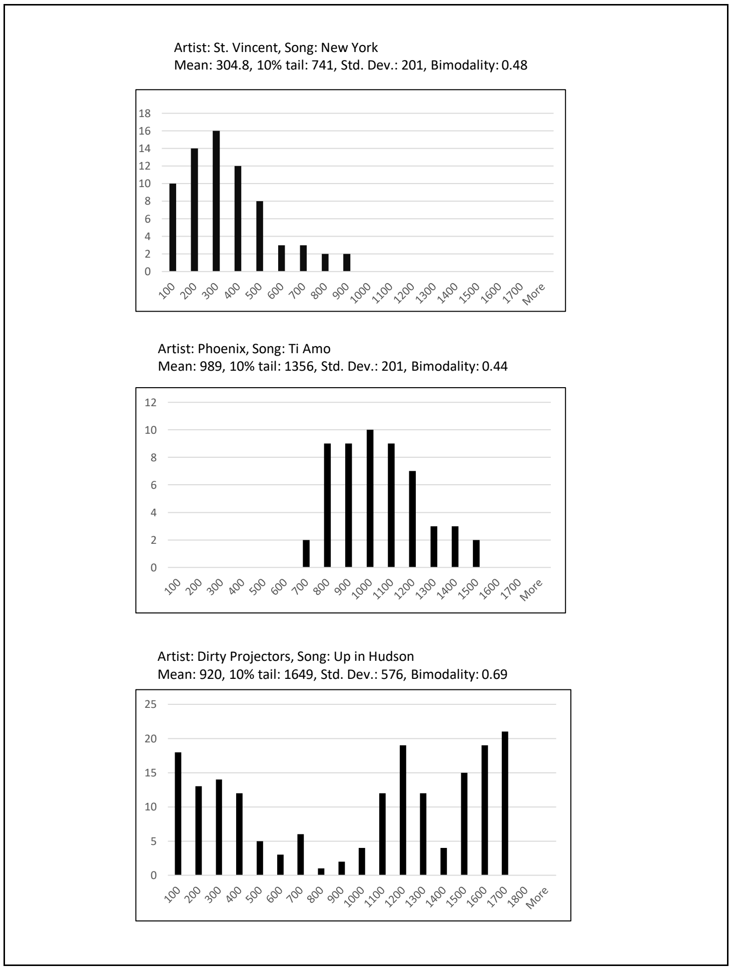
Source: http://everynoise.com/engenremap.html

Table 2 – Derivation of distance statistics for a sample song.

Artist: Gorillaz	**Cited inspiration: Michael Jackson**	**Cited inspiration: Hall and Oates**			
Song: Andromeda	Song: Billie Jean	Song: I can´t go for that			
Genre	**Genre**	**Genre**	**Genre pairs**		**Distance**
Alternative hip hop	Pop	Soft rock	Alt. hip hop	Pop	242
Art pop	Dance pop	Mellow gold	Alt. hip hop	Dance pop	251
		Album rock	Alt. hip hop	Soft rock	564
		Rock	Alt. hip hop	Mellow gold	792
		Folk Rock	Alt. hip hop	Album rock	848
			Alt. hip hop	Rock	736
			Alt. hip hop	Folk Rock	820
			Art pop	Pop	791
			Art pop	Dance pop	932
			Art pop	Soft rock	209
			Art pop	Mellow gold	287
			Art pop	Album rock	281
			Art pop	Rock	263
			Art pop	Folk Rock	528
				Mean	538.86
				10% tail	932
				Std. Dev.	274.9
				Bimodality	0.522

Source: Calculated using data from Song Exploder, http://songexploder.net/, and Every Noise at Once, http://everynoise.com/engenremap.html. *Andromeda*, the song by the Gorillaz is the original song. *Billie Jean* by Michael Jackson and *I can't go for that* by Hall and Oates are cited as inspiration. The genre lists for each artist are taken from the Every Noise at Once homepage. There are 14 genre pairs linking the original song to its inspirations. The distance is calculated using the similarity function in the Every Noise at Once homepage to order the list of 1800-plus genres in terms of similarity to the original song's genres. The distance is measured as the rank of the target genre from the original song' genre.

Figure 3 – Histogram and statistics for a sample of song observations



Source: Calculated using data from Song Exploder, http://songexploder.net/, and Every Noise at Once, http://everynoise.com/engenremap.html.

Table 3 - Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
YouTube-views	104	6,878.84	19,059.02	3.99	104,385.5
Mean distance	104	358.77	393.38	0	1623.78
10% tail	104	686.56	585.67	0	1763.00
Std. deviation	104	206.82	179.82	0	642.31
Spotify-listens	104	1.49	2.03	0.003	10.75
Bimodal	104	0.39	0.29	0	0.91
Views per day	104	6,879	19,059	3.99	104,386
Year started	104	2000	9.07	1960	2015
Billboard200	104	98.59	91.81	1	250.0
Major label	104	0.42	0.50	0	1
# of genres	104	6.65	2.85	1	10
Pop	104	0.18	0.39	0	1
Rap	104	0.07	0.25	0	1
Rock	104	0.09	0.28	0	1
Folk	104	0.08	0.27	0	1
Indie	104	0.27	0.45	0	1
Soundtrack	104	0.13	0.34	0	1
Electronic	104	0.11	0.31	0	1
Metal	104	0.04	0.19	0	1
Other	104	0.05	0.21	0	1

Table 4 – Determinants of songs' impact

	[1]	[2]	[3]	[4]	[5]
Dependent variable: Average number of YouTube-views per day since song`s video debut (logs)					
Billboard200	-0.002		-0.003	-0.004	-0.004
	(-0.80)		(-1.35)	(-1.01)	(-1.12)
Major label	0.161		0.135	0.227	0.162
	(0.37)		(0.33)	(0.39)	(0.30)
Year started	0.083^{***}		0.088^{***}	0.060^{*}	0.077^{***}
	(3.66)		(4.05)	(1.77)	(2.73)
Spotify-listens	0.510^{***}		0.495^{***}	0.453^{***}	0.491^{***}
	(4.19)		(4.29)	(2.76)	(3.35)
Mean distance		0.004^{***}	0.004^{***}	0.004^{***}	0.004^{***}
		(2.79)	(3.44)	(2.87)	(3.26)
10% tail		-0.003^{**}	-0.003^{**}	-0.003	-0.003^{**}
		(-1.98)	(-2.25)	(-1.51)	(-2.02)
Std. deviation		-0.0004	-0.002	-0.003	-0.002
		(-0.11)	(-0.78)	(-0.88)	(-0.70)
Bimodal		3.027^{**}	2.754^{**}	3.312	3.341^{*}
		(2.14)	(2.31)	(1.61)	(1.75)
Rap				-0.724	
				(-0.73)	
Rock				-0.137	
				(-0.13)	
Folk				-0.822	
				(-0.64)	
Indie				-0.862	
				(-1.07)	
Soundtrack				-1.260	
				(-1.02)	
Electronic				-1.083	
				(-1.12)	
Metal				-0.731	
				(-0.46)	
Other				0.227	
				(0.18)	
Constant	-159.84^{***}	5.96^{***}	-169.99^{***}	-114.11^{*}	-149.60^{**}
	(-3.53)	(15.72)	(-3.91)	(-1.68)	(-2.63)
Observations	104	104	104	104	71
Adjust. R^2	0.260	0.065	0.346	0.269	0.327
Prob.>F	0.0000	0.0310	0.0000	0.0045	0.0000

Minimum least-square estimation. t-stats in parentheses.