

Mueller, Bernardo

Article

Where'd you get that idea? Determinants of creativity and impact in popular music

EconomiA

Provided in Cooperation with:

The Brazilian Association of Postgraduate Programs in Economics (ANPEC), Rio de Janeiro

Suggested Citation: Mueller, Bernardo (2021) : Where'd you get that idea? Determinants of creativity and impact in popular music, EconomiA, ISSN 1517-7580, Elsevier, Amsterdam, Vol. 22, Iss. 1, pp. 38-52,
<https://doi.org/10.1016/j.econ.2021.02.002>

This Version is available at:

<https://hdl.handle.net/10419/266972>

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/>



Where'd you get that idea? Determinants of creativity and impact in popular music

Bernardo Mueller

Department of Economics, Universidade de Brasília, Brasília, DF 71665-185, Brazil



ARTICLE INFO

Article history:

Received 11 March 2019

Received in revised form 17 January 2021

Accepted 22 February 2021

Available online 27 February 2021

JEL classification:

O3

Z1

D16

D7

Keywords:

Popular music

Creativity

Recombination

Innovation

ABSTRACT

What are the determinants of creativity, novelty, and innovation? 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 uses the average number of daily *YouTube* views per day since the songs' video release as a measure of the song's impact, and tests how this impact is associated with the seven theories. 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 1800 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 used to measure different recombination strategies in a regression that seeks to explain songs' relative success, while controlling for other determinants, 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 success.

© 2021 The Author. Production and hosting by Elsevier B.V. on behalf of National Association of Postgraduate Centers in Economics, ANPEC. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

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

E-mail address: bmuel@unb.br

URL: <http://bpmmueller.wixsite.com/bernardo-mueller>.

<https://doi.org/10.1016/j.econ.2021.02.002>

1517-7580/© 2021 The Author. Production and hosting by Elsevier B.V. on behalf of National Association of Postgraduate Centers in Economics, ANPEC. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

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 to and recombination of already existing ideas. This paper explores these and other hypotheses described below. I test for the determinants of creativity and innovation 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 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 capture attention.

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 music; [Uzzi et al. \(2013\)](#) and [Wang et al. \(2017\)](#) for scientific publications; [Youn et al. \(2015\)](#) for patents and inventions; and [Barron et al. \(2018\)](#) 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 (114 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 Spotify’s music-intelligence division Echo Nest (originally a spin-off from the MIT Media Lab). It 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 1800 genres of music, from a capella to zydeco, passing through brutal death metal and capoeira as well as crunk, fado and gnawa, among many others.⁵ 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.

This paper is related to the scholarly literature on the economics of music, which covers a wide variety of topics, from the demand for music, production of music, distribution, prices, labor markets, live music, competition and others (see [Cameron \(2018\)](#) for recent anthology). More specifically, it focuses on the theme of creativity in music and provides an attempt to

¹ For information on the film *Amadeus* see https://www.imdb.com/title/tt0086879/?ref_=fn_al_tt_1.

² <http://songexploder.net/>.

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/>.

⁵ The number of genres has now increased to 5169 (January, 2021), but when the original data analysis in this paper was done in early 2018 there were approximately 1800 genres.

empirically estimate the determinants of success, as have others in this literature (Crain and Tollison, 1997; Giles, 2007; Earl and Potts, 2013; Askin and Mauskapf, 2017, among others). Building on these earlier contributions I add the novel use of music-related podcasts and new metrics for measuring success and music genres derived from recent innovations in the digital distribution of music (*Youtube* views and the *Every Noise at Once* platform).

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, as there seems to be a lack of consensus in the broad field of Creativity regarding definitions, measurement and determinants.⁶ 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 is briefly described below, leaving details to the citations therein. I also describe the protocols used to classify each of the *Song Exploder* podcasts into the set of theories given the composers' narrative.⁷ 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 songs' relative success, conditional on the songs already being successful in terms of sales, airplay or notoriety.

2.1. 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 Galenson (2006, 2009a). 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 (Galenson, 2006; Ginsburgh and 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. For songwriters Galenson (2009b) contrasts Irvin Berlin and Cole Porter, who worked experimentally and peaked late in their careers, to Bob Dylan and the duo John Lennon and Paul McCartney, who worked conceptually and did their path-breaking work early on. Galenson (2009b: 25) notes that during the 1970s:

...like Bob Dylan, Lennon and McCartney had passed the time when innovative lyrics and melodies would simply be given to them, as if by spirits. Much like the aging Dylan, at 62 McCartney could only look back at his own early work with a mixture of wonder and pride: "you can't go on making *Pepper* all your life. It just can't be done. *Pepper* was a peak. ... It hit the right notes, the right moment in time. It was strange. It was weird. It was hugely popular. It embraced all the things you want to embrace as an artist. ... But it was always going to be impossible to top."⁸

Galenson's categories refer to the entire career of the creator, whereas my data refers to a specific song. The songs were classified under these theories if the composer's narrative suggested a creative process that was fast, rational and planned (conceptual) or open-ended, incremental and tentative (experimental).

2.2. 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 (Cerisola, 2018; Hong and Page, 2004; Page, 2008, 2011; Meiseberg and Ehrmann, 2013; and 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. Wuchty et al. (2007: 1036) have shown that in academic publications in science "research is increasingly done in teams across nearly all fields ... and teams typically produce more frequently cited research than individuals do, and this advantage has been increasing over time." The same is likely the case for music.

Songs were attributed to this theory when the composer's narrative explicitly cites the participation of someone outside the artist's usual collaborators (band members, old co-authors, usual producers).

⁶ For good reviews see Kozbelt et al. (2010); Runco and Jaeger (2012); Simonton (2012), and Batey (2012).

⁷ The text briefly describes how I ascribed the theories to different songs in the sample. For those interested in greater detail I make available upon request a spreadsheet with the transcripts of all the parts in the podcasts where a composer makes a statement that leads the song to be classified under a given theory.

⁸ The McCartney quote is cited from Sawyers (2006: 249–250).

2.3. Theory 4 – recombination

Novelty, creativity and innovation often do not come from scratch, but are the result of the recombination of known and tested elements, be they songs (Askin and Mauskopf, 2017), scientific papers (Uzzi et al., 2013; Wang et al., 2017), technologies (Arthur, 2009; Mokyr, 1990), phenotypes (Darwin, 1859), economic capabilities (Hidalgo and Hausmann, 2009), ideas (Weitzman, 1998), beliefs (Mueller, 2016); or property rights (Alston and 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).⁹

Songs were classified under the recombination theory 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. No distinction will be made in the empirical analysis whether one or more than one song or artist are cited as influence. The distinction is simply whether an influence was cited or not. The main analysis will be performed using the full sample of songs, including those where recombination was present and those where it was not. Nevertheless, the test will also be run using only the observations linked to recombination as a check on robustness.

2.4. Theory 5 – the medium is the message

McLuhan (1964) famously argued that ‘the medium is the message’, that is that the medium through which content is delivered is often more important and impactful than 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. For example, according to Galenson (2009b: 26) “the Beatles changed the sound of popular music, by introducing novel production techniques, instruments that had not previously been used for pop songs, and electronic music.”

This theory was ascribed to songs in the sample when the composer explicitly stated that the creative process was meaningfully affected by something related to an instrument, studio, recording equipment, or some related situation.

2.5. 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.” In *Music in the Marketplace*, Cameron (2015: 78–83) considers a production function for music where in addition to labor and capital, there are three more unusual inputs: serendipity, authenticity and drugs. I have included ‘serendipity’ as a theory, but not drugs or authenticity given the difficulty of detecting or measuring these aspects.

Songs were classified under this theory when the narrative explicitly referred to an accidental or serendipitous role in the composition of the song.

2.6. Theory 7 – adversity

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 associated to songs in which the composer’s narrative made an explicit link of the song’s composition to some adverse or negative experience.

2.7. 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

⁹ 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.

Table 1
Testing theories of creativity.

N = 114	Conceptual Innovator	Experimental Innovator	Diversity/ Teams	Recombination	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/>.

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.”¹⁰ Several others also cited existing songs and explicitly stated that they wanted to make something similar.

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 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 Killah¹¹)

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. Data

In the next section I test whether the seven theories described above are associated with impact as measured by the average daily views of the songs’ videos on *YouTube*. I regress the logarithm of this value against indicator values for each of the theories, remembering that each song can be associated with more than one theory. Ideally, I would like to have more detailed measures of how each song fits each theory rather than a simple dummy variable. For example, was a song classified under the ‘adversity’ category influenced by heartbreak or by death of a loved one or by something else? Similarly, if ‘serendipity’ played a role, what was the nature of the serendipity? However, further classifying the composers’ narratives to this level of detail would introduce too much subjectivity so dummies were used. For the ‘recombination’ theory however, there is an objective way to quantify the different recombination strategies used by the artists, so for this variable the measure will be more elaborate than a simple indicator variable. In this section I describe these measures of recombination.

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 the next section is not to uncover the formula for the perfect pop song, but rather to 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?¹²

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

¹⁰ *Song Exploder* podcast # 99 with Sleigh Bells, <http://songexploder.net/sleigh-bells>.

¹¹ *Song Exploder* podcast # 26 with Ghostface Killah <http://songexploder.net/ghostface-killah>.

¹² [Earl and Potts \(2013: 154\)](#), find that in the quest for novelty songwriters often go too far and overshoot with the changes they make, thus alienating their fans and customers.

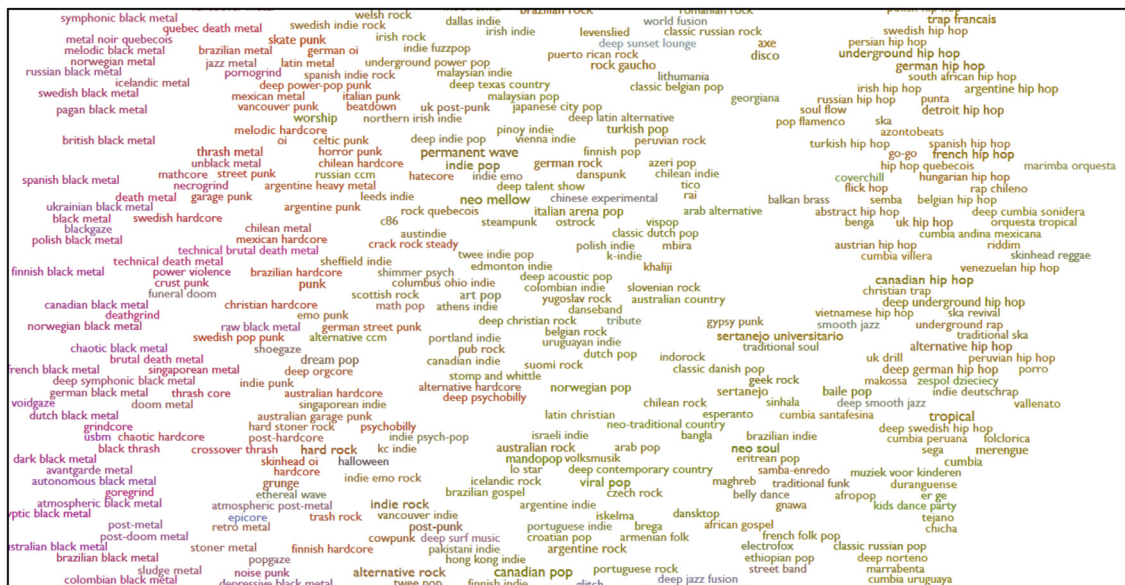


Fig. 1. Every noise at once genre map (partial excerpt).

Source: <http://everynoise.com/engrenmap.html>.

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 Mauskopf (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 and Mauskopf, 2017, p. 6)

It is important to note that although their result finds the same basic signature to creative success as mine, both papers differ in several of the data sources, sample of songs, explanatory variables, methods of measuring many of those variables, and especially in the fundamental driver of creativity investigated. While they rely on song features mechanically collected through music information retrieval, I rely on the composer’s accounts through the *Song Exploder* podcast.

In this paper I test for what Uzzi et al. (2013) call ‘atypical combinations’ and Askin and Mauskopf (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. Fig. 1 provides a snapshot of part of the musical-genre map, which contains more than 1800 different genres spatially placed according to their similarity based on algorithmically processed data from Spotify listeners’ habits.¹³ The dimensions in this figure are not determined ex-ante, but rather emerge from the data. The vertical dimension appears to have more electric genres at

¹³ View the entire map and see more information on how the map is built in <http://everynoise.com/> and in its creator’s (Glenn McDonald) blog <http://www.furia.com/>.

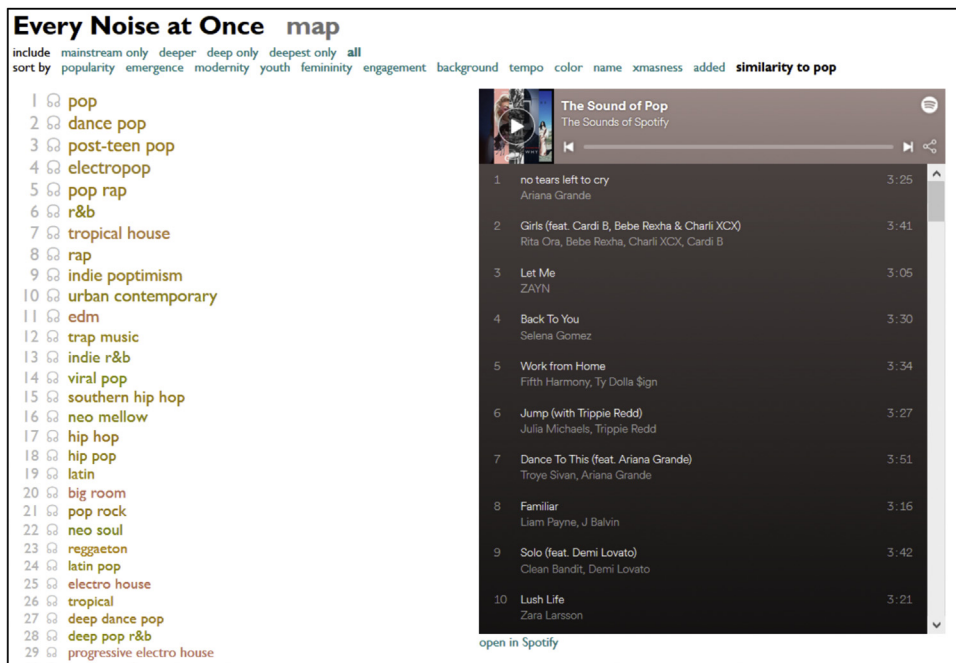


Fig. 2. Every noise at once (partial excerpt).

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

Table 2

Derivation of distance statistics for a sample song.

Artist: Gorillaz Song: Andromeda Genre	Cited inspiration: Michael Jackson Song: Billie Jean Genre	Cited inspiration: Hall and Oates Song: I can't go for that Genre	Genre pairs	Distance	
Alternative hip hop Art pop	Pop Dance pop	Soft rock Mellow gold Album rock Rock Folk Rock	Alt. hip hop	Pop	242
			Alt. hip hop	Dance pop	251
			Alt. hip hop	Soft rock	564
			Alt. hip hop	Mellow gold	792
			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's genre.

the top and more acoustic genres at the bottom, while the horizontal dimension varies from denser sounds on the left to sparser and spikier sounds on the right.

Fig. 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.¹⁴

Although the chart in Fig. 1 is often referred to as a 'map', that is not quite accurate. The figure is instead a readability-adjusted scatter plot that reduces the 12 dimensions into which each song is classified into two dimension that can be more easily perceived by the human eye. The list in Fig. 2 then reduces the map to a single dimension that is even simpler to apprehend. Yet the ordering in both renditions of the map are determined by the entire corpus of information used by the *Every Noise at Once* analytic process. I have not been able to find a rigorous technical description of what that process does exactly, probably because it is proprietary and since 2014 Echo Nest has been owned by Spotify.¹⁵ In his blog and in interviews, Glenn McDonald, the main developer, provides only vague notions of how figure is produced, for example; "This map builds on essentially everything else we do at the Echo Nest . . . from parsing the content of scraped web pages to analyzing the psycho-acoustic properties of the actual files."¹⁶

If the figure were a map, one would expect that the distance from point A to point B should be the same as the distance from point B to point A. That is, the distances between the genres should be symmetrical. But this is not the case. For example, ordering the list in Fig. 2 by similarity to *Motown*, the genre *Folk rock* is ranked at 53. Ordering it by similarity to *Folk rock*, leaves *Motown* at number 78. The way to understand this counterintuitive property is by considering the classification not as an expression of distance, but through the lens of Information Theory, which is the basis for techniques like machine learning, natural language processing, topic modelling, music information retrieval (MIR), among others (Shannon, 1948). Consider that you know the entire list of genres and I tell you that a given song is listed as a member of the *Rock* genre. This information does not tell you which song it is, nor exactly what it sounds like. But the information has narrowed down the possibilities, that is, it has reduced your level of uncertainty (reduced entropy). If I then tell you that the song was inspired by another song that belongs to a different genre, that will further reduce your uncertainty. But by how much? If the genre of the second song is very similar to that of the first, say *Classic rock*, this would provide less information than if the second genre were something like *Swedish synth*. In other words, information can be thought of as surprise, something you did not already know. More 'distant' genres are those that form more surprising pairs relative to expectations, and therefore represent more novelty. Note that switching the order of the genres is not symmetrical in terms of information or surprise. If you know a song's genre is *Classic rock*, you practically already know that it is also in the *Rock* genre, whereas many *Rock* songs are not considered *Classic Rock*.¹⁷ In the appendix of this paper I provide a more rigorous notion of the information theoretic approach through which the *Every Noise at Once* data is best understood.¹⁸

The purpose of compiling the various genre-pairs is to produce four different statistics that describe the genre-pair distribution that are then used to measure songs' success. The first is the mean, which captures how far on average the song reached out for inspiration. To assess the magnitudes, consider that the ranking runs from 1 to 1873. The second 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 third measure is the standard deviation of the distribution of genre-pairs, to capture the concentration vs dispersion of the influencing songs' genres. Finally, the fourth 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

¹⁴ 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.

¹⁵ The Echo Nest home page describes some of the services they sell and list some of the customers: "Leading music services (Clear Channel's iHeartRadio, Rdio, SiriusXM, Spotify), editorial, video and social media networks (The BBC, Foursquare, MTV, Twitter, VEVO, Yahoo!), connected device manufacturers (doubleTwist, Nokia), and big brands (Coca Cola, Intel, Microsoft, Reebok) all rely on The Echo Nest to reinvent the way people interact with music." <http://the.echonest.com/customers/>

¹⁶ <https://www.dailydot.com/debug/every-noise-at-once-echo-nest/>

¹⁷ In this example one genre is a subset of the other. But the asymmetry is not limited to such cases. The classification of genres considers 12 dimensions as well as other non-musical information, such as geography or main instruments, so that the link among genres can be complex.

¹⁸ Examination of the genre rankings sometimes reveals classifications that seem anomalous given our expectations. This is how Glenn McDonald referred to this issue in a post on a blog of the Royal Statistical Society of the UK: "Like most computational processes based on human data and human love, this one doesn't always work perfectly. Some of the genres that end up as neighbors on the map don't sound all that similar to human listeners. Some of the deductions the math makes are based on confused or erroneous data. Some of the songs are not ideal demonstrations of what they are trying to explain. Some of the internal checks and balances to mitigate wrong things also invisibly eliminate a few good things. None of this is intended to replace people loving music and sharing it with each other. But real maps aren't always perfect, either. The surprising thing is how well the whole thing works, despite all its inevitable flaws and binary obtuseness." <https://www.statslife.org.uk/culture/1095-every-noise-at-once-using-big-data-to-explore-new-music>

than one place.¹⁹ 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 Fig. 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.

4. Testing the determinants of songs' success

With these statistics we can proceed to estimate the determinants of creativity and innovation. The dependent variable is the average number of daily views on *YouTube* of the song's official (or most viewed) video since the day it was launched.²⁰ 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.²¹ Most videos follow a dynamic such that they attract many views immediately after their release and as time goes by the number of views decreases asymptotically. The average age of the songs in my sample is approximately three years, so most were already past the prime viewership phase. Nevertheless, I add the number of days since the song's release as a control for this effect.

The explanatory variables are the measures of which theories of creativity were associated with each song in the *Song Exploder* podcast. These are dummy variables for each theory except for 'recombination', where the four genre-pair distribution statistics were used. Also, it is necessary to control for artists' characteristics that may lead to *YouTube* video views independent of the song's intrinsic quality. If a famous band, such as *Metallica*, *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.²² I use six 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. Not all of the artists in the sample have managed to have a *Billboard* 200 hit, so the second variable uses the best placement in any of the secondary *Billboard* charts, for example, *Hot* 100, or *Hot Latin Song*. Some of the songs in the sample were used in soundtracks for films or TV and were thus not likely to be in *Billboard* charts. For these I have a dummy that is equal to 1 if they received an Oscar, Grammy, Emmy nomination or any other significant award. The idea of these variables is to capture an artist's popularity at the time it released the song in question.

The fourth 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 (Bishop, 2005). The fifth 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 sixth 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.²³

Another set of controls are dummies for some of the major genres: Pop, Rap, Rock, Folk, Indie, 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 and a summary description of each variable is provided in Table A1 in the appendix. There are 104 observations, as 10 of the cited influences were not actual artists, but a generic beat, a cartoon, a Christmas song, or other such inspiration that could not be found in the *Every Noise at Once* database. Of the 104 observations, 71

¹⁹ The bimodality coefficient is calculated through the following formula: $BC = \frac{m_3^2 + 1}{m_4 + 3\sqrt{\frac{(n-1)^2}{(n-2)(n-3)}}}$, 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 et al., 2013).

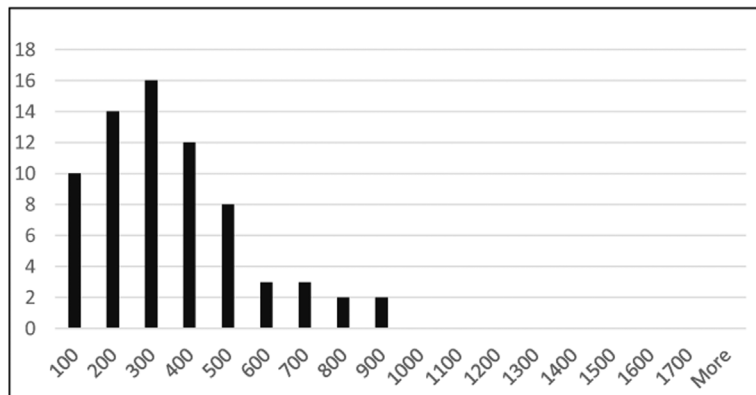
²⁰ I take the log of the average daily *YouTube* views to deal with the skewness that arises from having some very big hits in the sample. Results are similar without this transformation.

²¹ Gloor (2011: 78–79) discusses the advantages and disadvantages of using the *Billboard* charts as a measure of success. He notes that few artists ever make it to the charts and "of the artists who debut a song on the *Billboard* Hot 100 each year, almost half of them have never placed another song there again. Around 34–40% of the artists who placed an album on the *Billboard* Top 200 have not returned to the charts with a subsequent album."

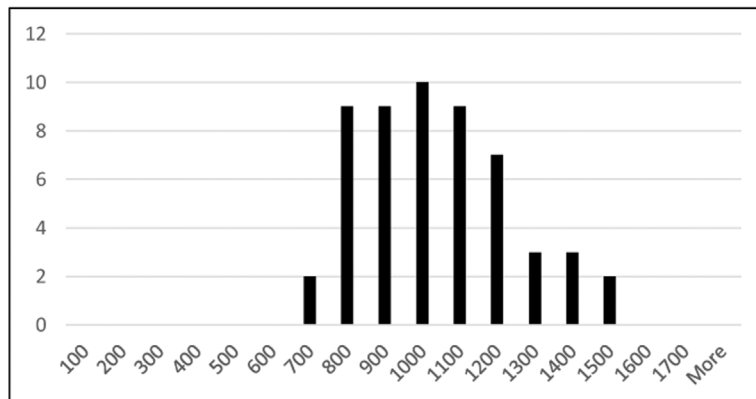
²² What I am calling 'inertial success' is the converse of what Hendricks and Sorensen (2009) call 'backward spillovers.' In the former, past success leads to greater impact of the new song. In the latter, a new song raises interest and sales of past releases.

²³ 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.

Artist: St. Vincent, Song: New York
 Mean: 304.8, 10% tail: 741, Std. Dev.: 201, Bimodality: 0.48



Artist: Phoenix, Song: Ti Amo
 Mean: 989, 10% tail: 1356, Std. Dev.: 201, Bimodality: 0.44



Artist: Dirty Projectors, Song: Up in Hudson
 Mean: 920, 10% tail: 1649, Std. Dev.: 576, Bimodality: 0.69

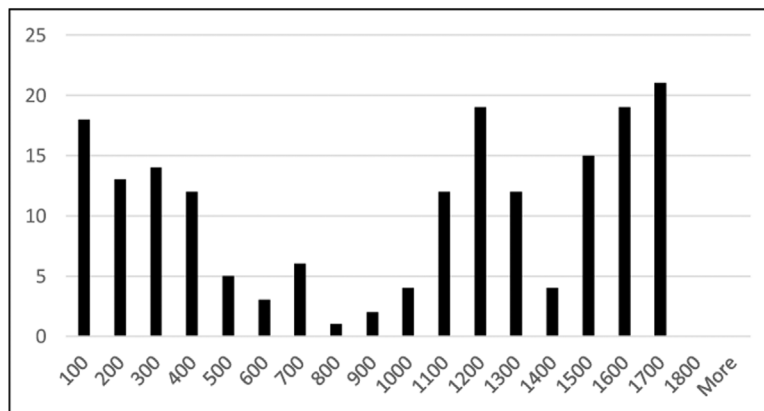


Fig. 3. Histogram and statistics for a sample of song observations.

Source: Calculated using data from Song Explorer, <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
Year started	104	2000	9.07	1960	2015
Billboard200	104	92.39	82.84	1	200
Other charts	104	90.81	25.93	3	100
Awards	104	0.13	0.33	0	1
Days since release	104	1227.9	814.07	62	4458
Major label	104	0.42	0.50	0	1
Conceptual	104	0.135	0.343	0	1
Experimental	104	0.260	0.441	0	1
Serendipity	104	0.183	0.388	0	1
Teams	104	0.510	.502	0	1
Medium/Message	104	0.212	0.410	0	1
Adversity	104	0.327	0.471	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

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. Column 1 shows the specification with only variables for the seven theories, plus the control for the number of days since the video's release. The variables that measure the genre-pair distribution statistics indicate that there is a pattern to how the recombination of existing ideas into a new song affects its success. All these recombination strategy 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. None of the other theories are found to have a statistically significant effect in this specification.

In column 2 the variables that measure the artists' popularity and stature at the time of the release of the song in question are included. The strongest effect comes from the artists' monthly Spotify-listens, a measure of artist popularity, which is statistically significant at 1%. The variable is in units of one-million listens, so the estimated coefficient of 0.47 indicates that an additional 1 million listens to the artist on Spotify is associated with 60% additional *YouTube* daily views of that song's video.²⁴ The major-label dummy is 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*.²⁵ None of the three variables that measure the artists' previous chart success or awards is significant. The inclusion of these controls does not change the results from column 1. Together, these variables explain almost one third of the variability of the songs' average daily *YouTube* views.

²⁴ Because the estimation is done in log-linear form the interpretation of the coefficients requires that they be transformed by the formula $\% \Delta dep. variable = (e^{\beta} - 1) * 100$ and interpreted as percent change in the dependent variable (where β is the estimated coefficient shown in Table 4).

²⁵ The average starting year in the sample is 2000, the minimum is 1960, and the maximum 2015.

Table 4
Determinants of songs' impact measured as *YouTube* view.

	[1]	[2]	[3]	[4]
Dependent variable: Average number of <i>YouTube</i> -views per day since video debut (logs)				
Mean distance	0.004** (2.49)	0.004*** (3.22)	0.004*** (2.93)	0.004*** (2.69)
10% tail	-0.003** (-2.14)	-0.003** (-2.39)	-0.003** (-2.08)	-0.003* (-1.91)
Std. deviation	-0.0001 (-0.03)	-0.001 (-0.45)	-0.001 (-0.42)	-0.001 (-0.37)
Bimodal	3.470** (2.39)	3.015** (2.33)	2.636* (1.83)	2.681 (1.21)
Conceptual	0.477 (0.69)	0.164 (0.26)	0.253 (0.37)	0.146 (0.16)
Experimental	0.227 (0.44)	-0.200 (-0.40)	-0.280 (-0.53)	-0.382 (-0.48)
Serendipity	-0.921 (-1.53)	-0.475 (-0.85)	-0.359 (-0.60)	-0.481 (-0.58)
Team	0.297 (0.64)	0.370 (0.91)	0.457 (1.07)	1.142* (1.90)
Medium is Message	-0.250 (-0.45)	0.079 (0.16)	-0.017 (-0.03)	-0.163 (-0.22)
Adversity	-0.029 (-0.06)	-0.097 (-0.22)	-0.122 (-0.25)	0.119 (0.18)
Spotify-listens		0.467*** (3.69)	0.464*** (3.43)	0.421** (2.34)
Major label		0.063 (0.13)	0.016 (0.03)	-0.065 (-0.10)
Year started		0.081*** (3.27)	0.077*** (3.00)	0.068* (1.89)
Billboard200		-0.005 (-1.34)	-0.006 (-1.46)	-0.006 (-1.14)
Other charts		-0.006 (-0.63)	-0.004 (-0.35)	-0.008 (-0.58)
Awards		0.523 (0.63)	0.172 (0.18)	-0.812 (-0.63)
Days since release	-0.0007** (-2.45)	-0.0003 (-1.16)	-0.0002 (-0.76)	-0.0001 (-0.19)
Constant	6.854 (11.21)	-154.50*** (-3.12)	-147.50*** (-2.85)	-127.92* (-1.78)
Genre dummies	No	No	Yes	No
Observations	104	104	104	71
Adjust. R ²	0.11	0.31	0.27	0.24
Prob. > F	0.0278	0.0000	0.0005	0.0248

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

* 10%.

** 5%.

*** 1%.

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 552. If the same artist had an additional 1 million Spotify listens, then the number of views would increase to 881. Similarly, increasing the year since first active from 2000 to 2010 would yield an increase in views to 1195. Naturally, this is a statistical manipulation with 'average' artist characteristics. The regression's adjusted R-squared of 0.31 indicates that more than two-thirds of the variation in views is not explained by this specification, so for any specific artist there are many other factors influencing success and creativity.

The effects of the recombination 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 1875 genres (from 358 to 458), the number of daily *YouTube*-views increases from 552 to 833 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 less *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. 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 900 to 1217 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 innovation.

In column 3 genre dummies are added to the specification. The results from columns 1 and 2 remain very much the same and none of the genre dummies are statistically significant. The adjusted R-squared falls with the inclusion of the genre dummies indicating that these do not contribute much towards the explanation of the variability of *YouTube* views.

Perhaps the best way to understand these results is in terms of the genre-pair histograms presented in Fig. 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. 1356), 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 2 would be 572. Keeping the six 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 521 for St. Vincent's New York, 934 for Phoenix's Ti Amo, and 332 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 70% 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 4 repeats column 2 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 estimated coefficient signs and magnitudes. In this specification bimodality is no longer significant, but the dummy for songs classified under 'teams' has become significant at 10%.

5. Conclusion

The test in this paper sought to explain the determinants of success for a sample of songs for which I had detailed information about how they were conceived. 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 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 success and impact of optimal differentiation and atypical combinations (Askin and Mausekapp, 2017; Barron et al., 2018; Uzzi et al., 2013).

There are several avenues for future research to extend this investigation. One is to wait for more episodes of *Song Exploder* to air and then see if the extended sample confirms the results.²⁶ Another is to look for other sources of data that allow the same type of exercise to be realized. There are several other podcasts that explore how songs are created that can be tapped into, though each has their own peculiarities and adaptations would be necessary. Some, for example, use unknown artists and songs that have not achieved widespread notoriety. These might lead to additional insights. Another idea is to look at the history of recorded music exploring different recordings of the same song by different artists or by testing similarity and success of follow-up songs by a given artist after a big hit. A final suggestion is an analysis of the distinction and blurring of recombination and plagiarism as a songwriting strategy.

²⁶ By January 2021 there were already 200 episodes. The idea has been so successful that the podcast has been transformed into a Netflix series, already in its second season.

Acknowledgements

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

Appendix A1

An information theoretic perspective on measuring genre distance across songs

This appendix illustrates more rigorously how a measure of novelty can be derived from the genre pairings through an information theoretic perspective. Let P_i be the probability of a song being of genre i , and similarly for P_j . Let $P(i,j)$ be the probability that a song simultaneously belongs to genre i and to genre j . A measure of how likely a song is to belong to both genres i and j , above the probability that the song would independently belong to each category can be written as:

$$S_{ij} = \log \frac{P(i,j)}{P(i)P(j)} \quad (1)$$

If S_{ij} is greater than zero, these two genres are co-occurring more than what would be expected by chance. If we add over all the genre pairs in which a given song's genre i participates (for example, all the pairs that contain *Alt. hip hop* in Fig. 2) we get an expression that measures how predictive knowing that a song is part of genre i is of all the other pairings:

$$S_i = \sum_{j=1}^N P(j|i) \left(\log \frac{P(i,j)}{P(i)P(j)} \right) \quad (2)$$

Finally, if we average over all genres for a song (for example, all 14 genre pairs in Fig. 2), then we have a measure of how predictive all the genre pairs for the song are of each other. In other words, it would be a measure of conventionality or lack of novelty.

$$C = \sum_{i=1}^N S_i P(i) \quad (3)$$

The statistics used in this paper were derived directly from the *Every Noise at Once* listings and not through these equations. The equations have been shown here to illustrate the type of information theoretic reasoning through which those listings are better understood than through the metaphor of a map.²⁷

Appendix A2

Table A1

Data description and sources.

Variable	Description	Source
YouTube-views	Total number of views of a song's video on <i>YouTube</i> from the day it was released to June 26, 2018, divided by the number of days in that period.	https://www.youtube.com/
Mean distance	Mean of the distribution of genre-pairs. See Section 3.	Calculated using genre map in http://everynoise.com/
10% tail	Rank distance in genre list from the new song genre to the 90% furthest inspiration song genre.	Calculated using genre map in http://everynoise.com/
Std. deviation	Standard deviation of the genre pair distances. See Section 3.	Calculated using genre map in http://everynoise.com/
Bimodal	Bimodality coefficient of the song's genre-pair distribution. See Section 3.	Calculated using genre map in http://everynoise.com/
Spotify-listens	Number of listens on artist's Spotify page circa June 2018.	https://www.spotify.com
Year started	Year in which the artist or band's career started.	https://www.wikipedia.org/
Billboard200	Highest position reached by artist previous to the songs release.	https://www.billboard.com/ and https://www.wikipedia.org/
Other charts	For artists that never made it to the Billboard 200 other Billboard charts were used.	https://www.billboard.com/ and https://www.wikipedia.org/

²⁷ For more details see <https://www.youtube.com/watch?v=-M2S7dqS0Xg>.

Table A1 (Continued)

Variable	Description	Source
Awards	Dummy = 1 if artists had received a major award previous to the song's release, eg, Grammy, Tony.	https://www.wikipedia.org/
Days since release	Number of days since the song's release to the date when the YouTube views variable was measured.	https://www.youtube.com/
Major label	Dummy = 1 if the artist was signed to a major label versus independent at the time of the song's release.	https://www.wikipedia.org/

References

- Alston, L., Mueller, B., 2015. Towards a more evolutionary theory of property rights. *Iowa Law Rev.* 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. *Am. Sociol. Rev.* 82 (5), 910–944.
- Barron, A.T., Huang, J., Spang, R.L., DeDeo, S., 2018. Individuals, institutions, and innovation in the debates of the French Revolution. *Proc. Natl. Acad. Sci.* 115 (18), 4607–4612. <http://dx.doi.org/10.1073/pnas.1717729115>.
- Batey, M., 2012. The measurement of creativity: from definitional consensus to the introduction of a new heuristic framework. *Creat. Res. J.* 24 (1), 55–65.
- Bishop, J., 2005. Building international empires of sound: concentrations of power and property in the 'Global' music market. *Pop. Music. Soc.* 28 (4), 443–471.
- Cameron, S., 2015. *Music in the Marketplace. A Social Economics Approach*. Routledge, Abingdon, UK.
- Cameron, S. (Ed.), 2018. *The Economics of Music*. Edward Elgar Publishing.
- Cerisola, S., 2018. Multiple creative talents and their determinants at the local level. *J. Cult. Econ.* 42 (May (2)), 243–269.
- Crain, W.M., Tollison, R.D., 1997. Economics and the architecture of popular music. *J. Econ. Behav. Organ.* 32 (February (2)), 185–205.
- Darwin, C., 1859. *On the Origin of Species by Means of Natural Selection, or Preservation of Favoured Races in the Struggle for Life*. John Murray, London, 1859.
- DeDeo, S., 2018. The data science of culture: from the French Revolution to string theory. In: Presentation at the University of California Merced, April 29, 2018, YouTube video: <https://www.youtube.com/watch?v=S7MYdqzOcy4>.
- Earl, P.E., Potts, J., 2013. The creative instability hypothesis. *J. Cult. Econ.* 37 (2), 153–173.
- Galenson, D.W., 2006. *Old Masters and Young Geniuses: Two Life Cycles of Artistic Creativity*. Princeton University Press, Princeton.
- Galenson, D.W., 2009a. *Conceptual Revolutions in Twentieth-Century Art*. Cambridge University Press, Cambridge.
- Galenson, D. (2009). "Innovators: Songwriters." NBER Working Paper Series, Working Paper No. 15511. <http://www.nber.org/papers/w15511>.
- Giles, D.E., 2007. Survival of the hippest: life at the top of the hot 100. *Appl. Econ.* 39 (15), 1877–1887.
- Ginsburgh, V., Weyers, S., 2006. Creativity and life cycles of artists. *J. Cult. Econ.* 30 (2), 91–107.
- Gloor, S., 2011. Just how long is your 'fifteen minutes'? An empirical analysis of artists' time on the popular charts. *J. Music. Entertain. Ind. Educ. Assoc.* 11 (1), 61–82.
- Hendricks, K., Sorensen, A., 2009. Information and the skewness of music sales. *J. Polit. Econ.* 117 (April (2)), 324–369.
- Hidalgo, C., Hausmann, R., 2009. The building blocks of economic complexity. *Proc. Natl. Acad. Sci. U. S. A.* 106 (26), 10570–10575.
- Holland, J., 1992. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence*. MIT Press, Cambridge.
- Hong, L., Page, S.E., 2004. Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proc. Natl. Acad. Sci. U. S. A.* 101 (46), 16385–16389.
- 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*. Penguin Group, New York.
- Kauffman, S.A., 1995. *At Home in the Universe: The Search for Laws of Self-organization and Complexity*. Oxford University Press, New York.
- Kozbelt, A., Beghetto, R.A., Runco, M.A., 2010. Theories of creativity. In: Kaufman, J.C., Sternberg, R.J. (Eds.), *Cambridge Handbook of Creativity*. Cambridge University Press, New York, pp. 20–47.
- McLuhan, M., 1964. *Understanding media: The Extensions of Man*. MIT Press.
- March, J.G., 1991. Exploration and exploitation in organizational learning. *Organ. Sci.* 2 (1), 71–87.
- Meiseberg, B., Ehrmann, 2013. Diversity in teams and the success of cultural products. *J. Cult. Econ.* 37 (February (1)), 61–86.
- Mokyr, J., 1990. *Lever of the Riches: Technological Creativity and Economic Progress*. Oxford University Press, New York.
- Mueller, B., 2016. Beliefs, institutions and development on complex landscapes. *Econ. Anal. Law Rev.* 7 (2), 474–495.
- Page, S.E., 2008. *The Difference How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*. Princeton University Press, Princeton.
- 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. *Front. Psychol.* 4, 700.
- Runco, M.A., Jaeger, G.J., 2012. The standard definition of creativity. *Creat. Res. J.* 24 (1), 92–96.
- Sawyers, J.S. (Ed.), 2006. *Read the Beatles: Classic and New Writings on the Beatles, Their Legacy, and Why They Still Matter*. Penguin Books.
- Shannon, Claude E., 1948. A mathematical theory of communication. *Bell Syst. Tech. J.* 27 (July & October), pp. 379–423 & 623–656.
- Simonton, D.K., 2012. Quantifying creativity: can measures span the spectrum? *Dialogues Clin. Neurosci.* 14 (1), 100–104.
- Taleb, N.N., 2007. *The Black Swan: The Impact of the Highly Improbable*, first edition. Random House, New York.
- Thomson, P., Jaque, S.V., 2018. Childhood adversity and the creative experience in adult professional performing artists. *Front. Psychol.* 9 (111), <http://dx.doi.org/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. *Res. Policy* 46 (8), 1416–1436.
- Weitzman, M.L., 1998. Recombinant growth. *Q. J. Econ.* 113 (2), 331–360.
- Wuchty, S., Jones, B., Uzzi, B., 2007. The increasing dominance of teams in production of knowledge. *Science* 316 (5827), 1036–1039.
- Youn, H., Strumsky, D., Bettencourt, L.M.A., Lobo, J., 2015. Invention as a combinatorial process: evidence from US patents. *J. R. Soc. Interface* 12 (106), <http://dx.doi.org/10.1098/rsif.2015.0272>.