



# Originality, influence, and success: a model of creative style

Karol Jan Borowiecki<sup>1</sup> · Caterina Adelaide Mauri<sup>2</sup>

Received: 25 May 2023 / Accepted: 30 May 2023 / Published online: 22 September 2023  
© European Union represented by the European Commission, Joint Research Centre 2023

## Abstract

Creative workers strive to achieve success and influence by producing original output. In this paper we define and measure originality and influence, based on a new model of style. We apply the methodology to Western classical music composed since the 15th century, and test it using extensive data on the content of musical compositions, popular success, and biographical information. We find that more original composers tend to be more influential upon the work of their later peers and more successful with present-day audiences. A positive association between originality and influence also holds across works by a given composer.

**Keywords** Creativity · Innovation · Influence · Style · Music history

**JEL Classification** J24 · N30 · O31 · Z11

An eminent artist will bring about a considerable change in the established modes of each of those arts, and introduce a new fashion of writing, music, or architecture.— Adam Smith (1759)

## 1 Introduction

Many scientists, entrepreneurs or artists are driven by the desire to be original, influential or successful. But little is known about how these variables interact. Adam Smith was the first economist to approach these topics, but since then few others have ventured into related inquiries, perhaps because originality is such an

---

✉ Karol Jan Borowiecki  
kjb@sam.sdu.dk

Caterina Adelaide Mauri  
caterina.mauri@ec.europa.eu

<sup>1</sup> Department of Economics, University of Southern Denmark, 5230 Odense, Denmark

<sup>2</sup> European Commission, Joint Research Centre, Via E. Fermi, 2749, 21027 Ispra, VA, Italy

elusive concept. And yet, originality is fundamental to economic growth. Without the originality of an inventor, there would be no inventions, and hence no technological progress. And without the desire to become influential or successful, some inventors would never become original.

This paper sheds new light on how these concepts are related by formalizing a model of creative style, based on which we define and compute measures of originality and influence. We seek to understand in particular, whether originality, influence, and popular success are conflicting goals in the creative process—such that creatives trade them off against one another—or whether they go hand in hand, perhaps as by-products of the “quality” of creative work.

The concept of style is of general interest and has been studied in various contexts, including business, politics, economics, sociology, and finance (e.g., Bertrand & Schoar, 2003; Chan et al., 2018; Malmendier et al., 2021; Aggarwal & Woolley, 2019). Style has also a particular relevance in the arts, where it captures patterns of artistic production that are distinctive of an artist, place, or period. Our model of style will be applied to the arts, but its framework and insights are of general relevance. The model builds on a probability distribution over features (attributes) of creative works. Based on these distributions, we identify and measure the degree of originality and influence of a body of a creative’s work. The originality of a creator is the degree of uniqueness of her work relative to past work. This means that the reference point is the past: The more one deviates from the style of previous creators, the higher the originality. In contrast, influence looks to the future: The more predictive one’s work is of the style of future creators, the higher the influence.

We apply and test our framework in the context of Western classical music between the 15th and 20th centuries, which provides an ideal setting to analyze the interplay between originality, influence, and long-term success for several reasons: First, the data enables us to extract key features from a large number of musical themes—short but important excerpts of musical compositions. We are thus able to identify and measure the style of a composer or composition, going far beyond mere placement within a musical period. Second, the structured nature of classical music provides a setting where measures for such concepts as originality and influence can be constructed consistently and convincingly, and studied over the very long-term. Third, qualitative evidence from musicology and music history enables us to validate our methodology in a number of tests, and also to support our findings with various anecdotal accounts.

The core of our data has been collected from dictionaries of musical themes by Barlow and Morgenstern (1975, 1976) and includes 18,074 melodic themes from 6352 classical and operatic works by over 750 composers. The source provides sequences of musical notes, which are conveniently transposed to a common key, as well as a staff for each theme showing the original key and time signatures. We combine this data with measures of individual composers’ long-term success from modern consumption data based on Spotify streams as well as two measures of distinction: The Murray index (Murray, 2003) and the length of biographical entries in Grove Music Online, an encyclopedic dictionary of music and musicians. Finally, we collect information on the lives and work of composers from their biographies.

We find that more original composers tend to be more influential upon the work of their later peers and more successful with present-day audiences or experts. Influence and success increase monotonically with originality, though the relationship is more uncertain for very high levels of originality. The results hold also at the level of the individual theme, and are stable across different specifications as well as robust to the inclusion of various controls. The magnitude of the effects suggests an important role played by originality in determining both influence and success.

Finally, we develop a simple algorithm to randomly generate compositions. These tend to score highly on our measure of originality. We conclude that while it is not difficult to generate original works, it is a challenge to create original compositions that successfully pass through the selection implied by the composers' creative process and of the (unknown) process by which compositions enter a compendium of famous works. In other words, maximizing originality is not a sufficient criterion to become influential or successful, but instead one's output needs to conform to norms and standards of a given time.

Our work relates to scholarship on innovation (see Hall and Rosenberg, 2010), particularly to two studies that measure the influence of technological innovation (Kelly et al., 2021; Bussy and Geiecke, 2021), and the increasingly often studied concept of creativity. In a model of the creative process, Feinstein (2011) describes how creators explore and gather elements before finding ways to combine and reconfigure these elements into new original forms. Akcigit et al. (2018) model and analyze sources of the creator's knowledge, which comes from interaction with other people or from external sources related to own explorations. Borowiecki (2022) explores whether teachers in creative fields leave an imprint on their students that shapes the style of their future work, using parts of the dataset presented in the underlying paper. In recent years, economists have also begun to study creativity empirically. For example, Graddy and Lieberman (2018) explore how artists' creativity is affected during bereavement, while Borowiecki (2017) shows the overall influence of psychological well-being on creative production.

Related to creativity is the concept of style, which has traditionally been restricted to the arts (Munro, 1946). However, in recent years, style has become increasingly relevant in several other contexts. Business scholarship includes studies of leadership style (Rotemberg and Saloner, 1993; Bertrand and Schoar, 2003), style in product design (Chan et al., 2018), and cognitive style in teams (Aggarwal and Woolley, 2019). The communication style may denote voting outcomes of Federal Reserve committee members (Malmendier et al., 2021) and constitutes a particular feature of populism (De Vreese et al., 2018). Style characterizes the behavior of children (Caspi et al., 2003), parents (Kim et al., 2015), and financial investors (e.g., Keim & Madhavan, 1997).

This work also relates to research on innovation in history. The supportive intellectual environment in Europe since the 1500s as well as its political fragmentation explain the stark rise in innovations in Europe, which triggered the Industrial Revolution and economic progress (Mokyr, 2016). Freedom to pursue one's own paths of inquiry results in increased originality of people (Feinstein, 2006; Aghion et al., 2008; Simonton, 2004). Relevant is also the important literature on the upper-tail human capital and economic growth (Mokyr, 2009; Meisenzahl and Mokyr, 2011;

Squicciarini and Voigtländer, 2015)).<sup>1</sup> The modern economic growth in Britain and abroad was possible due to an increased desire of innovators to influence others as they spread their improving mentality (Howes, 2017). Creative activity has also been mapped and measured over time and place. For example, Murray (2003) selects leading innovators in the arts and sciences from 800 to 1950, whereas Ger-gaud et al. (2017) document the geographic spread of innovative individuals and cities. In cities where many similar individuals cluster, visual artists (Hellmanzik, 2010) and music composers (Borowiecki, 2013, 2015) are more productive.<sup>2</sup>

This paper provides two main contributions. First, it introduces a new model that enables not only a formalization of style but also the measurement of the degree of originality as well as influence. The framework is versatile and can be applied in many contexts, including to individuals, products, or time periods, which makes it relevant for a large number of unanswered important research questions. Second, it provides an analysis based on this framework that shows originality to be positively associated with influence and several success measures among the composers in our sample. We conclude that originality, rather than being traded-off against influence and success, is a key driver of both.

The paper proceeds as follows. Section 2 introduces a model of style and proposes how to measure originality and influence. Section 3 presents the data sources and descriptives. Section 4 shows the results. Conclusions are in Sect. 5.

## 2 A model of style, originality, and influence

The aim of this section is to introduce credible measures of originality and influence. In order to do so, we first outline a model of style. Style is the way one typically does something. For example, style can be a particular manner or technique by which something is created, written, or performed. It is a distinctive characteristic of a person, group of people, place, or period.

Style plays a particular role in the cultural and creative sectors, because it enables the grouping of creative output or creatives into categories. The observation of style can thus be a useful tool in the study of the arts and music, advertising, architecture, design, film, publishing, video games, and many other creative domains. While any creative output is to some extent unique (otherwise it would not be creative), the manner in which it has been created can be categorized within a certain style.

In this section, we first outline our model of style, before turning to how we conceive of originality and influence on the basis of our model of style.

---

<sup>1</sup> If one accepts that originality is the ability to generate an idea that is unique, this paper also becomes relevant to the vast and important literature on economic growth, which defines growth as a function of the generation and transmission of new ideas (Lucas 2009; Jones 2005).

<sup>2</sup> The structure of music has been also studied in various interdisciplinary settings. Authors have sought to trace the evolution of structural regularities (e.g., Serrà et al., 2012), explore how music changes over time (e.g., Foster et al., 2014), how its structure relates to popular taste (e.g., Mauch et al., 2015) or explore similarities across individual composers (citealmsmith2014,smith2015). In psychology, work by Dean K. Simonton explores whether musical structure can reveal the psychology of musical aesthetics and creativity (Simonton 1980, 1984).

## 2.1 Modeling style

We propose to capture the style of creative output as a function of its *features* or key attributes. For example, in the visual arts such features may include the painting technique, materials used, or motifs painted. In the music it may be specific sequences of music notes, the key in which a piece is written, or more aggregated measures such as rhythm or mood.

In this framework, a style is linked to the frequency of these features. To make this empirically workable, the concept of style must be linked to creative output for which the frequency of features can be measured. We conceptualize style as a probability distribution over features and assume that the frequency of these features is the result of drawing from this probability distribution.

Formally, consider creative works by different people. Let  $F$  be the set of features that occur in at least one of the works under consideration, with  $f = |F|$  denoting the number of such features. The *style* of an individual creator  $i$ ,  $s_i$ , is a probability mass distribution over all features in  $F$ . In other words, style,  $s_i$ , is a vector of length,  $f$ , the components of which sum to one.

Let  $W_i$  be the set of works by individual  $i$ . Then each work,  $w \in W_i$ , is characterized by its associated vector of feature frequencies,  $x_w$ . We assume that  $x_w$  is the result of the composer drawing from  $s_i$ . The frequency distribution of features over all of individual  $i$ 's works,  $x_i = \sum_{w \in W_i} x_w$ , is thus a noisy proxy for style,  $s_i$ .

## 2.2 Modeling originality

We conceptualize *originality* in terms of deviation from past creative output. The style of an individual  $i$ ,  $s_i$ , is more original the more it deviates from *past style*, whereas a perfectly unoriginal individual would have an identical style to past style and the degree of originality would be zero.

In terms of the framework outlined at the start of Sect. 2, we first define a creator's *innovation* in terms of her own style and past style.

**Definition (Innovation):** Let  $s_i$  be individual  $i$ 's style and  $s_p$  be the past style. Then  $i$ 's innovation is defined as:

$$v_i = \tilde{s}_i - \tilde{s}_p \tilde{s}_p' \tilde{s}_i \quad (1)$$

where  $\tilde{s}_i = \frac{s_i}{\|s_i\|}$  and  $\tilde{s}_p = \frac{s_p}{\|s_p\|}$  are the normalized versions of  $s_i$  and  $s_p$ ; that is, they are rescaled so as to have length one.

Like an individual's style, her innovation is an  $f$ -dimensional vector. But unlike style, it is not a probability distribution. Instead, it is the difference between an individual's (normalized) style,  $\tilde{s}_i$ , and that style's projection onto (normalized) past style,  $\tilde{s}_p$ . The innovation is orthogonal to past style by construction. It gives the direction of the individual's deviation from the past.

In order to measure the *degree* of originality, which we refer to simply as originality, the innovation needs to be translated into a scalar measure. We do this by taking the length of the innovation vector,  $v_i$ .

**Definition (Originality):** Individual  $i$ 's *originality*  $\omega_i$  is defined as the length of her innovation:

$$\omega_i = \|v_i\| \quad (2)$$

Originality will therefore be greater, the more  $s_i$  deviates from  $s_p$ . Since all entries of each style vector are non-negative, originality will equal zero if and only if  $s_i = s_p$ . This makes originality a measure of dissimilarity between  $s_i$  and  $s_p$ .<sup>3</sup> At most, originality may equal one if  $s_i \perp s_p$ , a condition that may occur only if the sets of features with positive entries in  $s_i$  and  $s_p$  are disjoint.

Figure 1 illustrates the definitions of both innovation and originality. The normalized styles,  $\tilde{s}_i$  and  $\tilde{s}_p$ , are vectors of length one. The innovation,  $v_i$ , is the difference between  $\tilde{s}_i$  and its projection onto  $\tilde{s}_p$ . Its direction is given by the feature space in which individual  $i$ 's style deviates from the past. That is, it shows which features are either more or less common in individual  $i$ 's style than they are in past style and to what extent they are more or less common. If these differences are large, then  $v_i$  will be longer, which is why its length defines our measure of originality,  $\omega_i$ .

The figure also illustrates another view of our originality concept, namely in terms of a *decomposition* of style,  $s_i$ , into two components: One in the direction of past style,  $s_p$ , and another which orthogonally departs from the past, the innovation. The relative lengths of the two components indicate to what extent individual  $i$  follows the past or innovates.

### 2.3 Modeling influence

While originality measures the extent to which an individual's style departs from the past, influence relates an individual's style to the future. The individual is influential if she "bends" future work in the direction of her style.

The definition of influence relies on three concepts introduced above: Past style,  $s_p$ , future style,  $s_f$ , and individual  $i$ 's innovation,  $v_i$ . Using these, we measure the extent to which future style follows individual  $i$ 's innovation.

To this end, we first project  $s_f$  onto the span of past style and innovation (which is equal to the span of past style and the individual's style). The span of  $s_p$  and  $v_i$  is described by the matrix

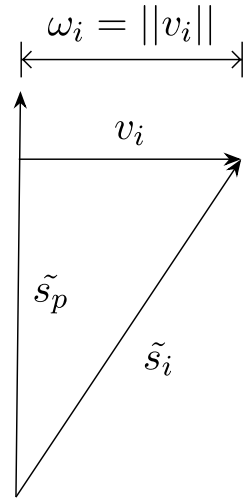
$$A = \left( \tilde{s}_p, \frac{v_i}{\|v_i\|} \right), \quad (3)$$

in which the two vectors are already normalized. The projection of  $\tilde{s}_f$ , the normalized vector describing future style onto the plane defined by  $A$  is then given by

$$p_f = AA' \tilde{s}_f. \quad (4)$$

<sup>3</sup> In addition, the measure is symmetric and the triangle inequality holds. Originality is therefore a metric in the mathematical sense.

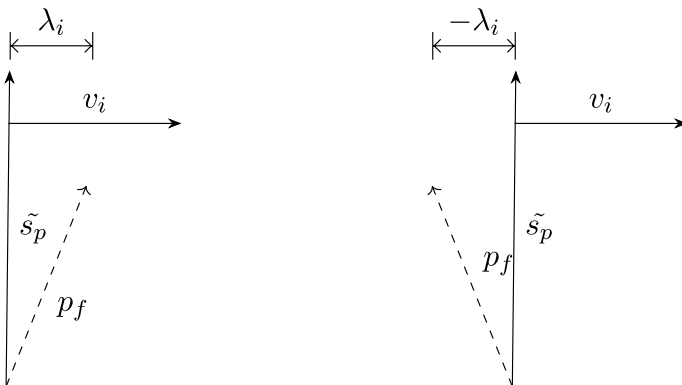
**Fig. 1** A diagram of innovation and originality. *Notes:* Originality,  $\omega_i$ , is the length of individual  $i$ 's innovation,  $v_i$



This projection can be understood as the “shadow” of  $\tilde{s}_f$  on the plane which contains both  $s_p$  and  $v_i$ . Having thus brought all three objects into the same plane, what remains to be done is to evaluate the deviation of  $p_f$  from  $\tilde{s}_f$  in terms of innovation  $v_i$ .

Figure 2 illustrates this. The vectors,  $\tilde{s}_p$  and  $v_i$ , are drawn as in Fig. 1. The projection,  $p_f$ , of  $\tilde{s}_f$  is dashed. It is in general shorter than one, since  $s_f$  will not perfectly align with the plane defined by  $A$  except by coincidence. In general therefore, there is a component of  $\tilde{s}_f$  that points away from the plane, following neither the past nor individual  $i$ 's innovation.

In the left panel of Fig. 2,  $p_f$  deviates from  $\tilde{s}_p$  in the same direction as  $v_i$ . In a sense,  $v_i$  points to the future. In this case, the creator’s influence is positive. In the



**Fig. 2** A diagram of positive and negative influence. *Notes:* Influence,  $\lambda_i$ , is the (directed) distance between the projection of future style,  $p_f$ , and the vertical line, the direction of which is given by past style

right panel, by contrast,  $p_f$  deviates from  $\tilde{s}_p$  in the opposite direction. Here  $v_i$  points away from the future. In this case, the creator's influence is negative.

To formalize this measure, we first project again, this time  $p_f$  on  $\tilde{s}_p$ , and obtain individual  $i$ 's contribution to future style:

$$u_i = p_f - \tilde{s}_p \tilde{s}_p' p_f \quad (5)$$

This contribution to future style is analogous to the innovation defined in Sect. 2.2. It too can be viewed as part of an orthogonal decomposition, here of  $\tilde{s}_f$ , where the contribution to future style represents the component that runs parallel to innovation,  $v_i$ .

But, as can be seen in the right panel of Fig. 2,  $u_i$  does not need to run in the same direction as  $v_i$  and the distinction is crucial. For this reason, we define influence as a directed length of  $u_i$ , where the direction is given by  $v_i$ .

**Definition (Influence):** Let  $u_i$  be individual  $i$ 's contribution to future style and  $v_i$  her innovation. Then individual  $i$ 's influence is defined as the directed length of  $u_i$ :

$$\lambda_i = ||u_i|| \frac{u_i' v_i}{|u_i' v_i|} \quad (6)$$

As with  $\omega_i$ , individual  $i$ 's originality, influence cannot exceed one. This extreme case occurs when future style is equal in direction to the creator's innovation,  $v_i$ ; that is, if the novel part of individual  $i$ 's style "becomes" the future.<sup>4</sup> At the other extreme, the influence of an individual  $i$  can theoretically approach minus one, though it cannot be exactly equal to minus one. Extreme cases occur when innovation is very small and future style is nearly orthogonal to past style.<sup>5</sup>

Another instructive aspect of the behavior of influence as defined above concerns the length of  $p_f$ . As can be seen intuitively in Fig. 2, if  $p_f$  were to shorten, while retaining its direction, this would move  $\lambda_i$  closer to zero. The length of  $p_f$  varies with the extent to which future style,  $s_f$ , is aligned with the plane defined by  $A$ : If the future is very different both from the past and from individual  $i$ 's style,  $p_f$  will be short and influence will tend to be small in absolute value. In the extreme, if the set of features used in the future is disjoint from that used in the past and by individual  $i$ , then influence is necessarily zero.

## 2.4 Working with frequencies

So far, we have defined a creator's innovation, originality, and influence in terms of style. Yet in reality style, defined as a probability distribution over features, is not known. Instead, any style will have to be substituted by its frequency equivalent.

<sup>4</sup> Such case can only occur if the sets of features used with positive probability in past and future style are disjoint. This is because styles cannot have negative entries and innovation is by definition orthogonal to past style.

<sup>5</sup> For instance, suppose half of the features in  $F$  have probability  $\epsilon$  in past style, individual  $i$ 's style differs from the past only in that it assigns probability zero to these features, and future style assigns positive probability only to these features. Then  $\lim_{\epsilon \rightarrow 0} \lambda_i = -1$ .



The frequency equivalent of a style is the *empirical style*, a frequency distribution over features drawn from an observed corpus of creative work. For instance, while individual  $i$ 's own style,  $s_i$ , cannot be observed, her empirical style,  $x_i$ , can. It is the frequency distribution in individual  $i$ 's collected works over all features in  $F$ . Like  $s_i$ ,  $x_i$  is thus an  $f$ -dimensional vector whose entries sum to one. Empirical past style,  $x_p$ , and empirical future style,  $x_f$ , are defined analogously.

To avoid cumbersome references to empirical style, we will refer to it as “style” in the empirical section of the document, but denote it  $x_i$ ,  $x_p$ , and  $x_f$  for clarity. Our baseline measures of composer  $i$ 's originality,  $\omega_i$ , and composer  $i$ 's influence,  $\lambda_i$ , are calculated on the basis of empirical style, using all features used by composer  $i$  and the full corpora of past and future themes.

In our empirical application, we observe many composers for whom only very few themes are available. The resulting sparsity of their empirical style  $x_i$  induces a correlation between the number of themes available for a composer and their originality estimate. To address this issue, we correct our baseline originality measure for sparsity. Appendix C describes this correction.

In addition to the baseline measures, we compute alternatives in which different themes are taken into consideration for  $x_i$ ,  $x_p$ , and  $x_f$ . Among these alternatives are measures computed at the level of the individual theme, such that each theme is considered to have its own style. In addition, we also compute an influence measure at the composer level, but restricting the corpus considered for future style.

#### 2.4.1 Theme-level originality and influence

The theme-level approach yields estimates of originality,  $\omega_{it}$ , and influence,  $\lambda_{it}$ , for each theme,  $t$ . To aggregate these values at the composer level, we compute simple averages over all themes of a given composer. While this may seem redundant, the resulting composer-level average originality,  $\bar{\omega}_i$ , has a key advantage. It does not suffer from the sparsity problem described above. This is because the themes tend to be of similar length. No correction is therefore applied for theme-level originality.

A similar theme-level measure is computed for those themes for which a composition year (or interval) is available. For each of these 3,038 themes, the corpus of past themes is assembled from all themes which were either composed before the start of a given theme's time interval, or which were composed by a composer who died before the start of the interval. The future corpus is assembled analogously, including those themes composed after the end of a given theme's time interval or by composers who were less than ten years of age at the end of the interval. This measure is only available for a reduced sample but has the advantage that past and future corpora contain works closer to the date of composition of a given theme.

#### 2.4.2 Near-term influence

To account for the possibility that a composer's influence may change over time, we create a near-term influence measure to be compared with overall influence. A

composer might be highly influential on the next generation of composers but then be mostly forgotten, or ignored by the next generation and rediscovered much later. The near-term influence of composer  $i$  measures her influence on the first generation of composers who were born less than 30 years after composer  $i$ 's death and were not yet ten years of age at composer  $i$ 's death.

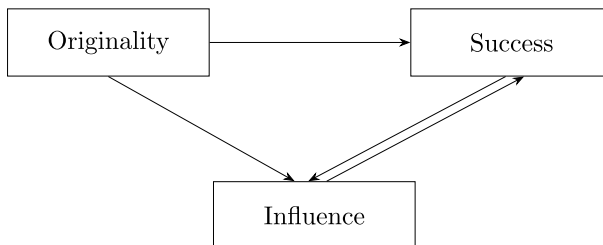
## 2.5 Modeling relationships

Originality, influence, and success are related in various ways. Before investigating these relationships empirically, it is useful to discuss them conceptually.

Figure 3 represents our thinking diagrammatically. By its construction, originality precedes the others in time. It is constructed on the basis of a creator's own style and of past style, meaning that all information on which originality is based was available at the end of the creator's productive life. The other two measures are only realized later, since influence depends on future style and success is measured in the present day. For this reason, we are mainly interested in the consequences of being original and treat it as exogenous to influence and long-term success.

More original creators may be more influential by being more inspirational to their future peers. At the same time, they may not be to the taste of their audience or critics, making them less well known to both future creators and future audiences. The relationship between originality and influence or success is thus *ex-ante* unclear and remains to be explored empirically in Sect. 4.

The relationship between influence and measures of success, however, is hard to interpret. By baseline measures the two are positively correlated, though not strongly (20%). Various causal explanations are plausible. Both could be affected by common factors, such as originality as just discussed; success may bring a creator to the attention of others, driving influence; or a creator's influence on others may become public knowledge, increasing her fame and thus her success. Of course, negative causal effects are also possible. Creators may avoid drawing on successful past creators, so as not to be seen as pedestrian.



**Fig. 3** Relationship between originality, influence, and success

### 3 Data and descriptives

#### 3.1 Features of musical themes

We collect data on features of a large number of musical themes from two volumes of musical dictionaries by Barlow and Morgenstern (1975, 1976). The authors declare that they cover “all the themes the average and even the more erudite listener might want to look up” and include 18,074 musical themes from 6352 classical and operatic works written by 769 composers.

These dictionaries have been compiled for readers to check details on a musical theme (e.g., the title, who composed it, etc.) based solely on the melodic tune. Therefore, for each theme the dictionaries provide a sequence of musical notes intended to enable the reader to uniquely identify each theme. We use this information to obtain *n-grams*—sub-sequences of *n* notes (duplets, triplets, quadruplets). Conveniently, the sequences are transcribed into a notation index, which is transposed to the key of C (either major or minor) and reports each note as a letter with a trailing symbol to indicate a sharp (half-step raise) or a flat (a half-step lowering).<sup>6</sup>

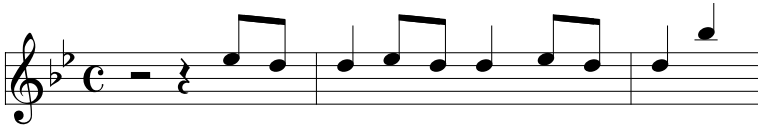
For example, the theme of Symphony No. 40 in G Minor by Wolfgang Amadeus Mozart is included in Barlow and Morgenstern (1975) as notes on a staff (Fig. 4). In addition, the first six notes of the theme are included in the notation index after being transposed to a common key: C minor as the original theme is in a minor key. The first six notes of the original theme in G minor  $Eb - D - D - Eb - D - D$  become, after being transposed,  $Ab - G - G - Ab - G - G$ .

Along with the sequence of notes, we collect time signature and key signature for each theme. A time signature, which can be indicative of rhythm, specifies how many beats are contained in each measure (segment of time) and which note value is equivalent to one beat. In Symphony No. 40, the C symbol at the beginning of the music staff means “common time” and is another way of writing the 4/4 time. A key signature indicates notes that are to be played higher or lower than the corresponding natural notes. In our example, the very title of the work tells us the key, and other hints come from the key signature, two flats reported at the beginning of the staff, and the harmonic structure of the opening.

Yet the key could not be successfully recorded for all themes. While the number of flat and sharp notes reported at the beginning of the staff can normally be observed for all themes, two possible keys—one major and one minor—are associated with the same set of sharp or flat symbols. For example, the key signature for G minor has two flats and its relative major is B-flat major.

To identify the key we perform a series of tests that look at the presence of certain notes in the musical theme (see Appendix A for details on how we identify the key signature). Themes for which the key identification process led to ambiguous results were dropped from the main dataset. Our baseline focus on 2-grams (sets of two consecutive notes), time signature, and key signature, enables us to cover

<sup>6</sup> While music normally gets transposed to best fit the pitch of a particular musical instrument, here transposing to a common key is necessary to enable comparisons across melodies.



**Fig. 4** Theme of the Symphony No. 40 in G Minor by Wolfgang Amadeus Mozart

10,890 musical themes and 684 composers. The total number of unique features extracted in this way is 193.<sup>7</sup> Including the key and time signature then, the final list of musical features for Mozart's Symphony No. 40 is [*AbG*; *GG*; *GAb*; *AbG*; *GG*; *4/4*; *G minor*].

When applying the model presented in Sect. 2 to this data, we can observe that the two vectors of style of the past,  $x_p$ , and future,  $x_f$ , are generally similar (i.e., the angle between them, at around  $30^\circ$ , is small). This is due to considerable continuity in this art form over the centuries. The two vectors also typically have few zeros, since any feature that occurs at all is likely to be seen at least once in both the past and the future.

Information on the years of compositions is obtained from *Petrucci Music Library* (2020). Unfortunately, the years of composition are only available for a small subset of 3038 themes, typically as an interval of several years, during which the associated work was composed.<sup>8</sup> For this reason, we do not rely on this data in our main analysis, but we provide results for a theme-level approach based on composition years. In the main analysis we rely on the long time span of the data, defining past and future corpora using the birth and death years of our composers. If a theme was composed by someone who died before a given composer turned ten, we consider it to be in the composer's past.<sup>9</sup> On the other hand, if a theme was composed by someone who was not yet ten when a given composer died, we consider it to be in the composer's future.

### 3.2 Success

To capture a composer's success (or eminence as one of our sources insists on calling it) we collect data from three different sources. We are thus able to cover popular success as well as success designated by experts.

First, we rely on modern consumption data from the music streaming service Spotify. We collect the number of followers of each composer and a popularity

<sup>7</sup> Two alternative datasets are created, each covering 16,826 musical themes and 757 composers: The first does not consider the key signature as a feature; the second retains only the observable information on the set of sharp and flat notes as features without making distinctions between a major key and its minor relative. Results from these alternative datasets are consistent with the baseline findings and are available upon request.

<sup>8</sup> Composition periods are available for an additional 1483 themes that do not have an associated key signature, which is used as a feature in our analysis.

<sup>9</sup> In Barlow and Morgenstern (1975, 1976), only Wolfgang Amadeus Mozart had composed before his tenth birthday.

index created directly by Spotify. The latter ranges from zero to 100. Both Spotify measures are retrieved for a period of 12 consecutive months and then an average of the scores is retained. This should limit concerns about seasonality of music or short-term shocks affecting the measures. As the two measures are highly correlated, we will, in the following sections, employ the number of followers as our baseline measure for success, but the results would remain qualitatively the same if any of the other success measure was used instead.

Second, we obtain the Murray quality index from “Human Accomplishment” (Murray 2003), which scores individuals from many disciplines who lived between 800 BCE and 1950. For Western music, the author combines 16 different international sources including encyclopedias, dictionaries, and surveys and aggregates them providing a list of the 522 most significant composers along with a weighted index that ranks these composers on a scale from one to 100. We assign a Murray index of zero to composers not listed by Murray but covered by Barlow and Morgenstern (1975, 1976). In this historiometrical effort the author argues that it is excellence in accomplishments that underlies the measures of eminence compiled in the overview, rather than mere fame.

Finally, we count the number of words contained in the main text of each composer biography published in *Grove Music Online* (2018)—a leading encyclopedic resource on music and musicians. The biographies are written accounts of the lives of composers including life events and their creative outputs.

### 3.3 Biographical information

We obtain detailed biographical information for all composers for whom we have information on musical themes and on measures of success. We collect individual-level data from the Grove Dictionary of Music and Musicians (now *Grove Music Online* 2018), including information on the years of birth and death as well as place of birth and nationality.<sup>10</sup>

For composers for whom information on the birth or death years or nationality are missing we used data from Pfitzinger (2017) and from Barlow and Morgenstern (1975, 1976), who occasionally report years of birth and death of the composers, and, if needed, other reference and online resources. For very few remaining composers—only seven—for whom either birth or death years were still not available, we estimated these dates using the information available (e.g. birth, death, or “*floruit*”, the period during which the artist was reported active) as well as information from the lives of their contemporaries.

We have also obtained for each composer data from the biography in *Grove Music Online* (2018) on her teachers and students, including information on which music conservatories she attended and when, as well as traveling data, which includes the list and number of cities visited. These data are used in additional analyses and robustness tests in the Appendix E.1.

<sup>10</sup> The latter is mapped in Fig. 17, Appendix D.

### 3.4 Descriptive statistics

Table 1 reports the century and country of birth for the 684 composers included in our main dataset. Though some of these composers were born as early as the 15th century, 62% were born in the 19th century. Most of our sample covers European composers, while a considerable share of composers in the 19th century is North American.<sup>11</sup>

Our baseline measures for influence, originality, and success are summarized in Table 2. Panel (A) summarizes originality measures (see Sect. 2.4 for descriptions). The first measure is computed for each composer, the second for each theme. Both measures are standardized and range from a minimum of zero to a maximum of 100. While the two variables cannot be directly compared, once theme-level originality is collapsed to a composer-level mean, the correlation between the two measures is positive and higher than 50%.

Influence measures are shown in Panel (B) of Table 2 (see Sect. 2.4 for descriptions). The measures are standardized while preserving the meaning of the zeros. The process results in influence variables having different negative boundaries and a maximum of 100. Influence measures are available for only 432 composers. This is because influence can only be evaluated when a corpus of future compositions is available, which is not true of composers who live at the end of the sample period. Influence highly correlates with both near-term influence and theme-level influence (> 85%).

Success is shown in Panel (C) of Table 2. All success measures are taken in their logarithmic form and, despite measuring different aspects of the popularity of a composer, are all highly correlated (around 70%). Before taking the logarithm, we set the index to zero for all composers in our dataset who are not explicitly mentioned by Murray and add one to all values.

Histograms of originality, corrected originality, and theme-level originality as well as the three influence measures are available in Appendix C (Figs. 14 and 16).

Table 3 reports summary statistics for originality, influence, and success by historical-musical period into which Western music is traditionally divided. Average values of influence decrease over time, while average originality remains fairly stable, with the Medieval, Baroque, and Romantic periods being characterized by the highest degree of originality. Success is on average much higher for composers of the Romantic period, which is also the one for which we have the highest number of composers in our dataset.

The changes in originality and influence over time are shown in Fig. 5. Instead of simply aggregating the originality and influence measures by year we create a “rolling window” indicator, which measures the originality and influence of the work of all composers who were at least ten years of age in a specific year—the contemporaneous composers. Their combined corpus is compared against past work by composers who died before the youngest of the contemporaneous composers turned ten and against the combined corpus of future work by composers who were at most ten years old

<sup>11</sup> Additional maps and descriptive statistics are available in Appendix B.

**Table 1** Descriptive statistics for composers by century and region of birth

Century	1200	1300	1400	1500	1600	1700	1800	1900	Total
Europe	1	5	12	63	64	90	339	9	583
N. America	0	0	0	0	0	2	75	10	87
Other	0	0	0	0	0	0	11	1	12
Avg. age	38.0	69.0	62.6	59.7	65.9	64.5	68.7	75.3	67.0
Avg. themes	1.0	2.4	3.0	3.5	24.0	27.7	14.9	9.70	15.9

The table reports the number of composers available in our dataset by geographical area and century as well as the average age and the average number of themes of a composer. Europe includes composers from Russia. The “Other” category includes mostly composers from South America. Two composers for whom the country of origin is unknown are dropped from this table. The data was collected by the authors (see Sect. 3 for details)

**Table 2** Descriptive statistics for originality, influence, and success

Variable	Mean	SD	Min.	Max.	N
Panel A: Originality					
Originality	36.67	15.97	0	100	682
Theme-level originality	74.02	15.55	0	100	10,883
Panel B: Influence					
Influence	18.15	23.08	− 38.74	100	431
Near-term influence	13.93	18.59	− 28.10	100	431
Theme-level influence	8.46	17.70	− 68.38	100	8606
Panel C: Success					
Ln Spotify followers	4.98	3.06	− 2	14	606
Ln Murray index	0.69	1.04	0	5	684
Ln Word count	7.18	1.31	4	11	583

The table reports in Panel A summary statistics for corrected originality,  $\omega_i^c$ , and originality at the theme level,  $\omega_{it}^c$ ; both measures are standardised. Panel B reports influence,  $\lambda_i$ , near-term influence and theme-level influence. Panel C reports success measures: the log number of Spotify followers, the log of the Murray index and the log of the word count. The data was collected by the authors (see Sect. 3 for details)

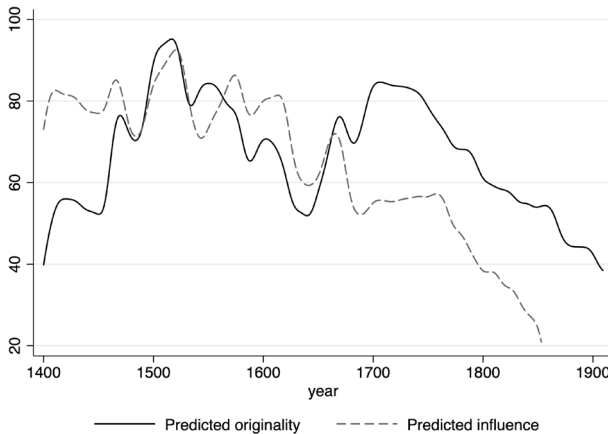
when the oldest of the contemporaneous composers died. The beginning and end of the sample period have to be interpreted with caution: The number of past themes is limited for composers near the beginning of our sample. Similarly, the number of future themes is lower for composers born near the end of the sample. This could limit the accuracy of our measures at the beginning and end of our observation window.

Focusing on the more central parts of the sample period, we observe two peaks in originality—around 1500 and at the turn of the 1700s, the Baroque period. The first peak is followed by almost two centuries of relatively low originality. In this period music by others was often appropriated for re-use, and such borrowing would not raise any concerns of plagiarism. The second peak—around the mid-18th

**Table 3** Descriptive statistics for composers by musical period

Period	Year	Originality	Influence	Success	Composers
Medieval	1150–1390	45.23	72.01	5.28	5
Renaissance	1391–1559	38.67	42.01	4.60	43
Baroque	1560–1710	40.27	28.72	5.57	106
Classical	1711–1762	31.73	23.08	5.03	47
Romantic	1763–1861	36.69	8.30	5.02	234
Modern and postmodern	1862–1946	35.60	0.90	4.75	249

The table reports the composer average for the influence measure,  $\lambda_i$ , the corrected originality measure,  $\omega_i^c$ , and the success measure (log number of Spotify followers), as well as the total number of composers available in our dataset by musical period. The categorisation of a composer into a musical period is based on the birth year. The exception is Leos Janacek (born 1854), who is considered here a modern composer, in line with scholarship. The data was collected by the authors (see Sect. 3 for details)



**Fig. 5** Originality and influence in Western music over time. *Notes:* The plot presents originality and influence of the work of all composers who were at least ten years of age in a specific year. The data was collected by the authors (see Sect. 3 for details)

century—comes at a time when originality had become increasingly praised and eventually became regarded as superior to imitation (*Grove Music Online* 2018).<sup>12</sup> As a result, the quality of a composer was assessed by the inventiveness of new music as opposed to a skillful manipulation of existing material.

Figure 5 also shows that the peaks in originality are not closely followed by peaks in the influence measure. Moreover, both measures clearly decrease during the last roughly two centuries, which is possibly a reflection of the “low-hanging fruit” phenomenon. A more formal explanation is provided by the humanities’ theory of anxiety of influence. The body of musical works increases with time, which instills in

<sup>12</sup> It was also around that time when Adam Smith wrote that “the disparity between the imitating and the imitated object is the foundation of the beauty of imitation. It is because the one object does not naturally resemble the other, that we are so much pleased with it, when by art it is made to do so” in *Essays on The Imitative Art* (Smith 1980).



young composers a form of anxiety as they struggle against past musical predecessors to create something original and to achieve success. The correlation between originality and influence is equal to 0.31.

## 4 Results

In this section we first present the results of our baseline approach, which involves estimating originality and influence at the composer-level and considers all work before a given composer turned ten years of age to be the past and all work after her death to be the future. We then present the results based on the variants of the approach discussed in Sect. 2.4 and investigate the relationship between a composer's age and her originality and influence. Finally, we assess the relationship between near-term influence and overall influence.

### 4.1 Originality, influence, and success

This section explores how the baseline measure of originality is related to, on the one hand, the success measures and, on the other, the baseline measure of influence.<sup>13</sup> We first explore these relationships graphically. The graphs in Fig. 6 plot baseline originality against our different measures of success: The number of Spotify followers, the Murray index, and word count measures.<sup>14</sup> In all cases, the graphs suggest a positive relationship between originality and each of the success measures.

Panel (A) of Fig. 7 shows non-parametric regressions of the success measure—the (log) number of Spotify followers—on baseline originality. Panel (B) shows an analogous regression of baseline influence measure on baseline originality. The graphs suggest that both success and influence are approximately linearly related to originality. Interestingly, rising originality increases influence on average, but the effect is more volatile.

Next, we explore these relationships using regression analysis. In Table 4, the success measures are regressed on baseline originality. Estimates report results for a Seemingly Unrelated Regression (SUR), where the error terms are allowed to correlate across the three columns of each set. The results show that originality is positively associated with each of the three success measures.<sup>15</sup> These relationships are barely affected by the inclusion of controls for the number of past and future themes or nationality fixed effects (columns 4–6).<sup>16</sup> To quantify the effect, looking

<sup>13</sup> The baseline measure of originality is what we term *corrected* originality in Appendix C and denote as  $\omega_i^c$ .

<sup>14</sup> As discussed in Sect. 3.2, we take the logarithm of all these measures to obtain less skewed distributions. For the case of the Murray index, we replace zeros by ones before taking the logarithm.

<sup>15</sup> It is encouraging to observe that for both expert-based measures (Murray index and word count) very comparable coefficients are estimated.

<sup>16</sup> The number of themes of the composer is not included here for a simple reason: It is closely related to success in its own right. The more well known a composer, the greater the number of themes that we find in the data. To address potential concerns that our corrected originality measure may itself still be affected by the number of themes of a given composer, we present theme-level results in Sect. 4.2.

**Fig. 6** Scatterplots of originality and different success measures. *Notes:* Scatter of baseline originality and different success measures: The log number of Spotify followers (correlation of 0.56), the log of the Murray index (correlation of 0.54), and the log of the word count (correlation of 0.48). Dots are labelled for top-5 composers based on the success measures

at column (4) in Table 4, we can read that an increase in originality by 1% is associated with an increase in the number of Spotify followers by close to 10%.<sup>17</sup> To express this differently, for a composer with 145 thousand followers (the log-mean of the distribution), this would imply an additional about 14,000 followers.<sup>18</sup>

Table 5 reports results for an OLS regression in which baseline influence is regressed on baseline originality. Different specifications progressively introduce fixed effects for the half-century of the composer's birth, fixed effects for the nationality of the composer, and controls for the number of each composer's themes as well as the number of themes in each composer's past and future. Reading the results from column (4), Table 5, we note that an increase in originality by 1% is associated with an increase in influence of about 0.3%. An attempt to visualize the relationship between originality, success, and influence, is reported in a three-dimensional heat map in Appendix F.

In summary, originality is shown to be positively associated with success or influence, and remains stable across various specifications. Even when we include half-century and nationality fixed effects, as well as the number of past, future, and composer themes, the coefficients remain large and positive. Next, we turn to theme-level data.

## 4.2 Theme-level data

Exploiting theme-level data has two advantages. First, it allows us to reassess the results previously obtained, that originality is positively associated with success and influence. Second, by including composer fixed effects, it becomes possible to measure the effect of originality on influence over the lifetime of a single creator.

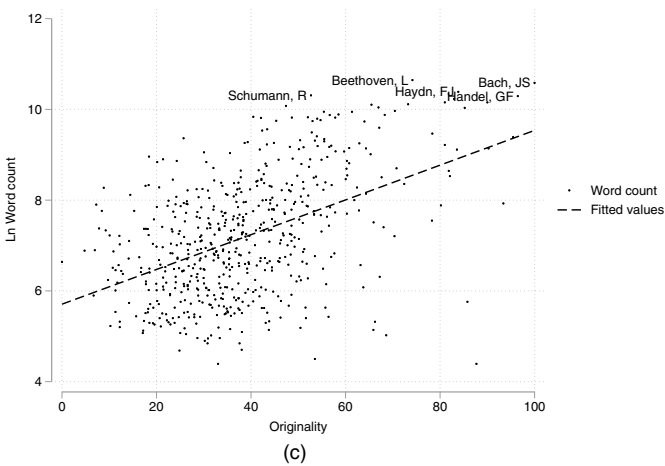
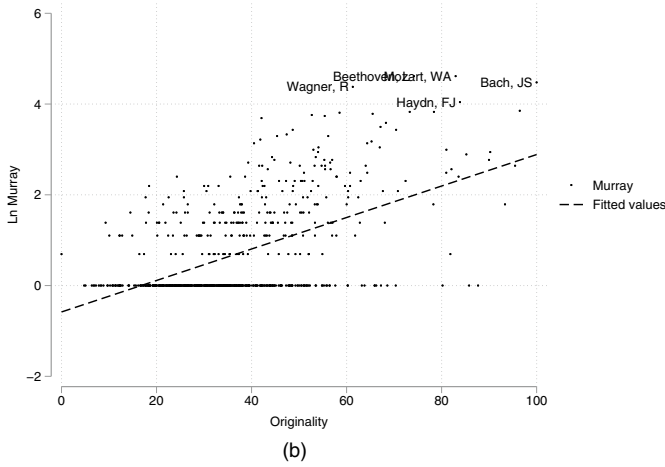
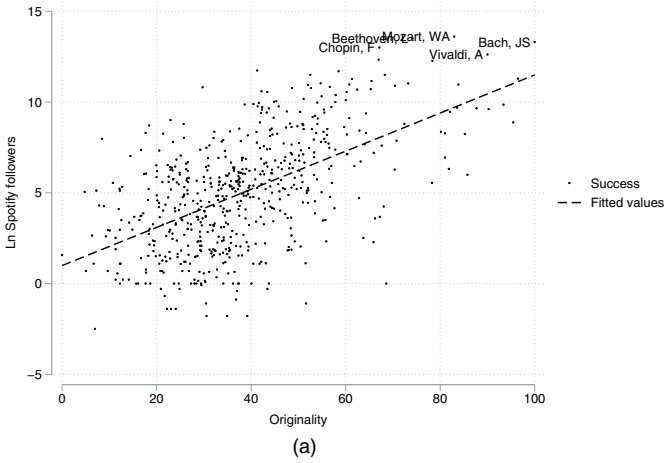
Table 6 reports OLS regressions as before, but where influence and originality are calculated at the theme level and every observation is a unique theme. As the table shows, the effects of originality on both influence and success remain positive and highly significant. For the within-composer effects, shown in columns (2) and (3), this means that the more original themes of a given composer tend to be more influential on future music than less original themes of the same composer.

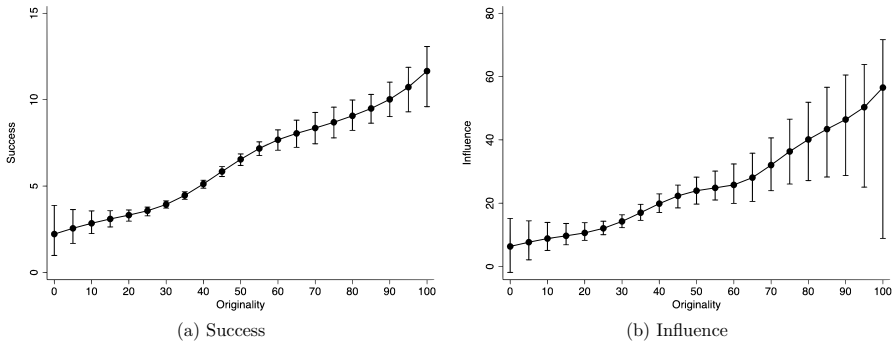
Comparable results emerge in Table 7, which provides a more restricted estimation of the previously shown model. In particular, we consider now only themes for which a date of composition is available.<sup>19</sup> As past and future are defined using the

<sup>17</sup> An increase in the log of the number of followers by 0.092 translates to a percentage increase of  $\exp(0.092) - 1 = 9.6\%$ .

<sup>18</sup> A related question is whether and how the *variation* of originality matters for success. This is discussed and explored in the Appendix E.2.

<sup>19</sup> Not all themes for which the composition date is available are included in this regression as for many of them we do not have the key signature, see Appendix A for details. A small number are also missing from the regression samples because past or future corpora could not be assembled for them.





**Fig. 7** Originality, influence, and success. *Notes:* Panel A shows non-parametric regressions of the success measure—the (log) number of Spotify followers on baseline originality. Panel B shows the same for baseline influence on baseline originality. Both graphs are the result of kernel regressions with a bandwidth of ten and using a Gaussian kernel. Adjusted predictions with 95% confidence intervals

**Table 4** Originality and success

	(1)	(2)	(3)	(4)	(5)	(6)
	Spotify foll	Murray	Word count	Spotify foll	Murray	Word count
Originality	0.0992*** (0.0066)	0.0359*** (0.0024)	0.0388*** (0.0029)	0.0921*** (0.0065)	0.0341*** (0.0024)	0.0366*** (0.0029)
Past themes				0.0004* (0.0002)	0.0000 (0.0001)	−0.0000 (0.0001)
Future themes				0.0003* (0.0002)	0.0001* (0.0001)	0.0002** (0.0001)
Half-century FE	Yes	Yes	Yes	Yes	Yes	Yes
Nationality FE	No	No	No	Yes	Yes	Yes
N	546	546	546	546	546	546
R-Square	0.329	0.338	0.313	0.378	0.370	0.352

The dependent variables are measures of success as follows: the log number of Spotify followers (columns 1 and 4), the log of the Murray index (columns 2 and 5) and the log of the word count (columns 3 and 6). The table reports Seemingly Unrelated Regression (SUR) estimates. Standard errors in parentheses. \*  $p < .1$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ . The data were collected by the authors (see Sect. 3 for details)

composition period of each theme, instead of the life of the composer, compositions of contemporaries are included in the calculation of influence and originality. This may explain results in columns (4) and (5), where originality loses its significant relationship with composer’s success.

It should be noted that themes for which a year of composition was retrieved were more likely to be composed by relatively successful composers. Figure 8 clearly shows that the two success distributions are only partially overlapping, with the distribution of success for composers with known composition years extending further to the right.

**Table 5** Originality and influence

	(1)	(2)	(3)	(4)
	Influence	Influence	Influence	Influence
Originality	0.4592*** (0.0618)	0.4508*** (0.0503)	0.4466*** (0.0523)	0.3458*** (0.0610)
Composer's themes				0.0422*** (0.0159)
Past themes				– 0.0003 (0.0026)
Future themes				0.0028** (0.0012)
Half-century FE	No	Yes	Yes	Yes
Nationality FE	No	No	Yes	Yes
N	431	431	431	431
R-Square	0.114	0.474	0.489	0.510

The dependent variable is a standardized coefficient that measures influence at the composer-level. Standard errors in parentheses. \* $p < .1$ , \*\* $p < .05$ , \*\*\* $p < .01$ . The data was collected by the authors (see Sect. 3 for details)

**Table 6** Theme-level originality and influence

	(1)	(2)	(3)	(4)	(5)
	Influence	Influence	Influence	Success	Success
Originality	0.0734*** (0.0131)	0.0560*** (0.0116)	0.0579*** (0.0123)	0.0128*** (0.0017)	0.0116*** (0.0021)
Theme length	0.7280*** (0.1581)		0.0659 (0.1459)		0.1307*** (0.0252)
Composer FE	No	Yes	Yes	No	No
N	8606	8606	8606	10,737	10,737
R-Square	0.005	0.003	0.003	0.003	0.004

The dependent variables are influence (columns 1–3) and success measured with the log number of Spotify followers (columns 4–5). Both originality and influence are theme-level variables. Standard errors in parentheses. \* $p < .1$ , \*\* $p < .05$ , \*\*\* $p < .01$ . The data was collected by the authors (see Sect. 3 for details)

### 4.3 Composer's age, originality, and influence

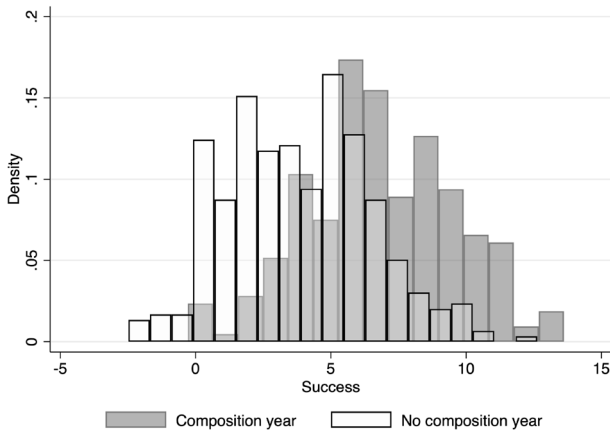
We exploit the available sample of themes for which the composition year was recorded and calculate composer's age at the time of composition.<sup>20</sup> Figure 9 shows the relationship between composer's age when the theme was composed and

<sup>20</sup> When the composition period is longer than one year, we calculate the composer's age at the mean.

**Table 7** Theme-level originality and influence using theme composition period

	(1)	(2)	(3)	(4)	(5)
	Influence	Influence	Influence	Success	Success
Originality	0.0506** (0.0203)	0.0486** (0.0196)	0.0573*** (0.0209)	− 0.0017 (0.0027)	0.0012 (0.0029)
Theme length	0.4779* (0.2827)		0.3404 (0.2882)		0.1093*** (0.0408)
Composer FE	No	Yes	Yes	No	No
N	3028	3028	3028	3029	3029
R-Square	0.002	0.002	0.003	0.000	0.002

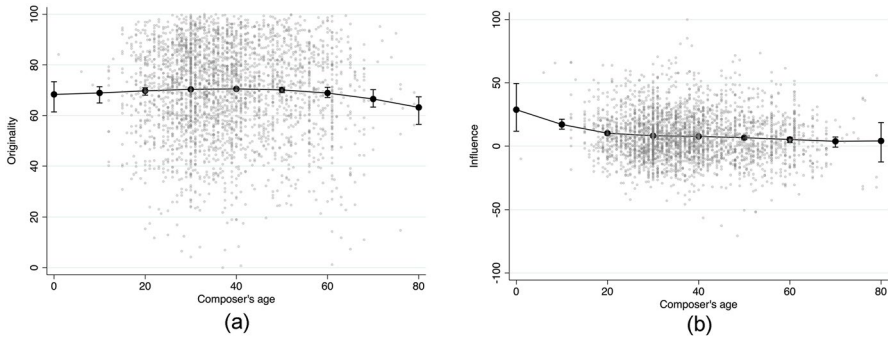
The dependents are influence (columns 1–3) and success measured with the log number of Spotify followers (columns 4–5). Both originality and influence are theme-level variables. Considered are only themes for which a date of composition is available. Standard errors in parentheses. \* $p < .1$ , \*\* $p < .05$ , \*\*\* $p < .01$ . The data was collected by the authors (see Sect. 3 for details)



**Fig. 8** Distribution of success by availability of composition periods. *Notes:* The histograms show the distribution of success for two groups of composers: Composers for whom at least one composition period could be retrieved and the remaining composers. For details on the data see Sect. 3.1

theme-level originality (Panel A) and theme-level influence (Panel B). In both panels, the fitted line is estimated using non-parametric regressions of the composer's age on originality (Panel A) and influence (Panel B). The scatter plots show weak but suggestive patterns whereby originality appears to be smaller for themes composed very early or very late in a composer's life and influence appears to decrease with the composer's age.

To test for these patterns, we perform quadratic regressions for each relationship. The results are shown in Table 8. While the concave relationship between originality and composer's age is not significant with composer fixed effects and controls for the number of past and future themes, the negative relationship between influence and composer's age remains consistent throughout both specifications.



**Fig. 9** Originality and influence over age. *Notes:* Panel **A** shows for each theme the age of composer when the theme was written and theme-level originality. Panel **B** shows the same for the theme-level influence measure. Each observation in the scatter represents a theme for which a composition period was retrieved and theme-level originality and legacy were computed. Both graphs are the result of kernel regressions with a bandwidth of ten using a Gaussian kernel. For details on the data see Sect. 3.1

**Table 8** Originality, influence, and composer’s age

	(1) Originality	(2) Originality	(3) Influence	(4) Influence
Composer’s age	− 0.0076 (0.0272)	0.0039 (0.0400)	− 0.1160*** (0.0283)	− 0.1388*** (0.0415)
Composer’s age <sup>2</sup>	− 0.0050*** (0.0018)	− 0.0024 (0.0021)	0.0017 (0.0018)	0.0000 (0.0022)
Past themes		− 0.0086 (0.0298)		− 0.0094 (0.0308)
Future themes		− 0.0057 (0.0160)		− 0.0053 (0.0165)
Composer FE	No	Yes	No	Yes
N	3038	3038	3028	3028
R-Square	0	0	0	0

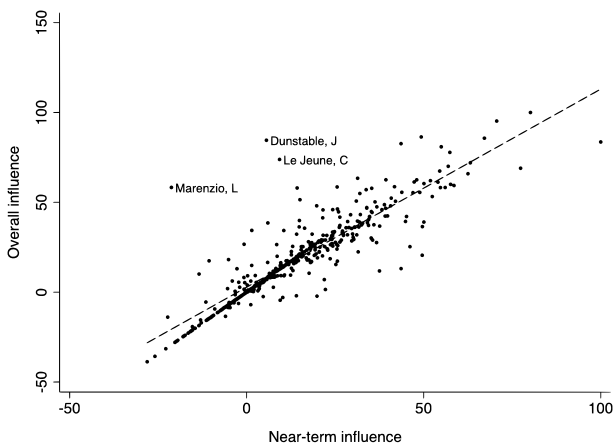
The dependent variables are originality (columns 1–2) and influence (columns 3–4). The variables of interest are the composer’s age and its square. All columns report OLS estimates. Standard errors in parentheses. \* $p < .1$ , \*\* $p < .05$ , \*\*\* $p < .01$ . The data was collected by the authors (see Sect. 3 for details)

#### 4.4 Near-term influence

The relationship between near-term influence (described in Sect. 2.4) and baseline influence (or  $\lambda_i$  in Eq. 6) deserves some attention. While near-term influence reflects the most immediate impact of a given composer (i.e., on the next generation of composers), overall influence should capture the influence over future generations.

For a fraction of composers (30%), the two measures coincide as they were born at the end of the sample. For those for whom the two measures differ, we can observe several cases in which the overall influence is much larger than the near-term one. These composers are overall quite influential, but they seem to be much more appreciated in the long-term than by the next generation of composers. Figure 10 illustrates this relationship. Composers such as Luca Marenzio, Claude Le Jeune, and John Dunstable have an overall influence measure above 50, much higher than their near-term influence. They all appear to be composers whose style was transitional in between two movements. While recounting the style change in Italy from the late Renaissance to the early Baroque, Giustiniani (1961) describes Marenzio's work as follows: "In a short space of time the style of music changed and the compositions of Luca Marenzio appeared with delightful new inventions, either that of singing with several voices or with one voice alone accompanied by some instrument, the excellence of which consisted in a melody new and grateful to the ear, with some easy fugues without extraordinary artifices." Note that Marenzio scores high in our measure of originality,  $\omega_i$ , placing himself among the 25% most original composers.

In the same period, Claude Le Jeune, criticized by his contemporaries for his use of the counterpoint, profoundly influenced some composers of the 20th century. For example, in 1948 Olivier Messiaen composed "Cinq rechants" an homage to Le Jeune's "Printemps" (Dobbins and His, 2019). Dunstable, too, was a transitional composer who "influenced the transition between late medieval and early Renaissance music" (Encyclopaedia Britannica, 2020).



**Fig. 10** Near-term influence and overall influence. *Notes:* The scatter plot shows the relationship between near-term influence and overall influence. Near-term influence of composer  $i$  restricts the observation window of influence to the first generation of composers after the death of composer  $i$ . Overall influence considers influence to any composer after the death of composer  $i$ . See Sect. 2.3 for details on calculation and Sect. 3 for data sources. Dots are labelled for influential composers whose overall influence is more than five times larger than their near-term influence. The dashed line shows a linear fit



## 4.5 Simulating themes

As discussed in Sect. 1, we view the compositions in our data as having originated through, first, the creative process of each composer and, second, a selection process by which they are included in the canon of sufficiently well known works as proxied by Barlow and Morgenstern (1975, 1976). Our results above show that originality is positively associated with influence and several success measures among the composers in our sample. Therefore, originality appears to determine influence and success, rather than being traded-off against either. Does this mean then that a composer should at any rate maximize originality or are there any barriers to achieving high originality? To shed some light into the extent of such barriers, we simulate musical themes similar to those we find in the data in order to evaluate their originality against the corpus of themes in the data.

Proceeding conservatively, we draw notes one by one to assemble random themes, excluding relatively rare notes and evaluate each simulated theme's originality.<sup>21</sup> Fig. 11 shows the resulting distribution of originality scores over simulated themes and real themes of equal length.<sup>22</sup> Despite being composed of only the most common notes, the simulated themes tend to score higher in originality than those in the data.

This suggests that it is not the case that highly original themes are rare per se. Rather, it appears that the creative and selection processes impose constraints that disfavour themes which are as if randomly generated. These constraints may include a need for a theme to conform to norms that are not built into our model or to be pleasing to hear in order to achieve some initial success with audiences, peers, and critics. The creative artist should thus not maximize originality at any cost, but rather optimize it in consideration of—unobservable to us—norms and standards of a given time.

## 5 Conclusion

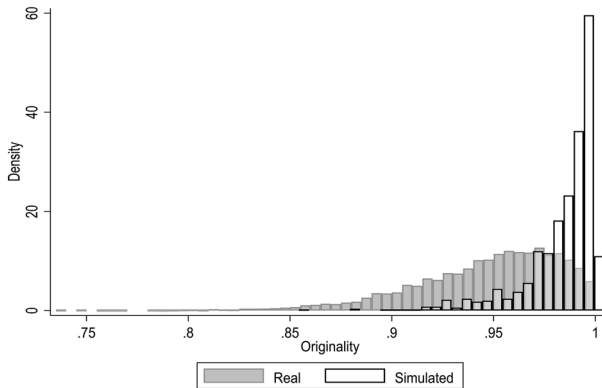
Identifying the style of a creative person or her work has long been of interest across several domains and is widely regarded as difficult. It is even more challenging to identify the extent of a person's influence upon others, or her originality. Measurement of these concepts is practically non-existent due to the elusive nature of style.

In this paper we overcome some of the difficulties by developing a novel approach to define and measure style. This enables us to rigorously identify and estimate the originality and influence of a creator or her creative output. Based on this framework, we study the interdependencies between originality, influence, and success in the context of music composition.

---

<sup>21</sup> We exclude 11 rare notes that appear in our data fewer than 0.2%, so that we draw only from the most common 12 notes. Including rare notes would naturally lead to even higher values of the originality measure.

<sup>22</sup> The simulated themes are each six notes long; the most common length in the data. This choice maximizes the corpus of real themes available for comparison.



**Fig. 11** Distribution of originality for real and simulated themes. *Notes:* The histograms show the distribution of originality for actual themes of six notes (the most common length) and simulated themes of the same length. The simulation (see Sect. 4.5) draws uniformly from the set of notes that appear in our data, excluding rare notes that are less common in our data than 0.2%. For details on the data see Sect. 3.1

Our results show that more original composers tend to be more influential upon the work of their later peers and are more successful with present-day audiences or experts. Originality is also predictive of influence and, to a lesser extent, success at the theme-level. We conclude that originality, rather than being traded-off against influence and success, is a key driver of both.

The empirical context and background of this study are distinctive, and so is the focus on a small, albeit important group of composers. However, the model proposed and mechanisms examined here are likely applicable to most settings where creative output is produced, especially since creative processes are likely comparable across domains and disciplines (van Broekhoven et al., 2020). This means that, for example, musicians and scientists may create two different kinds of work with different intentions and outcomes, but the process they use to get there appears to be similar (see Azoulay et al. 2017). Moreover, the method presented and tested in this paper, by enabling the measurement of the abstract concepts of style, originality and influence, opens new areas for future research, including on the role of design, leadership, and investment styles, or styles of political speeches and corporate communication. Finally, it is not only the measurement of originality or influence that is important, but also the very specific question of how these concepts are related to success. In addition to being of great importance to a composer, her success also brings benefits to her audience and to cultural production in her artistic field. On aggregate, success in the creative industries is relevant for economic growth.

Many creators, including artists, scientists, and entrepreneurs, are driven by the desire to become successful and influential. How to achieve that success or influence is fairly unknown. The *ex ante* uncertainty of the outcome (or the nobody knows theorem) is a notorious feature of cultural production and applies also more broadly to many areas of the creative industries (e.g., Schulze, 2005). However, by considering the mechanics of originality, some creators may be better capable to navigate

their mainstream success or have a greater chance to become influential by bringing a “change in the established modes” (Smith, 1976) in their fields.

## Appendix

### A Identifying key signature

Every key signature stands for either a major key or a minor key. For example, an empty key signature is either C Major or A Minor. In order to identify whether it is a major and minor key, we conduct various estimations. First, we look for early note in theme matching the major vs. minor key (often, first note is the tonic note of key). Second, we count tonic notes of major vs. minor key in the theme. Third, we count tonic chord notes of major vs. minor key. The predictive power of these estimations is then validated by using a sample of ca. 850 themes for which the true major or minor key is known from title of the work (e.g., “Prelude in C# Minor”). Using combinations of tests with high predictive power enables us to estimate the true key signature with a relatively high precision. In doing this, we obtain accuracy rates of >90%.

### B Originality, influence, and success for selected composers and countries

Table 9 shows originality and influence values for the top-30 composers based on the number of Spotify followers, and ranked in accordance with it. Column (3) reports the composer’s year of birth and column (4) the number of themes recorded for the composer. For Rachmaninov, Stravinsky Shostakovich, Ravel, Gershwin, and Prokofiev the influence measures cannot be computed as the composers are at the end of our sample and no musical themes composed after them are recorded in our data. Originality and influence measures for composers with very few themes, such as Erik Satie, should be considered as less reliable estimates.

Focusing on the geographical aspect, mean originality, influence, and success at the country level are plotted in three panels of Fig. 12. While, on average, Russian composers seem to be highly original and successful, their influence does not seem to be at the same level. Among the Russian composers, Igor Stravinsky stands out with his high originality, while Modest Mussorgsky is most influential, and Piotr Tchaikovsky is the most successful, closely followed by Sergei Rachmaninov and Igor Stravinsky.

European composers score high on average in all measures, in particular in influence. For example, Guillaume Dufay, composer and theorist of the early Renaissance, stands out for being the most influential composer of the region across the centuries: According to Brown (1974), Dufay “not only established the principles

**Table 9** Originality and influence of top-30 composers

(1) Rank	(2) Composer	(3) Birth	(4) N. of themes	(5) Originality	(6) Influence
1	Mozart, Wolfgang Amadeus	1756	641	83.10	60.48
2	Beethoven, Ludwig van	1770	471	73.51	61.24
3	Bach, Johann Sebastian	1685	525	100.00	62.53
4	Chopin, Fryderyk	1810	120	66.99	27.63
5	Vivaldi, Antonio	1678	47	89.00	43.09
6	Tchaikovsky, Piotr Ilyich	1840	138	66.73	4.13
7	Debussy, Claude	1862	106	78.05	– 8.01
8	Satie, Erik	1866	10	42.61	– 21.67
9	Schubert, Franz	1797	333	57.93	45.92
10	Rachmaninov, Sergei	1873	56	53.34	.
11	Brahms, Johannes	1833	371	67.91	– 2.89
12	Handel, George Frideric	1685	421	96.31	70.07
13	Wagner, Richard	1813	154	60.56	21.34
14	Liszt, Franz	1811	102	65.17	28.43
15	Dvorak, Antonin	1841	154	53.97	16.38
16	Stravinsky, Igor	1882	87	72.67	.
17	Grieg, Edvard	1843	110	60.19	21.72
18	Haydn, Franz Joseph	1732	368	83.52	55.37
19	Mahler, Gustav	1860	68	52.24	25.86
20	Verdi, Giuseppe	1813	182	70.53	33.44
21	Shostakovich, Dmitri	1906	46	55.73	.
22	Humperdinck, Engelbert	1854	37	29.84	8.85
23	Mendelssohn, Felix	1809	164	48.84	38.75
24	Schumann, Robert	1810	219	52.97	47.53
25	Ravel, Maurice	1875	71	65.27	.
26	Gershwin, George	1898	15	60.46	.
27	Saint-Saens, Camille	1835	100	41.87	16.45
28	Paganini, Niccolo	1782	27	57.91	25.06
29	Prokofiev, Sergei	1891	96	43.75	.
30	Puccini, Giacomo	1858	55	56.94	7.01

The table reports the birth year, number of themes, corrected originality measure,  $\omega_i^c$ , and the influence measure,  $\lambda_i$ , for the top-30 composers based on the number of Spotify followers. The data was collected and compiled by the authors (see Sect. 3 for details)

of style for the music of the Renaissance, but in his latest works, he also pointed the way to the future.”

Among American composers, worth of notice is Stephen Foster, highly influential composer of the 1800s who is often referred to as the “father of American music” (Songhall, 2020) and the “one of the first who made professional songwriting profitable.”

**Fig. 12** Originality, influence, and success by country



(a) Originality



(b) Influence



(c) Success

## C Originality and influence: Correction and standardisation

### C.1 Correcting for sparsity

In contrast to past style and future style, data on the empirical style of composer  $i$ ,  $x_i$ , tends to be sparse. This is because for most composers only a few features are recorded in the data and so the frequency is equal to zero for many features. Around 20% of composers have only one recorded theme and the average number of notes per theme is 6.9.

Figure 13 shows a scatter of the relationship between uncorrected originality,  $\omega_i$ , on the vertical axis, and the logarithm of the number of composer  $i$ 's themes (Panel A) and the corrected values of originality,  $\omega_i^c = \omega_i - \hat{\omega}_i^0$  (Panel B).

Composer  $i$ 's originality,  $\omega_i$ , is observed to be high when the number of themes is low (see Panel (A) of Fig. 13). This can be understood as a small-sample problem. We consider  $x_i$  to reflect the composer's style, but if only a few themes are available per composer, most entries of  $x_i$  are zero. This "noise" is captured in our projection as a departure from the past, resulting in an overestimation of originality.

To obtain a measure of originality that is unrelated to composer  $i$ 's number of themes,  $n_i$ , we want to adjust the uncorrected originality,  $\omega_i$ , by subtracting its *unoriginal expectation*,  $\omega_i^0 = \mathbf{E}_0(\omega_i | n_i, x_p)$ . This term captures the expected value of uncorrected  $\omega_i$  given the number of themes,  $n_i$ , written by composer  $i$  and the vector  $x_p$  describing past style under the assumption that the composer is entirely unoriginal, meaning that he samples only from the feature distribution of the past,  $x_p$ . In practice, we obtain a numerical approximation,  $\hat{\omega}_i^0$  to  $\omega_i^0$ , by sampling from the probability mass function associated with  $x_p$ .

We obtain the corrected values of originality as follows:  $\omega_i^c = \omega_i - \hat{\omega}_i^0$ . The resulting values are plotted in Panel (B) of Fig. 13. The originality values are now positively correlated with the number of themes (positive correlation of 0.49). While values of originality are expected to be positive, their realisations are sometimes below zero due to the correction (minimum of  $-0.14$ ). In the composer-level analyses that follow we use corrected originality and, for simplicity, refer to it as originality.

In order to facilitate the interpretation and the comparison among variables, all measures of originality and influence are standardized. The re-scaling is so that the lowest value is zero and the highest is 100.

Figure 14 shows histograms of uncorrected originality,  $\omega_i$  (A), corrected originality,  $\omega_i^c$  (B), and theme-level originality,  $\bar{\omega}_i$  (C).

Figure 15 shows how originality depends on the number of past themes (Panel A) and how influence depends on the number of future themes. While our baseline originality measure is uncorrelated with the number of themes in the past of composer  $i$  ( $-0.05$ ), a positive correlation ( $0.58$ ) can be observed between influence and the number of themes in the future of composer  $i$ .

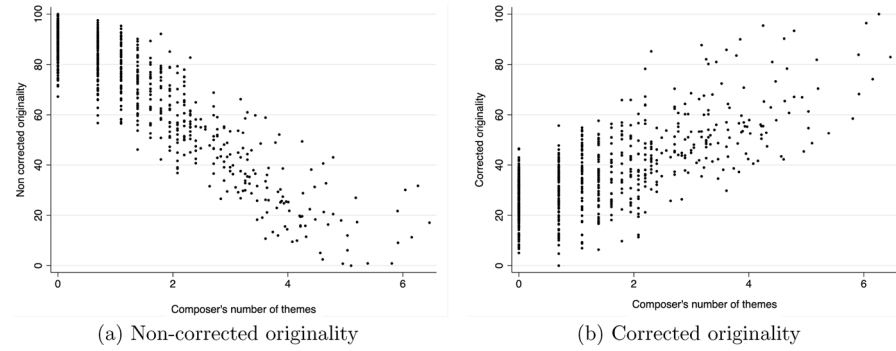


Fig. 13 Correction of the originality measure

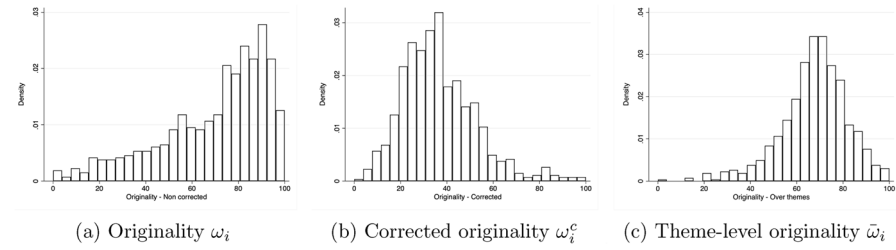


Fig. 14 Originality measures

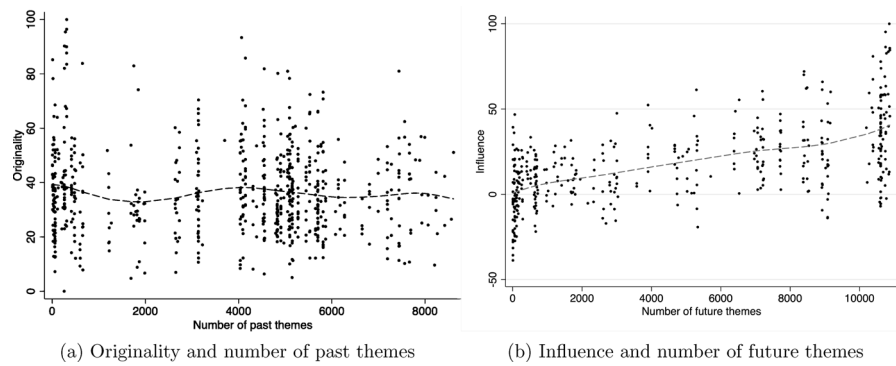
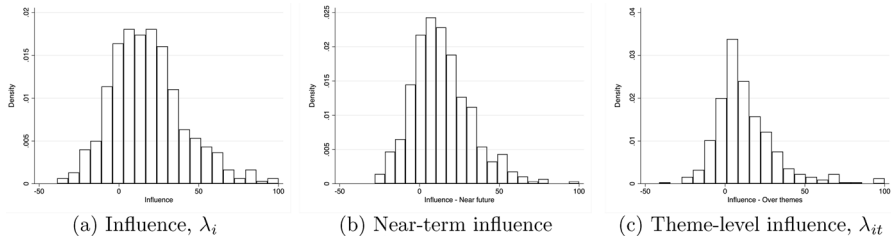


Fig. 15 Originality, influence, and the number of past or future themes

### D Additional figures

See Figs. 16, 17.



**Fig. 16** Influence measures. *Notes:* The histograms show three influence measures. The baseline influence measure,  $\lambda_i$ , is calculated using all features ever used by composer  $i$  and the full corpus of future music (Panel A). We extend this by a measure of near-term influence of composer  $i$  (Panel B), which shows composer  $i$ 's influence on the first generation of composers after composer  $i$ 's death (see details in Sect. 2.4). Panel (C) shows the theme-level influence,  $\lambda_{it}$



**Fig. 17** Nationality of composers

## E Additional regression results

### E.1 Determinants of originality

Many studies report on the effects of personality traits, life events or interpersonal relationship on individual originality (see overview by Simonton (2004)). Yet



what factors and in which combination determine the originality of creative output remains an important and unsolved question.

To address it with our measures of originality we made use of an additional dataset that, for a reduced sample of composers, contains information on teacher-student relationships as well as on composers' travelling patterns. The additional data was provided by Borowiecki (2022), where it is also described.

Table 10 reports results from an OLS regression in which composer's originality, as defined in Sect. 2.2, is regressed on characteristics of the composer, such as the

**Table 10** Determinants of originality

	(1) Originality	(2) Originality	(3) Originality
Teacher's originality	0.1581** (0.0674)	0.1461 (0.1155)	0.2914* (0.1506)
Composer's themes	0.0890*** (0.0133)	0.1006*** (0.0172)	0.1422*** (0.0273)
Past themes	0.0014 (0.0010)	− 0.0002 (0.0028)	− 0.0012 (0.0031)
Future themes	0.0010 (0.0007)	− 0.0031 (0.0022)	− 0.0015 (0.0028)
N. of teachers		0.0615 (1.9910)	− 1.3435 (2.2636)
Teacher's success		− 0.3322 (0.8059)	− 0.4353 (0.8984)
Teacher's influence		− 0.0898 (0.1169)	0.0253 (0.1339)
N. cities visited			− 0.3618 (0.4428)
N. conservatories as student			− 2.5807 (3.0218)
Years at conservatory			0.2413 (0.4599)
Half-century FE	No	Yes	Yes
Nationality FE	No	Yes	Yes
N	143	116	97
R-Square	0.318	0.517	0.570

The dependent variable is composer's originality, which is regressed on the originality of the most original teacher of a given composer (column 1), number of teachers, and success/influence of the most successful/influential teacher of a given composer (column 2). Column (3) includes additional controls for the number of cities visited, the number of conservatories at which the composer studied and the total years of conservatory studies. The table reports OLS estimates. Standard errors in parentheses. \*  $p < .1$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ . The data was collected by the authors (see Sects. 3 and E.1 for details)

number of cities he visited, the number of conservatories at which he studied and for how many years he studied. Moreover, using the data on teacher-student connections, we are able to include levels of originality, success, and influence of the teacher.

The most interesting result is a positive and significant effect of the originality of the teacher on the composer's own originality (column 1 and 3). This suggests that the originality of an instructor is likely to boost instead of hinder the creativity of her pupils.

## E.2 Variation of originality and success

We have observed in Sect. 4.1 that originality is conducive towards success. An emerging question is whether and how the variation of originality matters for success. Is it the case that particularly successful composers deliver consistently highly original output? Alternatively, it may be the case that successful composers produce output of mixed levels of originality, so that some works are path-breaking, while others are less progressive and hence perhaps more digestible by audiences and critics. This concern is studied in more depth here.

We calculate the standard deviation of originality for each composer using theme-level originality,  $\omega_{it}$ . Such a obtained coefficient represents the dispersion of the originality measure relative to its mean and could be thought as informative of how diverse in originality levels the various themes of a composer are.

In Table 11 we reproduce the results of Table 4, while adding the standard deviation of originality at the composer-level as a variable of interest. The standard deviation of

**Table 11** Robustness: Originality and success

	(1)	(2)	(3)	(4)	(5)	(6)
	Spotify foll	Murray	Word count	Spotify foll	Murray	Word count
Originality	0.0919*** (0.0073)	0.0356*** (0.0029)	0.0368*** (0.0033)	0.0846*** (0.0072)	0.0332*** (0.0029)	0.0339*** (0.0033)
Sd. Originality	0.0332 (0.0213)	0.0244*** (0.0085)	0.0202** (0.0097)	0.0331 (0.0204)	0.0230*** (0.0082)	0.0205** (0.0093)
Past themes				0.0004* (0.0002)	0.0000 (0.0001)	-0.0001 (0.0001)
Future themes				0.0005*** (0.0002)	0.0002*** (0.0001)	0.0003*** (0.0001)
Half-century FE	Yes	Yes	Yes	Yes	Yes	Yes
Nationality FE	No	No	No	Yes	Yes	Yes
N	411	411	411	411	411	411
R-Square	0.309	0.336	0.307	0.375	0.386	0.361

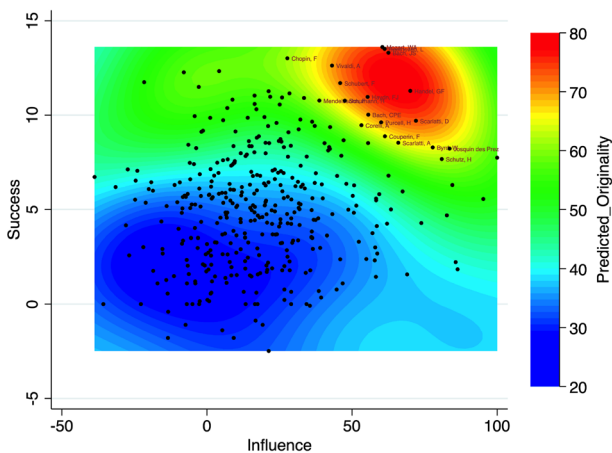
The dependent variables are alternative measures of success, the log number of Spotify followers (columns 1 and 4), the log of the Murray index (columns 2 and 5) and the log of the word count (columns 3 and 6). The table reports Seemingly Unrelated Regression (SUR) estimates. Standard errors in parentheses. \*  $p < .1$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ . The data was collected by the authors (see Sect. 3 for details)

originality is positive and significant for the Murray index and the word count success measures (columns 5–6). This suggests that not only originality matters for success but that some variation in originality among a composer's themes could also be beneficial. In other words, the prescription for success may require the production of both more and less original works.

### F Three dimensions: originality, success, and influence

Finally, we extend the graphical exploration to three dimensions: Originality, success, and influence. Figure 18 is an attempt of visualizing this relationship by the means of the three-dimensional heat map. The horizontal axis shows the measure of influence. The measure of success, the log number of Spotify followers, is on the vertical axis. The blue to red color scale represents different levels of predicted originality, which is computed by fitting a smooth function of success and influence to our originality measure, using a two-dimensional kernel. Consistently with what we have seen before, top composers in terms of long-lasting success show medium to high levels of originality.

In this contour graph, composers in the bottom-left area score low in influence, exhibit low originality, and have low success as captured by the log number of Spotify followers. The bottom-right quadrant has composers with average levels of originality and high influence but whose influence has not translated into long-term success. The highest levels of originality tend to be attained by composers with very high levels of both influence and success, as can be viewed in the top-right quadrant.



**Fig. 18** Relationship between composers' originality, success, and influence. *Notes:* The figure shows a contour plot resulting from a kernel regression of originality on success, and influence. For details on the measures used see Appendix C

**Acknowledgments** We thank Daniel P. Gross, Andrea Ichino, John O’Hagan, Noam Yuchtman and audiences at the European Historical Economics Society Conference (Paris), the 24th Dynamic Economics, Growth, and International Trade Conference (Odense), the 9th European Workshop on Applied Cultural Economics (Copenhagen), and seminar participants at the University of Bergen for many helpful comments. We also thank Harvard Business School Division of Faculty and Research and the University of Southern Denmark for financial support.

**Funding** Open access funding provided by University Library of Southern Denmark.

#### Declaration

**Conflict of interest** There is no conflict of interest.

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

## References

- Aggarwal, I., & Woolley, A. W. (2019). Team creativity, cognition, and cognitive style diversity. *Management Science*, 65(4), 1586–1599.
- Aghion, P., Dewatripont, M., & Stein, J. C. (2008). Academic freedom, private-sector focus, and the process of innovation. *The RAND Journal of Economics*, 39(3), 617–635.
- Akcigit, U., Caicedo, S., Miguelez, E., Stantcheva, S., & Sterzi, V. (2018). *Dancing with the stars: Innovation through interactions*. National Bureau of Economic Research: Technical Report.
- Azoulay, P., Liu, C. C., & Stuart, T. E. (2017). Social influence given (partially) deliberate matching: Career imprints in the creation of academic entrepreneurs. *American Journal of Sociology*, 122(4), 1223–1271.
- Barlow, H., & Morgenstern, S. (1975) *A dictionary of musical themes - Revised Edition*. Crown Publishers, Inc.
- Barlow, H., & Morgenstern, S. (1976). *A dictionary of opera and song themes - Revised Edition*. Crown Publishers, Inc.
- Bertrand, M., & Schoar, A. (2003). Managing with style: The effect of managers on firm policies. *The Quarterly Journal of Economics*, 118(4), 1169–1208.
- Borowiecki, K. J. (2013). Geographic clustering and productivity: An instrumental variable approach for classical composers. *Journal of Urban Economics*, 73(1), 94–110.
- Borowiecki, K. J. (2015). Agglomeration economies in classical music. *Papers in Regional Science*, 94(3), 443–468.
- Borowiecki, K. J. (2017). How are you, my dearest Mozart? Well-being and creativity of three famous composers based on their letters. *Review of Economics and Statistics*, 99(4), 591–605.
- Borowiecki, K. J. (2022). Good reverberations? Teacher influence in music composition since 1450. *Journal of Political Economy*, 130(4), 991–1090.
- Brown, H. M. (1974). Guillaume Dufay and the early renaissance. *Early Music*, 2(4), 219–233.
- Bussy, A., & Geiscke, F. (2021). *A geometry of innovation*. SSRN: Technical Report.
- Caspi, A., Harrington, H. L., Milne, B., Amell, J. W., Theodore, R. F., & Moffitt, T. E. (2003). Children’s behavioral styles at age 3 are linked to their adult personality traits at age 26. *Journal of Personality*, 71(4), 495–514.
- Chan, T. H., Mihm, J., & Sosa, M. E. (2018). On styles in product design: An analysis of US design patents. *Management Science*, 64(3), 1230–1249.

- De Vreese, C. H., Esser, F., Aalberg, T., Reinemann, C., & Stanyer, J. (2018). Populism as an expression of political communication content and style: A new perspective. *The International Journal of Press/Politics*, 23(4), 423–438.
- Dobbins, F., & His, I. (2019). “Le Jeune, Claude. Grove Music Online,” oxfordmusiconline.com.
- Encyclopaedia Britannica, “John Dunstable,” britannica.com 2020.
- Feinstein, J. S. (2006). *The nature of creative development*. Stanford Business Books.
- Feinstein, J. S. (2011). Optimal learning patterns for creativity generation in a field. *American Economic Review*, 101(3), 227–32.
- Foster, P., Mauch, M., & Dixon, S. (2014). Sequential complexity as a descriptor for musical similarity. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 22(12), 1965–1977.
- Gergaud, O., Laouenan, M., & Wasmer, E. (2017). A brief history of human time: Exploring a database of notable people.
- Giustiniani, V. (1961). Discorso Sopra la Musica. *Musica Disciplina*, 15, 209–225.
- Graddy, K., & Lieberman, C. (2018). Death, bereavement, and creativity. *Management Science*, 64(10), 4505–4514.
- Grove Music Online, oxfordmusiconline.com 2018.
- Hall, B. H., & Rosenberg, N. (2010). *Handbook of the economics of innovation* (Vol. 1). Elsevier.
- Hellmanzik, C. (2010). Location matters: Estimating cluster premiums for prominent modern artists. *European Economic Review*, 54(2), 199–218.
- Howes, A. (2017). *The spread of improvement: Why innovation accelerated in Britain 1547–1851*. Brown University: Unpublished.
- Jones, C. I. (2005). Growth and ideas. In *Handbook of economic growth* (Vol. 1, pp. 1063–1111). Elsevier.
- Keim, D. B., & Madhavan, A. (1997). Transactions costs and investment style: An inter-exchange analysis of institutional equity trades. *Journal of Financial Economics*, 46(3), 265–292.
- Kelly, B., Papanikolaou, D., Seru, A., & Taddy, M. (2021). Measuring technological innovation over the long run. *American Economic Review: Insights*, 3(3), 303–20.
- Kim, C., Yang, Z., & Lee, H. (2015). Parental style, parental practices, and socialization outcomes: An investigation of their linkages in the consumer socialization context. *Journal of Economic Psychology*, 49, 15–33.
- Lucas, R. E., Jr. (2009). Ideas and growth. *Economica*, 76(301), 1–19.
- Malmendier, U., Nagel, S., & Yan, Z. (2021). The making of hawks and doves. *Journal of Monetary Economics*, 117, 19–42.
- Mauch, M., MacCallum, R. M., Levy, M., & Leroi, A. M. (2015). The evolution of popular music: USA 1960–2010. *Royal Society open science*, 2(5), 150081.
- Meisenzahl, R. R., & Mokyr, J. (2011). The rate and direction of invention in the British industrial revolution: Incentives and institutions. In *The rate and direction of inventive activity revisited* (pp. 443–479). University of Chicago Press.
- Mokyr, J. (2009). Intellectual property rights, the industrial revolution, and the beginnings of modern economic growth. *American Economic Review*, 99(2), 349–355.
- Mokyr, J. (2016). *A culture of growth: The origins of the modern economy*. Princeton University Press.
- Munro, T. (1946). Style in the arts: A method of stylistic analysis. *Journal of Aesthetics and Art Criticism*, 5(2), 128–158.
- Murray, C. (2003). *Human Accomplishment: The Pursuit of Excellence in the Arts and Sciences, 800 BC to 1950*. Harper Collins.
- Petrucci Music Library, imslp.org 2020.
- Pfzinger, S. (2017). *Composer genealogies: A compendium of composers, their teachers, and their students*. Rowman & Littlefield.
- Rotemberg, J. J., & Saloner, G. (1993). Leadership style and incentives. *Management Science*, 39(11), 1299–1318.
- Schulze, G. (2005). Nobody knows anything-Or do we? Introduction to the Special Issue on the Movie Industry. *Journal of Cultural Economics*, 157–158.
- Serrà, J., Corral, Á., Bogañá, M., Haro, M., & Ll Arcos, J. (2012). Measuring the evolution of contemporary western popular music. *Scientific Reports*, 2(1), 1–6.
- Simonton, D. K. (1980). Thematic fame and melodic originality in classical music: A multivariate computer-content analysis 1. *Journal of Personality*, 48(2), 206–219.
- Simonton, D. K. (1984). Melodic structure and note transition probabilities: A content analysis of 15,618 classical themes. *Psychology of Music*, 12(1), 3–16.

- Simonton, D. K. (2004). *Creativity in science: Chance, logic, genius, and zeitgeist*. Cambridge University Press.
- Smith, A. 1976 [1759]. *The theory of moral sentiments*. Clarendon Press.
- Smith, A. (1980 [1795]). *Essays on philosophical subjects*. Clarendon Press.
- Smith, C. H., & Georges, P. (2014). Composer similarities through “The Classical Music Navigator”: Similarity inference from composer influences. *Empirical Studies of the Arts*, 32(2), 205–229.
- Smith, C. H., & Georges, P. (2015). Similarity indices for 500 classical music composers: Inferences from personal musical influences and “Ecological Measures”. *Empirical Studies of the Arts*, 33(1), 61–94.
- Songhall. (2020). “Stephen Foster,” [songhall.org](http://songhall.org).
- Squicciarini, M. P., & Voigtländer, N. (2015). Human capital and industrialization: Evidence from the age of enlightenment. *The Quarterly Journal of Economics*, 130(4), 1825–1883.
- van Broekhoven, K., Cropley, D., & Seegers, P. (2020). Differences in creativity across Art and STEM students: We are more alike than unlike. *Thinking Skills and Creativity*, 38, 100707.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.