UC Irvine
UC Irvine Previously Published Works

Title

Feature-Based Structures of Opportunity: Genre Innovation in the American Popular Music Industry, 1958 to 2016

Permalink

https://escholarship.org/uc/item/7267t6f5

Journal

American Sociological Review, 89(3)

ISSN

0003-1224

Authors

Kim, Khwan
Askin, Noah

Publication Date

2024-06-01

DOI

10.1177/00031224241246271

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Check for updates

American Sociological Association

American Sociological Review
2024, Vol. 89(3) 542–583
© The Author(s) 2024
DOI:10.1177/00031224241246271
journals.sagepub.com/home/asr
Sage

Feature-Based Structures of Opportunity: Genre Innovation in the American Popular Music Industry, 1958 to 2016

Khwan Kim[a] and Noah Askin[b]

Abstract

We offer a new perspective on how cultural markets are structured and the conditions under which innovations are more likely to emerge. We argue that in addition to organization- and producer-level factors, product features—the locus of marketplace interaction between producers and consumers—also structure markets. The aggregated distribution of product features helps producers gauge where to differentiate or conform and when consumers may be more receptive to the kind of novelty that spawns new genres, our measure of innovation. We test our arguments with a unique dataset comprising the nearly 25,000 songs that appeared on the *Billboard* Hot 100 chart from 1958 to 2016, using computational methods to capture and analyze the aesthetic (sonic) and semantic (lyrical) features of each song and, consequently, the market for popular music. Results reveal that new genres are more likely to appear following markets that can be characterized as diverse along one feature dimension while homogenous along the other. We then connect specific configurations of feature distributions to subsequent song novelty before linking the aesthetic and semantic novelty of individual songs to genre emergence. We replicate our findings using industry-wide data and conclude with implications for the study of markets and innovation.

Keywords

product features, innovation and diversity, cultural markets, the production of culture, music

Dating back at least as far as Schumpeter (1939, 1950), scholars have been interested in the relationship between market structure and innovation. Specifically, which configurations and concentrations of (typically large) firms in a market give rise to increased or diminished rates of innovation? Within sociology, there is a particular interest in the nature of this relationship in the creative industries (e.g., music, art, and film) due to their culture generating role (Caves 2000) and the acknowledgment that they "depend on continuous innovation" (Jones et al. 2016: 751). Indeed, work from both the production of culture perspective (Peterson and Berger 1975) and resource partitioning theory (Carroll 1985) has explored the effects of market concentration on innovation in various creative industries.

However, despite a common interest in the forces that drive product differentiation and conformity, these two streams of research

[a]INSEAD
[b]University of California-Irvine

Corresponding Author:
Noah Askin, University of California-Irvine, 4293 Pereira Drive, Irvine, CA 92697, USA
Email: naskin@uci.edu

generally reach opposite conclusions. The former, building on Schumpeter's (1950) premise, initially argued that a high concentration of large firms dampens innovation by reducing producer entrepreneurship (DiMaggio 1977; Dowd and Blyler 2002; Hirsch 1972; Peterson and Berger 1975). In contrast, the latter argues that a high concentration of large firms creates opportunities that increase foundings of smaller, specialist producers who, in turn, create novel product categories like craft beers and new film genres (Carroll and Swaminathan 2000; Mezias and Mezias 2000). Although the production of culture perspective has continually added nuance to its conceptualization of market structure and its consequences for innovation (e.g., Burnett 1992b; de Laat 2014; Lopes 1992), the two bodies of work generally differ in their perspectives and predictions (see Dowd 2004:1446).

One possible reason for the differing predictions is that across both streams of research, market structure has historically focused on the producers of innovation (i.e., firms and individual creators) and their characteristics. Yet work in the realm of popular music shows that different conceptions of market structure (e.g., looking at the degree of firms' decentralized decision-making in the market) negate much of the evidence connecting firm concentration to innovation (Burnett 1992b, 1993; Dowd 2004; Lopes 1992). Furthermore, the observation that market concentration, decentralized decision-making, and rates of innovation have all waxed and waned over the course of music history (Peterson and Berger 1996) has prompted scholars to posit that other features may structure markets (e.g., de Laat 2014). Following in this vein and acknowledging recent research pointing to the need to connect "the social, cultural, *and material* poles of innovation" (Cao, Chen, and Evans 2022, emphasis added; see also Wohl 2022), we build on the existing market structure research and propose adding *the products themselves* to the equation. Specifically, we want to understand how product features structure a market and promote or inhibit innovation in cultural industries.

We focus on product features because they are the locus of marketplace interaction between producers—the individual and organizational actors involved in the creation of a cultural product and its presentation to the market[1]—and consumers, the end-users of these products. Product features are particularly important for understanding new products because they are the "details" that "invoke the public's familiarity with [existing] technical artifacts and social structures" as a means of introducing and making comprehensible something new (Hargadon and Douglas 2001:477). Said differently, consumers use the features of new products to understand and situate them among existing products (Tversky 1977), leading to their adoption or rejection, whereas producers use features to distinguish their products and make them appeal to consumers. Therefore, the features of a cultural product (e.g., the ingredients in a dish [Rao, Monin, and Durand 2003] or the colors and styles used in clothing [Godart and Galunic 2019]) represent an intersection between consumer preferences and producer decisions.

It follows that aggregating the features of the available products in a market reveals an overall distribution: the landscape of consumer tastes and producer opportunity spaces. This feature distribution can be observed via regularly collected, updated, and interpreted information about market activity, referred to as market information regimes (Anand and Peterson 2000; Zanella, Cillo, and Verona 2022). In this article, we focus on two product dimensions, which we call aesthetic and semantic features (cf. Castañer and Campos 2002; Peterson and Beal 2001), and we propose that the distribution of these features provides another means of capturing the structure of a market that we can use to examine when and where innovation may emerge.

To answer our main research question—how product features structure a market and promote or inhibit innovation—we set out three objectives. Our first objective is to provide evidence that the feature-based market structure captured by dynamic market information regimes helps indicate when

innovation in the form of new genres is more likely to emerge, above and beyond the market- and producer-level factors that characterize previous research. As innovation can be defined as the successful implementation of creative output (Anderson, Potočnik, and Zhou 2014), and novelty is a necessary component of creativity (Amabile 1988), the appearance of a novel category (genre) is very much a sign of a successful innovation (van Venrooij 2015).[2] This is especially true when that appearance is in the mainstream. We argue that the emergence of new genres is more likely when the feature-based market structure is mismatched, meaning that between the two feature dimensions (i.e., aesthetic and semantic), one is more homogenous, and the other is more diverse.

Our second objective is to link these feature-based structural forces to an outcome of the cultural production process: the novelty of individual products. To draw theoretical and empirical connections from the market-level structure to individual product creation, we first classify markets into four different conditions based on the degree of diversity along each of the two feature dimensions (aesthetic and semantic): both diverse, both homogenous, and two conditions where one feature is diverse, and the other is homogenous. We then examine whether product novelty differs across these market conditions. As above, we highlight market-level feature configurations where one dimension is homogenous and the other is diverse as critical indicators of greater novelty in the individual products that subsequently appear in the market.

Our third objective is to "zoom in" on individual products to explore how their feature novelty plays a role in signaling innovation. More specifically, we ask whether and how a product's novelty within or across feature dimensions can be a catalyst for its becoming the originator of a ground-breaking category. In other words, does a product's feature novelty make it more likely to be the progenitor of a new genre, and what composition of novelty matters the most? To answer this question, we examine the aesthetic and semantic features of individual products and distinguish between two types of genre emergence: genre innovation, which reflects ground-breaking novelty that results in fundamentally new categories or new combinations of existing categories, and genre evolution, which also creates or recombines categories but without departing from the boundaries of a parent genre (i.e., new subgenres).[3] We hypothesize that both aesthetic and semantic novelty will be necessary for genre *innovation*, but that one or the other will be sufficient to generate genre *evolution*.

We explore our proposed phenomena in the American popular music market. The recent availability of fine-grained, song-level data in the form of sonic attributes and lyrical content means we can look in detail at the features of products in the music market in ways that have not historically been possible. We focus primarily on the *Billboard* Hot 100 charts, which capture the 100 most popular songs in the United States each week, from 1958 to 2016. These charts have been widely used to study the connection between market structure and innovation (Burnett 1992b; Dowd 2002; Dowd and Blyler 2002; Lena 2006; Lopes 1992; Peterson and Berger 1975), making them an ideal setting for our analyses.

Since the middle of the twentieth century, the *Billboard* charts have been the primary source for making sense of what was happening in the U.S. music industry (Anand and Peterson 2000). Even with increasing quantities of music released and made accessible over the course of our period of study, what happens on the Hot 100 chart still has an important signaling effect as a market information regime (Shi 2022; Zanella et al. 2022). Music industry stakeholders have limited capacity and cannot observe what is going on in the entire population, so while many niche charts exist (e.g., country, rock & alternative, R&B), the Hot 100 remains a focal point for "field participants to structure their beliefs about the success or failure of particular recordings, artists, and sub-genres" (Anand and Peterson 2000:281). The Hot

100 also generates the lion's share of field-wide media attention, which drives additional focus back to the chart and the genres that appear on it (see van Venrooij 2015:121). Our argument is that for something to truly be considered an innovation, it must achieve some degree of popular success such that it shifts a market's categorical structuring, making the Hot 100 an appropriate context in which to observe such phenomena.

However, despite its regular use in research on diversity and innovation, the Hot 100 is not without its issues. For example, it focuses only on the United States; prior to the introduction of Soundscan in 1991, it may not have accurately accounted for some genres; and it only captures a small, mainstream segment of the music market.[4] To address some of these concerns and test the robustness of our findings, we also collected a secondary dataset of over 1.5 million tracks, capturing a wide swath of the available music released during the same period as our Hot 100 data.

Our results suggest that product features provide another means of structuring markets, and they can offer clues as to when and where innovation is more likely to subsequently take place. We show that a market's potential for innovation varies with the degree of homogeneity and diversity along the semantic and aesthetic feature dimensions, even after controlling for other market- and producer-level factors likely to influence rates of innovation. We also draw a connection between market structure and product novelty, providing evidence of a mechanism connecting macro to micro market dynamics. Finally, we link product-level novelty back to innovation by showing that highly differentiated songs (i.e., those that are novel along both feature dimensions) are more likely to signal genre innovation, whereas songs that are novel along only one feature dimension are more likely to signal genre evolution. Taken together, these results reflect a cycle of innovation in the music market—one that connects market-level phenomena to cultural producers' output (i.e., the products) and back to market states. We therefore add a new layer to the long-running debate on market concentration and innovation: the role of product features as a market-structuring force.

THEORY

Market Information Regimes and Product Features in Creative Industries

In competitive fields like creative industries, organizations derive critical resources from their market (Chang and Chen 2020; Edelman and Yli-Renko 2010), but they must first find ways to make sense of the ambiguity and uncertainty typically surrounding markets (Salancik and Pfeffer 1978). To do so, it is critical to construct field-wide information regimes that help market actors "make sense" of the goings-on in the market. Anand and Peterson (2000:270) define these as "socially-constructed information regimes that compile reports about the market." By including "regularly updated information about market activity" (Anand and Peterson 2000:271), such information regimes create a common source of relevant information around which market actors can orient their understanding of the market.

In the context of the music industry, *Billboard*'s weekly charts create such a market information regime (Anand and Peterson 2000). Until recently, when the availability of daily charts provided by digital streaming services made weekly charts slightly less important, the commercial music industry relied heavily on *Billboard*'s chart data. In the absence of reliable, widely available market-level data, these charts embody a central mechanism of market dynamics, serving as short-hand descriptions of the competitive landscape in the industry, allowing record companies and musicians to continuously monitor each other and each other's products (White 1981)—as well as consumer demand—vis-à-vis their relative commercial success.

The charts provide a means for the producers and creators of music to spot opportunities for creativity and innovation, which are

determined by specific configurations of cultural and material elements (Godart, Seong, and Phillips 2020). In the realm of commercial music, those elements are song features—sonic attributes and lyrics. Artists create, and label executives select and promote, songs that vary along these feature dimensions in attempts to conform to or differentiate from the existing competitive landscape; consumers face choices about whether to listen to music that mirrors current trends or diverges from them. The *Billboard* charts provide regular snapshots of the industry from which producers and consumers make sense of the current musical landscape and then make choices. The features of the 100 most popular songs in a given week also shape the market context into which new genres arise.

New Genres as Innovations

Initially created as a means of segmenting the music market for the purposes of targeted advertising on radio (Negus 1999; Peterson 1990), genres have become *the* central categorizing force in the music industry (DiMaggio 1987; Frith 1996; Lena and Peterson 2008; McLeod 2001). As music's classification system, genres are a touchpoint for artists, consumers, critics, record labels, and audiences, with each party engaging each other in an ongoing conversation about the expectations, norms, boundaries, and commercial or artistic value of each genre (Lena 2012; Lena and Peterson 2008). Genres are highly contested spaces: the boundaries are debated and reinforced by various actors in the industry, each of whom is seeking to identify and expand or preserve what is "allowed" within a given genre (Becker 1982). As such, genres serve multiple purposes, helping to structure the music industry while also determining the appropriate sounds, behaviors, and identities for artists and audiences (Askin and Mol 2018; Becker 1982; DiMaggio 1987; Hsu and Hannan 2005; Phillips and Owens 2004). Genres create markets (Caves 2000; Peterson 1990) and generate status hierarchies (Bourdieu 1983, 1984). That genres go beyond merely dictating sounds and styles is reflective of their economic and social power. Frith (1996:76) points out their particularly critical role, writing that "genre is a way of defining music in its market, or alternatively, the market in its music."

Critically, when analyzing music industry dynamics, we must acknowledge that genres are not merely musical categories, they are a manifestation of deep-rooted power dynamics influenced by race, gender, and industrial frameworks. For example, from the mid-1920s to the mid-1940s, music was specifically aimed at different racial, ethnic, regional, and class groups. During this span, the correspondence between musical and social categories was obvious, most notably in the industry-wide consolidation around the "race" and "hillbilly" genres, driven by major labels' splitting of music along racial lines (Roy 2004). In the mid-1940s, these categories became known as "rhythm & blues" and "country" music, respectively, and they began to meld, with associated genres fusing and audiences mixing (Dowd 2003; Dowd and Blyler 2002; Roy 2004). Yet social categories and their respective power dynamics continue to play a significant role in shaping genre categorization, market structure, and the success of new genres and their pioneers. Within the hip-hop genre, for instance, lighter-skinned artists often achieve greater chart success, showcasing how race, intertwined with prevailing societal beauty norms, correlates with genre representation and success (Laybourn 2018). Intermediaries, such as critics, also demonstrate biases in categorizing musicians based on race, further perpetuating racial disparities in the music industry. Black artists are less likely to be categorized as boundary spanners, making it more difficult to be identified as an artist who recombines existing genres into something new (van Venrooij, Miller, and Schmutz 2022). We integrate racial and organizational dynamics into our analyses by considering a broad range of artist demographics and their overall representation in the market.

Because of the central position genres hold within the music industry, we propose

that the appearance of a new genre represents an innovation: something substantial enough to create a new market, generate economic activity, and change the organizing hierarchy in the industry (Dowd 2004; Lopes 1992; McLeod 2001). The mainstream appearance of a new genre—although a "discrete event" that is by no means guaranteed during the process of nascent genre development—is one of the "key events that 'inaugurate' the form at the field level as a new form and mark an important moment in its institutionalization" (van Venrooij 2015:110). Peterson and Beal (2001:233) note that three of the key processes involved in genre formation are determining the sound and style of what is deemed "acceptable" for the specific genre, understanding appropriate lyrical themes, and finding a target audience (see also Lena 2006). Naming the genre is a fourth key process, and it indicates a genre has created sufficient internal consistency and external differentiation to distinguish it from other genres (cf. Rosch 1978). Summing up, this work suggests innovation takes place when a genre receives a name that identifies distinct sonic and lyrical themes, *and* it finds or creates an appropriate market.

Importantly, the genre structure of the industry goes beyond "top-level" genres (e.g., "rock," "jazz," "country") and includes hundreds of subgenres—variations on existing themes or recombinations of existing styles into "streams" (Ennis 1992; see also McLeod 2001)—that are nested underneath the top-level genres (Lena 2006). New subgenres also represent important emergence, albeit on a less grand scale. To that end, we examine the two levels on which genres emerge: new top-level genres (as well as new combinations of existing genres), which represent innovations that shift the music market, and new subgenres (as well as new combinations of existing subgenres) that are more evolutionary but still represent significant structural and cultural developments.

Recall that we set out three objectives in order to answer our main research question about how product features may structure markets and promote or inhibit innovation: (1) demonstrating that product features provide a means of structuring markets and then drawing a connection between the aggregate feature-based market structure and subsequent innovation rates; (2) showing how certain configurations of the feature-based distributions are more likely to precede more novel products in the market; and (3) determining the extent of feature novelty that is more likely to lead to genre innovation versus genre evolution. We split each into its own "study," with its own theorizing and methodological approach. We connect them theoretically following the introduction to Study 1.

STUDY 1: PRODUCT FEATURES AND GENRE INNOVATION

We begin with the assertion that two feature dimensions are especially relevant to the emergence of innovation in cultural markets—story or narrative (semantics) and form (aesthetics). Castañer and Campos (2002) show that cultural producers generate novelty by differentiating along these dimensions (cf. Peterson and Beal 2001). A key assumption here is that an array of cultural products' features can be reduced to, and reasonably captured by, two dimensions. For instance, in the film industry, plots and screenplays are semantic features, whereas cinematography, sound effects, and casts are aesthetic features. In popular music, lyrics constitute semantic features,[5] and sonic attributes (e.g., tempo, key, timbre, mode) represent aesthetic features (Carey 1969; Hirsch 1971; Roy and Dowd 2010). When combined, the semantic and aesthetic features create a two-dimensional space within which cultural products can be positioned.

As a critical market information regime, the *Billboard* charts provide a dynamic context for observing the distributions of our two dimensions of interest over time. We argue that the configuration of the distributions of semantic and aesthetic features can serve as the macro structure of opportunities for both

product success (what is currently popular? [see, e.g., Askin and Mauskapf 2017]) and innovation (where is there untapped potential?). We focus on the latter question. Below, we define three possible configurations of market-level feature distributions based on the combination of homogeneity and diversity within each dimension (semantic and aesthetic): homogeneity in both, diversity in both, and homogeneity in one and diversity in the other. Drawing on competing predictions from research on differentiation and conformity, we theorize why the three different structures may present different levels of opportunity for innovation.

Differentiation–Conformity Tension

Research on differentiation in the creative industries suggests that product homogeneity in a market implies two conditions that may be conducive to innovation: (1) a high likelihood of unsatiated demand for novel, distinctive products, and (2) an existing hegemonic structure (organizational or product) that provides opportunities for new categories to differentiate from (Anand and Croidieu 2015; Anand and Peterson 2000; Carroll and Swaminathan 2000; Swaminathan 1995). Producers, sensing these conditions and market receptivity, will attempt to capitalize in the form of new organizations or novel products that can give rise to new categories.

Regarding unsated demand, the example of the popular music industry reveals that periods of high song homogeneity are followed by bursts of innovative sounds: in 1955, the rise of rock-n-roll was facilitated by pent-up demand from the 1940s and 1950s, when the commercial music market was dominated by just a few genres and record labels (Peterson and Berger 1975). In the realm of film, innovative genres are more likely to arise when big film studios dominate the market and focus on mainstream genres but fail to serve audiences' diverse tastes (Mezias and Mezias 2000). This unsated demand spurs the founding of small, specialized firms that increase the creation of new film genres.

Regarding hegemonic structure in a market, which indicates clarity around the organizations that have the most power or the products that are the most popular, research suggests that such conditions indicate a likelihood of greater product homogeneity. This homogeneity enables innovation in the form of new categories and genres because there is an existing frame of reference against which producers can more easily position a new category (Anand and Croidieu 2015; Lena 2012; van Venrooij 2015). This ecological approach explains category emergence across a range of creative industries. For example, various craft and specialty beer categories emerged in the U.S. brewing industry as a reaction to the homogenized offerings of the major breweries. Microbrewers used homogeneity in product features (i.e., a skewed feature distribution) as an opportunity structure providing guidance on where and how to innovate to meet consumer demand (Carroll and Swaminathan 2000). Similarly, new labels by smaller wineries in California emerged as they confronted factory-produced wine (Anand and Croidieu 2015; Swaminathan 1995); nouvelle cuisine arose as a challenge to classical French gastronomy (Rao et al. 2003); and modern architects fought over what differentiated modern architecture from traditional revivalist architecture (Jones et al. 2012). In summary, homogeneity in the marketplace suggests an opportunity for subsequent innovation.

However, when the market is instead characterized by feature diversity, it can be difficult to pinpoint a dominant market position. Such conditions make it challenging for producers to know whom to confront and where to differentiate. Furthermore, it may mean consumers have enough variety to satisfy varying tastes. Hence, this stream of research on differentiation suggests that feature diversity provides little information regarding where and how to innovate, only that such diversity is welcome. The implication is that subsequent rates of innovation are likely to remain flat.

Conversely, work on conformity and legitimacy speaks to a competing prediction. In

this view, feature homogeneity indicates that only a small number of legitimate schemas or feature configurations are guaranteed to gain acceptance and attention from audiences. Producers are pressured to conform around the taken-for-granted templates to gain legitimacy in hopes of increasing the likelihood of survival (Deephouse 1999; Greenwood and Suddaby 2006; Meyer and Rowan 1977; Scott 2003). Homogeneity thus suggests isomorphic pressure will drive producers to model their products after a "hit-making" schema rather than innovate. Those who fail to conform to the legitimate formulas are likely to be penalized and may not gain attention (DiMaggio and Powell 1983; Zuckerman 1999). Feature diversity, on the other hand, can be read as signaling more flexible legitimacy standards, as audiences will accept many different product features. From the perspective of conformity and legitimacy, when there is greater diversity in the market, attempts at innovation are less likely to be ignored or penalized (Phillips and Zuckerman 2001; Zuckerman 1999; Zuckerman and Kim 2003). Future innovation rates should rise. Taken together, work on differentiation, emphasizing the agentic behavior of producers, posits that innovation thrives under homogeneity, whereas research on conformity, focusing on the structural influence of market receptivity, suggests diversity is more conducive to innovation.

As discussed, two feature dimensions can combine into three different market configurations: homogeneity across both aesthetic and semantic features, diversity across both feature dimensions, and homogeneity along one dimension and diversity along the other. Regarding the first configuration, building on the differentiation and conformity literatures just discussed, we theorize that despite the potential benefits of capitalizing on unsated demand provided by differentiation, a *homogeneous cultural market in both dimensions* is unlikely to spur innovation. In markets where the reception of novel products is highly uncertain (Bielby and Bielby 1994; Leahey, Beckman, and Stanko 2017), and the market holds outsized rewards for successes (Rosen 1981), the risk of illegitimacy and its consequences is simply too great (Hsu 2006a, 2006b; Zuckerman and Kim 2003). We expect producers in such markets will play it safe and generally follow existing trends (see Figure 1a).

Similarly, although *diversity across both feature dimensions* implies relatively weaker isomorphic pressure from the market, we anticipate innovation in such a context will be less likely due to difficulties in determining where and how to differentiate (see Figure 1b). Said differently, when the market is already highly differentiated, it will be hard to differentiate enough to generate innovation.

As for the third configuration, *mismatched feature distributions*, in which there is homogeneity along one dimension and diversity along the other, the tension between differentiation and conformity suggests the following: Cultural producers seeking to innovate should prefer market conditions with a powerful, hegemonic reference point that provides something against which they can differentiate. Simultaneously, they should prefer some degree of diversity in the market because it indicates the market is more accepting of newness, thereby lowering the stakes associated with innovation and signaling that divergence is not only tolerated but potentially celebrated. Therefore, producers can aim to strike a balance, tailoring their innovations to stand apart from the homogeneous baseline while simultaneously capitalizing on the audience's demonstrated appetite for uniqueness. This leads us to argue that the market will see higher subsequent levels of innovation when there is a mismatched distribution landscape (see Figure 1c). A priori, we are agnostic about which dimension should be in each condition. Stated formally:

> *Hypothesis 1:* Rates of genre innovation will be higher following markets in which there is a mismatch between the distributions of semantic and aesthetic features—that is, one is homogenous and the other is diverse.

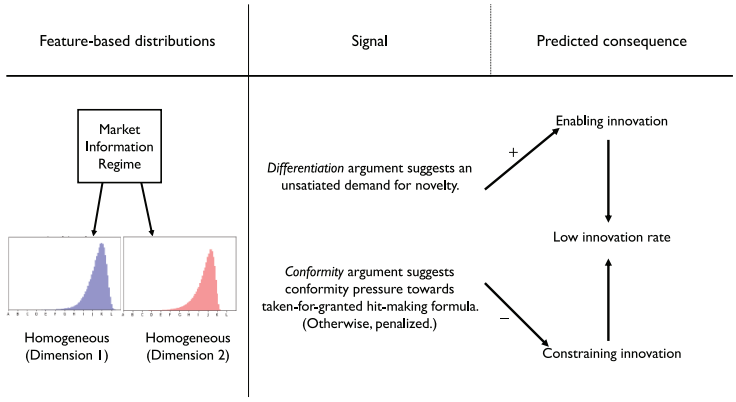

Figure 1a. Market Homogeneity in Both Feature Dimensions

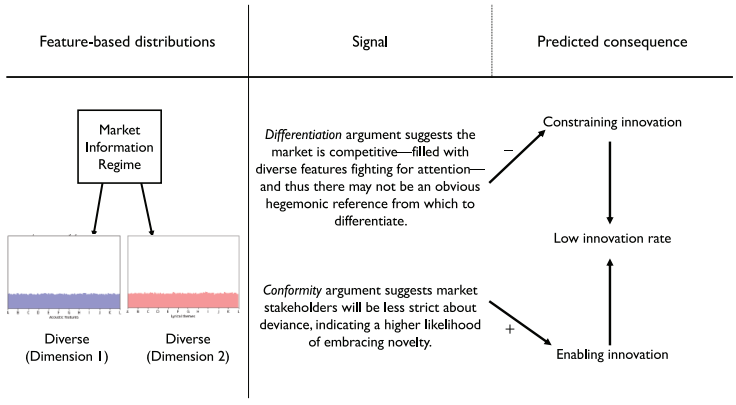

Figure 1b. Market Diversity in Both Feature Dimensions

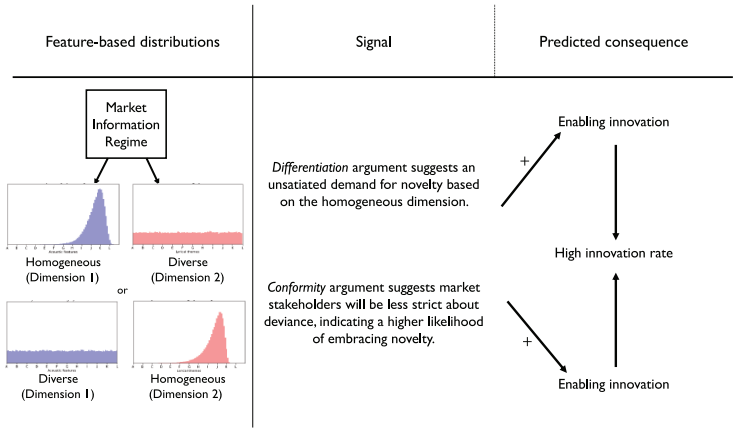

Figure 1c. Homogeneity/Diversity Combination: Mismatched Feature Dimensions

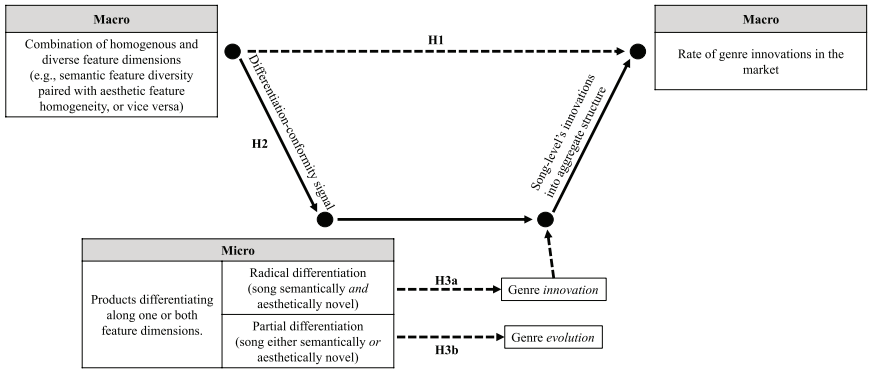

Figure 2. Theoretical Model Using Coleman's Boat

Connecting Market-Level Feature Distributions and Innovation in Cultural Markets

We have thus far argued that the composition of market-level feature distributions will influence rates of innovation in that market. However, for structural (i.e., macro) conditions to influence market-level outcomes, there must be micro-level changes that reflect reactions to those conditions that can ultimately be aggregated up to the subsequent outcomes (see Coleman 1986). We therefore use Coleman's (1986, 1994) "boat" or "bathtub" model and follow recent work that operationalizes it empirically as a means of theorizing these macro-micro links—the dynamics between social structure and individual or organizational behavior (e.g., Conti, Kacperczyk, and Valentini 2021).

As depicted in Figure 2, which summarizes the theoretical pathways we propose, we argue that the composition of macro-level feature distributions will point to subsequent rates of innovation in the market (Study 1). What underlies this relationship is that, first, certain combinations of feature distributions signal market conditions that are more amenable or conducive to novelty, which will catalyze producers to create and disseminate music with novel features (macro-to-micro link 1; Study 2). This generation of novelty at the product level, in turn, enhances the likelihood that a song becomes a progenitor of a new genre or subgenre, depending on whether that novelty reflects innovative or evolutionary differentiation (micro-to-micro link 2; Study 3) and there is sufficient audience convergence around that novelty (Boone et al. 2012; Lena 2012; van Venrooij 2015). Finally, when we observe these innovative activities across the micro-level behaviors and aggregate them, we witness a transformation in the social structure, reflected in a measurable change in the macro-level rates of innovation (micro-to-macro link 3). In summary, our "boat" model draws a systemic connection between the distribution of features within a cultural market and its subsequent innovation rate, mediated by the actions and receptivity of individual market participants.

STUDY 2: FEATURE DISTRIBUTIONS AND PRODUCT NOVELTY

Our model posits that diversity in one dimension and homogeneity in the other is most likely to result in higher market-level rates of innovation. This hypothesis consists of two mechanisms. First, producers will be more likely to create—and audiences will be more amenable to—more novel products, which will see their presence in the market increase following mismatched feature distributions. Second, those more novel, differentiated

products will be more likely to become the progenitor product of a new category—that is, an innovation.

The second mechanism requires further theorizing that will be addressed in the next section. In terms of the first mechanism, we follow the same logic as Hypothesis 1, except at the producer rather than the market level. Artists and label executives will view mismatched distribution landscapes (i.e., either diversity–homogeneity or homogeneity–diversity) as indicating greater consumer receptivity toward novelty and differentiation than aligned landscapes (i.e., diversity–diversity or homogeneity–homogeneity). Accordingly, we expect product-level differentiation vis-à-vis a comparison set of recently released products—our operationalization of novelty—to be higher when recent market conditions reveal a mismatched distribution landscape.

Hypothesis 2: Mismatched feature distributions are more likely to precede greater cultural product novelty than are aligned feature distributions.

STUDY 3: FEATURE DIFFERENTIATION, GENRE INNOVATION, AND GENRE EVOLUTION

Finally, we turn our focus to individual products and their features. Product features are crucial because innovation demands novelty (Becker 1982), and novelty comes from, among other things, the ideological and aesthetic distinctiveness of product features (Jones et al. 2012). Here, we are interested in drawing the connection between producers creating a novel product and that product's likelihood of finding an audience large enough to register as an innovation. More specifically, if our interest above was in the range of sounds and lyrics presenting opportunities for differentiated music, our interest here is in determining whether products on the more novel end of those dimensions are more likely to become innovations.

Innovation, however, is not a uniform construct: it can be decomposed into different varieties (see note 3). Indeed, extensive feature differentiation is the start of genre *innovation* (Lampel, Lant, and Shamsie 2000), but such de novo category creation is a rare event, representing the advent of an entirely new set of institutional arrangements. Once innovation in the form of the creation and naming of a new category takes place, the legitimacy conferred shifts the attention of producers and audiences (Anand and Croidieu 2015; Croidieu, Rüling, and Boutinot 2016; Lamont and Molnár 2002; Lena and Peterson 2008). Subsequently, rather than fighting for category-level recognition, producers focus on more individualistic identity claims, differentiating and competing for audience attention within the new category. This, in turn, spawns new products that do not disrupt but sustain the new category via *evolution* (Beverland 2006; Carroll and Wheaton 2009; Peterson 2005). In summary, genre innovation breeds a number of subgenres under a parent category (see McLeod 2001), which we call genre evolution, and which emphasizes processes of institutional exploration within an established category structure. Genre evolution occurs when feature differentiation does not violate the broader logic of a parent genre yet creates a new subgenre that extends its parent genre.

Two examples highlight the distinctions between genre innovation and evolution. First, in examining the emergence of recorded jazz in the 1920s, Phillips and Owens (2004) show that smaller record labels that had the will and capability to differentiate dramatically (in this case, because the dominant labels would not do so for fear of race-related backlash) were more likely to generate genre innovation (e.g., recording a new genre like radical jazz). This was despite the risk of an illegitimacy penalty.[6] In contrast, firms that aimed to create novel sounds but feared an illegitimacy penalty (or worse) were more likely to produce new *sub*genres, which contributed to the evolution of jazz more generally. Hip-hop presents a second example. While hip-hop had to revolutionize the way

its music was made and heard in order to establish itself as a genre in the 1970s, dozens of subgenres have emerged within hip-hop by delicately differentiating their core message (lyrics) or musical style (sonics) (Alridge 2005; Alridge and Stewart 2005).

These examples suggest there is value in contrasting trailblazing songs with songs whose contribution is smaller in scope, extending existing genres by creating subgenres. We therefore examine the effect of a song's differentiation along each feature dimension, as well as their interaction, on genre innovation and genre evolution.[7]

Hypothesis 3a (innovation): Genre innovation, represented by the appearance of a new parent genre (or new combination of parent genres), is more likely when a song's lyrical novelty and sonic novelty are both high.

Hypothesis 3b (evolution): Genre evolution, represented by the appearance of a new subgenre (or new combination of subgenres), is more likely when either the lyrical novelty or the sonic novelty of a song is high, but not when both are.

DATA AND METHODS

Continuing in the tradition of the production of culture's explorations of diversity and innovation in the music industry, we rely on the weekly *Billboard* Hot 100 chart, which began publishing week-to-week rankings of popular songs on August 4, 1958. Our data run from that date through March 26, 2016. The *Billboard* charts are among the most reliable and accurate barometers of success and popularity in the U.S. music industry (Lopes 1992) and have been widely used by sociologists, management scholars, and musicologists (Berg 2022; Burnett 1992a, 1996; Dowd 2000, 2004; Lee 2004; Lena 2006; Lopes 1992; Peterson and Berger 1975; Shi 2022; Watson and Anand 2006; Zanella et al. 2022). We use 22,319 songs' weekly chart appearances, which add up to 276,846 song-weeks, from 1958 to 2016.[8]

Using these songs as the foundation, we collected detailed data on the songs, their artists, and the firms that produced them. First, we capitalized on advances in music information retrieval, a field at the intersection of machine learning, computer science, and musicology, to capture 11 sonic features about each song. Table 1 describes the 11 features we collected. These include standard musicological attributes like "tempo" and "key" as well as more algorithmically-derived features like emotional "valence" and "danceability," which aim to better capture the subjective sonic nature of a song (for more detail, see Askin and Mauskapf 2017). These data, initially generated by algorithms created by a company called The Echo Nest, now help power Spotify's recommendation engines.

Second, we collected lyric data from various websites like MetroLyrics.com, lyrics.com, and Genius.com.[9] Third, we used Wikipedia, *The Encyclopedia of Pop, Rock & Soul*, recording labels' official websites, Allmusic.com (an online music encyclopedia), and Discogs.com (an online recorded music database and marketplace) to gather artist demographic information like gender, race, and a group indicator (i.e., whether a musician is a solo artist or group), as well as organizational information about firms (i.e., record labels) like organizational structure, firm age, whether founded by a musician, and so on.

Finally, we collected every song's primary genre—Discogs has 15 genres (e.g., rock, R&B, rap, pop), 14 of which appear in our data—and subgenre (258, including art rock, Chicago soul, jazz funk, and so on). Genres and subgenres in Discogs are user-generated, but the options for each are limited, as are the number of genres (three) and subgenres (six) that can be attributed, specifically to avoid confusion presented by "excess styles." Genre and subgenre attributions are community-approved, meaning any new release entered into the Discogs database has to receive sufficient supporting votes from the community to get listed on the site.[10] The fan-based nature of the genre attributions likely reflects industry designations of genre, meaning there is likely to be strong validity to the genres and subgenres associated with each song, but it also offers greater freedom and less corporate influence than would genre data obtained from a tightly-controlled

Table 1. The Echo Nest / Spotify Sonic Features

Feature	Scale	Definition
Acousticness	0–1	Represents the likelihood the song was recorded solely by acoustic means (as opposed to more electronic/electric means).
Danceability	0–1	Describes how suitable a track is for dancing. This measure includes tempo, regularity of beat, and beat strength.
Energy	0–1	A perceptual measure of intensity throughout the track. Think fast, loud, and noisy (i.e., hard rock) more than dance tracks.
Instrumentalness	0–1	The likelihood a track is predominantly instrumental. Not necessarily the inverse of speechiness.
Key	0–11 (integers only)	The estimated, overall key of the track, from C through B. We enter key as a series of dummy variables.
Liveness	0–1	Detects the presence of a live audience during recording. Heavily studio-produced tracks score low on this measure.
Mode	0 or 1	Whether the song is in a minor (0) or major (1) key.
Speechiness	0–1	Detects the presence of spoken word throughout the track. Such vocals are not considered for the model.
Tempo	Beats per minute (BPM)	The overall average tempo of a track.
Time Signature	Beats per bar/measure	Estimated, overall time signature of the track. 4/4 is the most common time signature by far and is entered as a dummy variable in our analyses.
Valence	0–1	The musical positiveness of the track.

Note: See https://developer.spotify.com/documentation/web-api/reference/#object-audiofeaturesobject.

music encyclopedia like Allmusic. Discogs is a common source of genre data in music industry studies (Askin and Mauskapf 2017; Shi 2022; van Venrooij 2015; van Venrooij and Schmutz 2018).

Given the different levels of analyses required for our three studies, we constructed three datasets. Across all studies, we also perform robustness checks with an expanded dataset capturing much of the recorded and released music in the United States during the same years covered by our Hot 100 data (see Appendix A for more details).

STUDY 1: PRODUCT FEATURES AND GENRE INNOVATION

Dependent Variable

Rate of genre innovation. Our first dependent variable is the percentage of songs in innovative genres appearing on the Hot 100 chart in each time window.[11] We count two types of genre creation as innovative—that is, the successful implementation of novelty. The first type, the conventional way of measuring innovation in studies of popular music (Coase 1979; Lopes 1992; Negus 1999; Peterson and Berger 1975), is the simple appearance of a song in a novel genre: for instance, the first time a hip-hop song appears on the chart. The second type of innovation takes its intuition from work on recombination (Fleming 2001; Hargadon and Sutton 1997; Schumpeter 1934) and is defined as the appearance of a song with a novel combination of existing genres. Songs, like films (Hsu 2006b), can be categorized with multiple genre denominations. If a song is labeled as both hip-hop and electronica, for example, and that combination of genres has not previously appeared, we also count it as an innovation.

As we do not have an exhaustive history of genres, it is difficult to determine whether a song's genre is new and thus innovative in

the early years of the *Billboard* chart. For example, although several jazz songs show up in the very first week of the Hot 100, we know jazz had long been in the popular realm (see Phillips 2013). Accordingly, genres that already existed in the early period serve as a baseline group instead of being included in our analyses. We only treat genres as new if they do not appear within the first eight weeks of data. Furthermore, while songs whose genre(s) appear on the charts for the first time are deemed innovations, we do not believe that only one such song is innovative. Other songs in the same new genre(s) were likely written or recorded around the same time and may have taken slightly longer to be released or to reach the chart. Following others who have studied innovation in music (de Laat 2014), we count a song as an innovation if it appears within one year of the genre's (or genre combination's) first appearance on the charts.

To calculate the rate of genre innovation in the field, we take the ratio of the number of songs in novel genres to the number of total songs over 52-week periods, but sliding 26 weeks at a time (meaning there are 26 weeks of overlap in consecutive 52-week windows, see the Analytic Strategy section).[12] In all our models, we include a lagged version of this dependent variable to account for autocorrelation. In our robustness checks, we replace this dependent variable with one measured at the population level (i.e., the wider music industry, not just the Hot 100) to see if any association between the feature landscape and genre innovation persists across the mainstream market of the Hot 100 and the industry in general.

Independent Variables

Market-level semantic (lyrical) homogeneity. To capture the level of semantic homogeneity in the popular music market, we encoded the lyrics of each song into a single vector using the pre-trained BERT (Bidirectional Encoder Representations from Transformers) model, a Natural Language Processing (NLP) model developed by Google (Devlin et al. 2018). BERT is widely used to extract features of text and represent them numerically via word or sentence-embedding vectors. What makes BERT particularly powerful is its use of contextual word embeddings: whereas Word2Vec and older NLP techniques produce the same word embedding for a word regardless of context, BERT takes context into account, providing more accurate feature representations and thus improving model accuracy. BERT "looks" in both directions (left and right surroundings of a target word) in a sentence simultaneously, using the full context to predict the target word. That is, unlike previous models, it takes both the previous and next tokens (i.e., words) into account at the same time—the most cutting-edge models pre-BERT were missing this "same time" part. BERT functions by randomly masking some words in sentences and then predicting those masked words, while also predicting the next sentence given the features of a focal sentence. In doing so, BERT models the association between two sentences, a feature that allows BERT to outperform other state-of-the-art algorithms (Yoshimura et al. 2019; Young et al. 2018; Zhang et al. 2019).

Effectively, BERT is a model that understands how to represent text. For our purposes, we use the BERT base uncased pre-trained model (24 layers, 1,024-dimension, 16 heads, 340M parameters) from BERT-as-service (Xiao 2018). We feed it lyrics, consisting of a sequence of words, and it scans to the left and right of each word simultaneously to produce a 1,024-dimension vector representation of each word. Next, we average the vectors for all words in a focal song to produce a single vector, yielding an averaged 1,024-dimension vector for each song. We then collect all songs in each 52-week period and average their song-level vectors to create a single vector that represents the market's lyrical vector for the time window. Finally, we calculate the cosine similarity between each song's vector and the market-level vector for its 52-week period. This similarity measure reflects how

similar each song is to the market-level centroid: the mean of these cosine similarities for all the songs in a given market represents the semantic homogeneity in the market for that period. Concretely, the average cosine similarity is high (i.e., the market is semantically homogenous) if most of the songs are concentrated around the market-level centroid; it is low if most of the songs are widely dispersed, indicating semantic diversity.

Market-level aesthetic (sonic) homogeneity. To measure aesthetic homogeneity, we begin with the 11 sonic attributes from Spotify that represent each song's aesthetic features. The attributes are captured numerically: most range between 0 and 1, and some are in different scales, such as tempo (e.g., 120 beats per minute) and key (e.g., C#, which is 1 on a 0 to 11 numeric key scale). To vectorize each song, we normalize all sonic attributes to the same 0 to 1 range and integrate them into a single, 11-dimension sonic vector. We then calculate the market-level aesthetic homogeneity similarly to semantic homogeneity. We collect all songs in a given 52-week period and average their song-level aesthetic vectors to create a single vector that represents the entire market's aesthetic vector for the time window. Next, we calculate the cosine similarity between each song's vector and the average market-level vector for its time period. This cosine similarity score reflects each song's similarity to the market-level centroid; the mean of the cosine similarities for all the contemporaneous songs represents the aesthetic homogeneity in the market for that period. Concretely, the average cosine similarity is high (i.e., the market is aesthetically homogeneous) for a 52-week period if songs are concentrated around the market-level centroid. It is low if songs are widely dispersed, indicating aesthetic diversity.

Control Variables

Lagged rate of genre innovation (lagged DV). Whether a technological breakthrough or a sociological change, innovation is path dependent (Inglehart and Baker 2000; Tavassoli and Karlsson 2015). This means current innovation can be explained by past innovation, raising the possibility of autocorrelation. We therefore control for the rate of genre innovation in the prior period (i.e., the lagged DV).

Market concentration. Our analyses include three variables to control for the effects of market-level firm composition that research suggests may affect product innovation. First, following the tradition in the production of culture literature (Alexander 1996; de Laat 2014; Dowd 2004; Dowd and Blyler 2002; Lee 2004; Peterson and Anand 2004), we compute and include the Herfindahl index—the sum of the squared market shares—among parent record labels represented on the Hot 100 in a given period to control for market concentration.

Prevalence of independent firms in the market. Second, we use a measure that captures the prevalence of independent firms: the ratio of independent record labels in the market. Prior literature suggests innovation is more likely in markets with small specialized organizations among the population (Baum and Mezias 1992; Carroll and Swaminathan 2000; Mezias and Mezias 2000). We therefore include the ratio of the number of independent ("indie") labels with songs on the charts to the total number of record labels placing songs on the charts in a given period.

Prevalence of decentralized organizational structure in the market. The above variable captures the pervasiveness of indie labels, but it does not consider whether those indie firms were purely independent, such that the label is responsible for all processes associated with the creation and distribution of a record. This includes producing, distributing, marketing, and copyrighting, among other responsibilities. One label does not necessarily handle everything. Some labels may be partially independent,

responsible for producing an album but leaving the work of distribution, marketing, and copyright to a parent label. In the debate about market concentration and innovation, scholars have shown that the expansion of decentralized production in the popular music industry (i.e., major labels using subsidiary labels to release certain content) effectively reduced the negative influence of market concentration on creating new music (Burnett 1992b; Dowd 2004; Lopes 1992). Big firms provide their subsidiaries with financial and marketing support, as well as sufficient discretion, enabling them to generate novel cultural products.

Examples of decentralized labels include acquired indie labels (e.g., Motown, which was an indie label until it was acquired by MCA Records in 1988 and kept releasing albums under the Motown name); indie labels with a major label distribution deal (e.g., Cash Money, which was run independently until a $30 million distribution deal with Universal Records in 1998); and original imprints established as subsidiaries by a larger parent firm (e.g., Polydor, which was created in 1954 as an imprint of Deutsche Grammophon). To account for the potential influence of this diffusion of decentralization, we build on Dowd's (2004) approach to measure decentralization. Because label decentralization can vary over time, as in the case of Motown or Cash Money, we collected information on organizational changes such as mergers, acquisitions, and distribution deals from labels' official websites, news articles, and Wikipedia. These data allow us to track any events that indicate the decentralization of a focal label. Concretely, in our data, Motown is coded as an indie label until its acquisition in 1988 and then as a decentralized firm from 1988 onward. We code 1 if a subsidiary or independent label is owned or distributed by a major label, deeming that major label and any subsidiaries to be "decentralized." We then include in our models the ratio of songs produced by decentralized firms to the total count of songs on the chart in a given period.

Genre homogeneity in the market. To avoid conflating genre presence and our feature dimensions (e.g., rock songs that all sound similar may be heavily overrepresented on the chart in a given period), we include controls for genre homogeneity. First, we count the number of songs in each genre in each period. To turn the counts into a single market-level homogeneity score, we used Teachman's entropy index (Teachman 1980), originally developed by Shannon (1948). Following Harrison and Klein (2007), we chose Teachman's entropy because our theoretical specification of homogeneity (or, inversely, diversity) is closely related to how they define variety. Entropy is at its maximum when all outcomes are equally likely or, in terms of genre variety, the number of songs from each genre in a given period is equal to one another. Inverting this, our homogeneity measure is (1 – normalized Teachman's entropy); genre homogeneity is high when the market is dominated by a few primary genres. To create this measure, we first calculated the entropy score of a genre i, E_i, reflecting genre diversity in each period measured as

$$E_i = -\sum_i P_i \ln P_i$$

where P_i is the proportion of genre i's occurrences in each market. We then normalized that score to put it on a 0 to 1 scale and subtracted it from 1, returning a homogeneity score. A market is highly genre homogeneous when this value is closer to 1, and highly diverse when closer to 0. Our models include this homogeneity variable for primary genres and subgenres.

We next consider demographic factors of the musicians and bands on the Hot 100, as they are likely to exert influence on the emergence of new musical genres. For example, research suggests there is a relationship between inequality and creativity (Godart et al. 2020; Storper and Scott 2008), pointing to an alignment between social distinctions (e.g., race, gender, and class) and musical differentiation (e.g., creating new genres) (Roy

and Dowd 2010). Race and gender also play a role in the genre attributions "assigned" to artists (Peterson 1997; Phillips and Kim 2009; Schmutz 2009; Singerman 1999) and whether artists are permitted to span categories (van Venrooij et al. 2022), both of which will affect the likelihood of being among the first artists to be identified as representing a new genre. Considering the power dynamics at play in the music industry, it is possible that minorities could be a source of innovation (Phillips and Owens 2004) or be restricted from innovation opportunities (van Venrooij et al. 2022). Furthermore, work on the relationship between innovation and group dynamics suggests innovation is more likely to come from teams than from individuals, as collaborative efforts allow for specialization and the efficient division of labor (Guimerà et al. 2005; Uzzi et al. 2013). To account for these factors, we introduce the following three control variables.

Female ratio. The female ratio captures the proportion of female artists, whether solo or within a group, relative to the total number of musicians on the Hot 100 during a specific time window. In our calculation of the female ratio, we classify a focal musician as female if she is either a solo female artist or part of a female-dominant group, excluding cases where group composition is equally split between male and female members (e.g., The Mamas & The Papas).

Non-White ratio. The non-White ratio represents the proportion of non-White artists relative to the total number of musicians on the same charts. We classify a focal musician as non-White if they are either a non-White solo artist or a non-White group. Cases where a group includes an equal number of White and non-White members (e.g., B.B. King and Eric Clapton for the album "Riding with the King," released in 2000) are excluded from this computation.

Group ratio. The group ratio reflects the ratio of non-solo artists to the total number of musicians within the Hot 100 in each period.

Thus far, our control variables are constructed at the level of the Hot 100 chart. However, the emergence of innovative genres that ultimately land on the charts might be catalyzed by market characteristics and information not just taking place on the Hot 100, the most popular marketplace, but also by dynamics playing out across the broader music industry. To address this potential influence, we collected every song from each of the record labels that put at least one song on the Hot 100 during the 59 years covered by our data, plus their subsidiary and imprint labels. This amounted to 1.53 million songs, for which we collected additional data from Discogs, Spotify, and Genius. We ultimately added roughly 374,000 songs (due to data availability) recorded in the U.S. market during our observation period. With these data, we added five controls (see Appendix A for details about the additional data and variable construction).

Release density in population. First, we compute the number of all releases—whether an album or a single—produced each year to address the density-dependence effect of the labels on the new market entries, which may in turn affect the emergence of new genres (Dowd 2004; van Venrooij 2015).

Rate of genre innovation in population. Second, similar to our creation of the rate of genre innovation on the Hot 100, we measure the rate of genre innovation at the population level to control for the possibility that genre innovation on the Hot 100 is associated with genre innovation in the population.

Aesthetic and semantic homogeneity in population. Third and fourth, like the aesthetic and semantic homogeneity measures created for the songs on the Hot 100, we measure the population-level homogeneity across our two dimensions by using all the songs available to control for the potential effect of the population-level feature landscape on genre innovation on the Hot 100.

Decade dummies. Finally, to capture heterogeneity in the ways music is produced and consumed due to time-related trends like technological advances and cultural evolution, we coded dummy variables for seven decades from the 1950s to the 2010s (de Laat 2014; see also Phillips and Owens 2004). There are numerous ways to account for temporal differences in the history of the music industry, but our primary analyses use these decade dummies because this is arguably the most common way—outside of a yearly approach, which we cannot use in light of the way our analyses are constructed—that music fans and followers of the charts conceive of eras (see, e.g., Whitburn 2008).

As noted, the history of the *Billboard* charts and the broader music industry can be divided in other ways. For example, the algorithm for determining songs' positions on the Hot 100 has changed several times over the chart's history (Anand and Peterson 2000), potentially influencing how new songs and genres appear. Moreover, the most popular format for personal music consumption (e.g., cassette, CD, MP3) has similarly changed over the 59 years covered by our data (RIAA 2023). Such changes, reflecting shifts from an industry that was driven by singles, then by albums, then again by singles—and that has gone through multiple format eras—could also influence which songs and genres rise to prominence. Therefore, in our robustness checks, we replace decade dummies with *Billboard policy dummies* and *popular format dummies* to see if our results are sensitive to different definitions of distinctive periods in the history of popular music (see Appendix B for details). In the online supplement, Tables S1a and S1b report descriptive statistics and correlation coefficients of the variables for Study 1, and Figure S1 shows the longitudinal trends of our control variables.

Analytic Strategy

Because our dependent variable, *rate of genre innovation*, is a proportion bounded between 0 and 1, ordinary least squares estimation may be biased and inconsistent. Instead, we use a fractional probit model to test our hypothesis (Papke and Wooldridge 1996). We use Stata's "fractional response regression" (*fracreg*) command with a probit link and robust standard errors, which is a standard modeling approach for handling fractional dependent variables (Adegbesan and Higgins 2011). Estimations using fractional logit models and OLS models provide similar results.

Our hypothesis is that the market-level distribution of semantics and aesthetics in time period $t - \alpha$ has an effect on genre innovation(s) in the market in time period t. We set α to 2: in the American popular music industry, recording contracts between artists and recording companies have normally contained a clause on the "traditional album production cycle" that requires the contracted artists to release a new album every 12 to 18 months (Negus 2011:42). Because this indicates a 1- to 1.5-year period to produce new songs, it is reasonable to assume producers—label executives and the musicians themselves—would spend roughly this amount of time to incorporate their read of the market-level feature distribution into the production and release of new songs. Also, as discussed in note 12 the lagging scheme is based on rough time frames in the music industry. Our use of a $t - 2$ timeframe for the independent variables and applying a 26-week sliding window to our analysis is illustrated in Figure S2 in the online supplement. This rolling window approach is common practice in management research, particularly when main variables are aggregated at the market level over consecutive time periods (e.g., Matusik and Fitza 2012; Rietveld and Ploog 2022; Rietveld, Schilling, and Bellavitis 2019).

We standardize our independent variables—semantic and aesthetic homogeneity—to better interpret the results of their interaction and reduce collinearity (Dalal and Zickar 2012). A test of the variance inflation factors returned no significant problems.

Results

Table 2 highlights the key results of our primary fractional probit models (Table S2 in

Table 2. Select Results from Fractional Probit Models for Rate of Genre Innovations (t) in Study 1

DV = Genre Innovation Rate in Year t	Model 3
Hypothesis 1: Mismatched Feature Distribution (t – 2)	
Aesthetic homogeneity × Semantic homogeneity	–.148***
	(.030)
Aesthetic homogeneity	.039
	(.034)
Semantic homogeneity	.078
	(.052)
Hot 100-Level Controls (t – 2)	
Genre innovation rate in Hot 100	–.882
	(2.254)
Market concentration (HHI)	–.160
	(.091)
Indie label ratio	–.142*
	(.059)
Decentralization ratio	.040
	(.056)
Primary genre homogeneity	.044
	(.064)
Sub-genre homogeneity	–.127*
	(.053)
Female ratio	–.076
	(.046)
Non-White ratio	–.140*
	(.068)
Group ratio	–.115*
	(.047)
Number of obs.	113
Log-likelihood	–5.952
χ^2	459.950

Note: Continuous variables are standardized except for the lagged DV. Robust standard errors are in parentheses. Full results appear in Table S2 in the online supplement.
$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$ (two-tailed tests).

the online supplement includes all variables, model specifications, and robustness checks). Our main interest is in the interaction term between aesthetic and semantic homogeneity in Model 3, which tests our first hypothesis. Although neither independent variable is individually a significant predictor of genre innovation, their interaction is strongly negative ($p < 0.001$), indicating that genre innovation is more likely to occur when there was previously homogeneity along one feature dimension and diversity along the other. This is consistent with our first hypothesis.

In terms of our control variables of interest, subgenre homogeneity shows a negative coefficient across these three models, suggesting the expansion of downstream genre scenes represented by increased subgenre diversity may subsequently promote the creation of new niches for a primary genre (Carroll and Swaminathan 2000; Mezias and Mezias 2000; van Venrooij 2015). Additionally, the consistently negative coefficient on the non-White ratio means a decline in the number of non-White musicians relative to their White counterparts in the market may precede genre innovation, suggesting genre innovation is more likely to emerge when the market was previously characterized by greater racial imbalance (i.e., when the market is highly

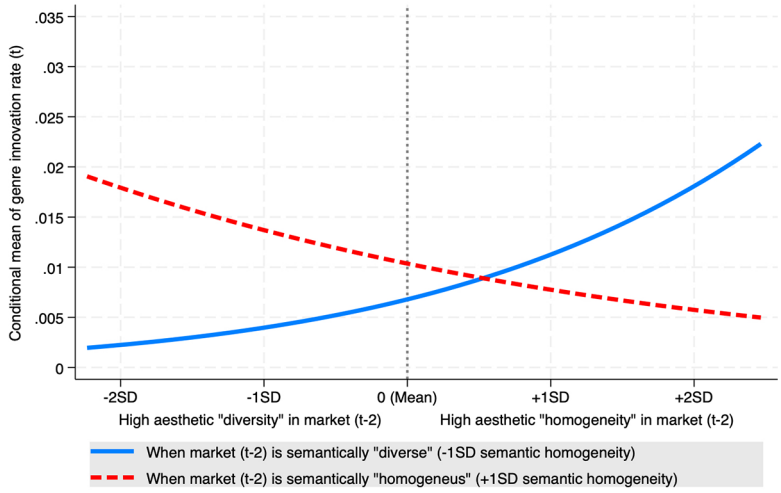

Figure 3. Predicted Marginal Effect of Aesthetic (Sonic) Homogeneity at Different Levels of Semantic (Lyrical) Homogeneity from Model 3 in Table 2

dominated by White musicians). Similarly, the negative coefficients on group ratio indicate that genre innovation is more likely when the prior mainstream chart was dominated by more solo artists than groups.

Figure 3 plots the marginal average effects on genre innovation based on Model 3. It shows that the conditional mean of the rate of genre innovation is low when aesthetic homogeneity and semantic homogeneity are *both* one or two standard deviations above or below the mean—that is, when they are both relatively homogeneous (the right side of the dashed line) or both relatively diverse (the left side of the solid line). By contrast, the rate of genre innovation is high when aesthetic homogeneity is high but semantic homogeneity is low, indicating a semantically diverse market (right side of the solid line). The inverse also appears to be true: a semantically homogeneous but aesthetically diverse market is similarly likely to see subsequent market-level innovation (the left side of the dashed line).

To see if one scenario is more likely than the other to lead to genre innovation, we visualize our data in Figure 4, which provides descriptive evidence that lyrical diversity and sonic homogeneity is more likely to precede innovation. Concretely, in the heatmaps, we divide the market landscapes into 25 conditions based on the intensity of aesthetic and semantic homogeneity (from very homogenous to very diverse across each dimension) and enter the rate (Figure 4) and count (Figure S3 in the online supplement) of genre innovation that follows each market condition in each cell. Across both the Hot 100 and the broader music industry, it is clear that genre innovation follows market situations where semantics are diverse but aesthetics are homogeneous, but not the other way around. We explore this dynamic in greater detail in Study 2.

Robustness Checks

We assess the robustness of the above results by running the following additional models in Table S2 in the online supplement. In Models 4 and 5, we replace our decade dummies with our two additional temporal measures, *Billboard* policy dummies and popular format dummies. In Models 6 through 10, the dependent variable is replaced by the rate of

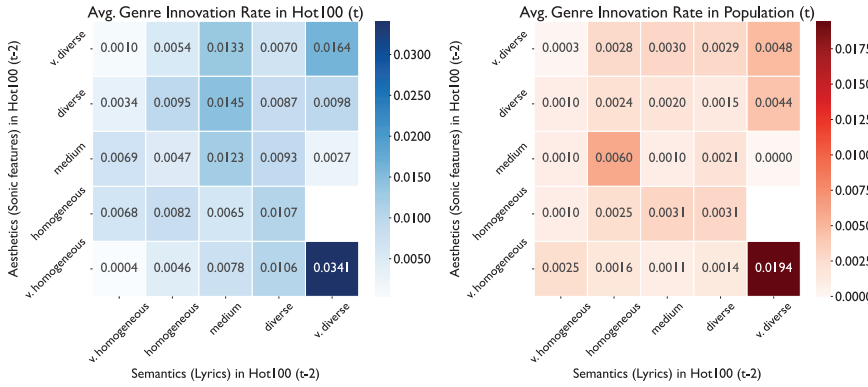

Figure 4. Heatmaps Showing Innovation Rates in the Hot 100 and Population across 25 Market Conditions Based on the Intensity of Aesthetic and Semantic Homogeneity
Note: The figures suggest that semantically diverse but aesthetically homogeneous markets anticipate genre innovations.

genre innovation measured at the *population* level while keeping the same set of covariates as our primary Hot 100 models. The negative coefficients of the interaction terms between aesthetic homogeneity and semantic homogeneity in Models 4 and 5 ($p < 0.001$) suggest our results are not sensitive to the different ways we account for time across our observation periods. Furthermore, goodness-of-fit measures suggest the decade dummies are best suited for the Hot 100 models, and the popular format dummies are best for the broader industry, although the differences between the two are small.

The consistent pattern of results when modeling our effects at the population level, although of smaller magnitude, indicates the feature landscape captured by the market information regime (i.e., *Billboard* Hot 100) may be relevant not only to genre innovation in the mainstream market but also to the entire industry. Conversely, the interaction effect of the feature distributions in the wider population are either not significant or show weak effects in the same direction as the Hot 100 feature distribution interaction. This further suggests that because the industry is so vast, the charts may provide a centralized source of information about market features and receptivity for many industry players regardless of their active participation in this niche. Finally, the results generally persist when we consider only population-level variables (see Table S3 in the online supplement). Appendix C provides two qualitative examples of our phenomenon.

STUDY 2: FEATURE DISTRIBUTIONS AND PRODUCT NOVELTY

Having established a connection between genre innovation and product feature distributions in the market for popular music, we move next to the question of whether novel songs are more likely to be produced and released under specific market conditions.

Dependent Variable

Song-level (composite) novelty at t. Song novelty is calculated via two steps. We first calculate a song's novelty score along each feature dimension, semantic and aesthetic, separately. Then, we standardize each and sum the two novelty scores to place them on the same scale. We call this summed novelty variable *composite song novelty.* We elaborate on each step.

First, using the song-level BERT-generated lyrical vectors, we compute each song's

semantic novelty by comparing its vector to those of the other songs on the charts over the prior five years. For example, the semantic novelty of a song released in 1983 captures how distant that song's lyrics are from the centroid of lyrical themes of the songs released between 1978 and 1982 (i.e., it does not include contemporaneous songs and we only account for each song once, even if it appeared on the chart for multiple weeks). We calculate this value by first creating an average 1,024-dimension vector from the comparison set, then subtracting the cosine similarity between the focal song's vector and that mean vector from one. The higher that novelty score, the more distinctive a song's lyrics, indicating its themes are different from the themes recently popular on the charts. We do not measure lyrical novelty for songs released during the first four years in our data, although those songs are used to calculate lyrical novelty for songs charting in 1963. Additionally, we ran the analysis with a 2-year window and a 10-year window for robustness checks; results are consistent.

Second, using the 11-attribute sonic vectors from Spotify that we use in Study 1, we apply the same approach to aesthetic novelty. We compute a focal song's *aesthetic novelty* relative to the songs on the charts over the prior five years. A song's aesthetic novelty is high (i.e., it "sounds" different) if a song has a highly distinctive sonic vector distant from its predecessors' sonic feature centroid. We then standardize and sum the two variables, *semantic novelty* and *aesthetic novelty*, yielding *composite song novelty*.

Independent Variables

Market-level feature configurations at t – 2. Our first argument suggested that market-level feature distributions can provide an opportunity for cultural producers—artists, record producers, and labels—to differentiate their output. Based on this theorizing, we use the (standardized) market-level homogeneity variables created for Study 1 and split them at their mean (i.e., 0).[13] Doing so creates four market conditions based on high or low semantic and aesthetic homogeneity: homogeneous semantics and aesthetics, diverse semantics and aesthetics, diverse semantics and homogeneous aesthetics, and homogeneous semantics and diverse aesthetics. We assign each song to one of the four conditions, based on the market two years prior to a song's release. As a result, 3,597 songs are assigned to markets where both features are homogenous, 5,866 songs to markets where both are diverse, 5,400 songs to markets with diverse semantics and homogenous aesthetics, and 6,711 to markets with homogenous semantics and diverse aesthetics.

Control Variables

For this set of analyses, we include several covariates constructed at four different levels: firm, musician, population, and *Billboard* Hot 100. First, we create six variables to account for the characteristics of the record labels responsible for producing new music. To do so, we referred to various archival sources—Wikipedia, labels' official websites, news outlets, and trade magazines. The first five of these are dummy variables intended to control for the many ways labels are founded and run, each of which plays a role in determining what gets produced.

First, we indicate whether a label is either independent (1) or major (0), and then whether it is decentralized (1, otherwise 0). Next, in line with prior literature that suggests collaborative organizational structures can enhance innovative performance (Walter, Auer, and Ritter 2006; Zhou and Li 2008), we delineate labels based on their adoption of one of the following organizational forms—spin-offs, joint ventures, or partnerships. We assign a 1 if a label falls within one of these categories, otherwise 0. Given the variations in the degree of creative control and artistic autonomy offered to musicians themselves that can affect creativity in the production of cultural products (Godart et al. 2020), we also indicate whether a label was founded by a performer (1, otherwise 0). Additionally, we

introduce a variable that indicates whether a label was founded by a producer (1 if yes, 0 if no), acknowledging the distinct role that producers may play in shaping the music creation process. The last firm-level variable is firm age, a continuous variable that reflects organizational maturity, represented numerically by the number of years since the firm's establishment.

Second, as in Study 1, we control for factors pertaining to the musicians and bands themselves that may influence the music they can produce and its novelty. We include a series of dummies to denote the gender and race of the creator(s) of each song, and whether they are a solo artist or a group. This amounts to five dummy variables: *gender (female)* for solo artists and groups in which female members outnumber male members; *gender (mixed)*, for groups with equal male and female representation; *race (non-White)* for non-White solo artists and groups with a majority of non-White members; *race (mixed)* for groups with equal numbers of White and non-White members; and *group* for distinguishing between solo artists and bands. *Gender (male)*, *race (White)*, and *solo artist* are the omitted categories as reference groups. At the musician level, we also include a control, *first entry to Hot 100*, for whether a song is an artist's or band's first song on the charts (de Laat 2014; Dowd 2004), and, in light of the importance of lyrics in our analyses, an indicator for *instrumental songs*.

Third, we again control for population-level variables: *release density in population, genre innovation rate in population, aesthetic homogeneity in population*, and *semantic homogeneity in population*. The operationalization of these variables is the same as in Study 1.

Fourth, we control for seven chart-level variables from Study 1 that likely influence the music that is produced and released: *genre innovation rate, firm concentration (HHI), primary genre homogeneity, subgenre homogeneity, female artist ratio, non-White artist ratio*, and *group ratio*.

Finally, we control for each song's primary genre and the era in which it was released. It is possible that, due to genre-specific characteristics, certain genres are more likely to create fine-grained distinctions that lead to a greater likelihood of putting out novel songs. Taking pop as our baseline genre, we include the following 14 parent genres as control variables: rock, folk-world-country, funk & soul, hip-hop, jazz, blues, electronic, Latin, children, stage & screen, reggae, classical, brass & military, and non-music. We also include controls for time periods: decade dummies in our main models and *Billboard* policy dummies and popular format dummies in robustness checks.

Analytic Strategy

To test Hypothesis 2, we ran pooled, cross-sectional OLS regressions for our dependent variable, *composite song novelty*, on the 19,442 songs for which we have complete data. The covariates at the firm or musician level were all measured in year t (i.e., the same year as DV is measured), and those at the population or Hot 100 level were all measured in year $t - 2$. We use heteroskedasticity-robust standard errors in the analysis. Tables S4a and S4b in the online supplement report descriptive statistics and correlation coefficients for the main variables, except for categorical variables. A test of the variance inflation factors reveals no significant collinearity concerns.

Results

We hypothesized that song novelty would be greater when the prior market was characterized by mismatched feature distributions (i.e., one feature was homogenous and the other was diverse) than when feature distributions were aligned. However, based on the results from Study 1 revealed in Figure 4, we further expect that a market characterized by lyrical diversity and sonic homogeneity will give rise to the most novel songs. Figure 5 depicts the standardized results of our primary model (Model 14 in Table S5 in the online supplement, which reports the full results of

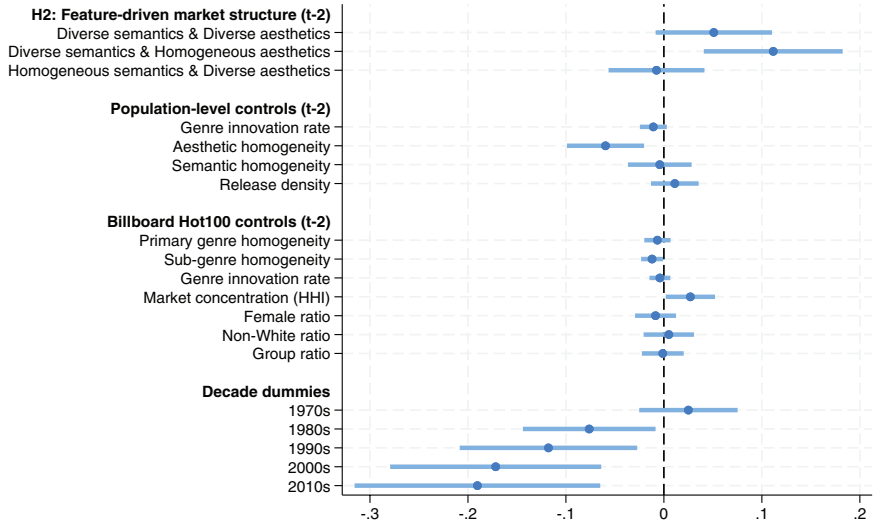

Figure 5. Selected Standardized Coefficients from OLS Model 14 Predicting Composite Song Novelty (at *t*) in Study 2
Note: Control variables measured at the firm level and the musician level; genre dummies are omitted due to space constraints.

OLS regressions testing this hypothesis). Our prediction is supported: we find a significant, positive relationship between a lyrically diverse but sonically homogeneous market at $t-2$ and higher song novelty at t, in comparison with the reference, a prior market where both feature dimensions are homogeneous. From these results, we not only find support for Hypothesis 2, but we find further evidence in support of Hypothesis 1: novelty is most likely in markets with a mix of diversity and homogeneity.

Robustness Checks

A similar pattern—greater subsequent novelty when the charts are semantically diverse but aesthetically homogenous—remains when we use our different demarcations of time (Models 15 and 16 in Table S5 in the online supplement). However, in these models, we also find that markets where both features are diverse give rise to subsequent novelty, although not to the same extent as the diverse-homogenous market. We explore this further at the population level, conducting similar regression models with the larger, more general dataset (Table S6a in the online supplement includes descriptive statistics for these analyses). Unlike the data collected at the Hot 100 level, for which we hand-coded some variables, the size of this population-level dataset—nearly 380,000 songs—is too large to hand-code. Hence, firm- and musician-level variables are omitted in these models. Nonetheless, the results show similar patterns consistent with the main findings: with our decade-level controls, we find the same positive results for lyrically diverse but sonically homogeneous markets, but with other measures of time, markets where both feature dimensions are diverse are also correlated with greater song novelty (for full results, see Table S7 in the online supplement).

STUDY 3: FEATURE DIFFERENTIATION, GENRE INNOVATION, AND GENRE EVOLUTION

In our final study, we explore the relationship between individual products' feature differentiation and genre innovation or evolution.

Whereas in Study 2, we created a composite novelty measure based on two product-level features—the sum of semantic and aesthetic novelty—here we keep the two dimensions separate, creating a novelty variable for each to test Hypotheses 3a and 3b.

Dependent Variables

Genre innovation (song-level). We created a dummy variable, *genre innovation*, to indicate songs that meet our criteria for representing a genre innovation. A song is coded 1 if it is categorized in a genre (or genre combination) that appeared on the charts for the first time within the 52 weeks prior to that song's debut, 0 otherwise.

Genre evolution (song-level). We created another dummy variable, *genre evolution*, to indicate songs that meet our criteria for representing genre evolution. A song is coded 1 if it is categorized in a *subgenre* (or combination of subgenres) that appeared on the charts for the first time within the 52 weeks prior to that song's debut, 0 otherwise.

Independent Variables

Our two independent variables are the two components of the *(composite) song novelty* variable from Study 2. We are therefore assessing the independent effect of each song's *semantic novelty* and *aesthetic novelty*, as well as their interaction, on *genre innovation* and *genre evolution*.

Control Variables

We include several control variables from Study 2: *independent firm dummy; decentralized major dummy; spin-off, joint venture, or partnership dummy; founded by performer dummy; founded by producer dummy; firm age; first entry to Hot 100 dummy; gender dummies; race dummies; group dummy; instrumental song dummy*, and *parent genres dummies*.[14] In addition, we include the 11 *sonic attributes* to control for potential effects of the individual sonic qualities of a song, as well as *artist tenure on the Hot 100* to account for artist incumbency (de Laat 2014; Dowd 2004; Lopes 1992). *Year dummies* are included in all models to account for time-specific effects on our dependent variables.

Analytic Strategy

To test Hypotheses 3a and 3b, which have binary outcome variables, we ran a series of binary logit regression analyses. We use heteroskedasticity-robust standard errors in both analyses. Tables S8a and S8b in the online supplement report descriptive statistics and correlation coefficients for the main variables in Study 3. Our independent variables—semantic novelty and aesthetic novelty—are log-transformed to correct for their right-skewed distribution and then standardized to make the coefficients and their interaction terms more interpretable. A test of the variance inflation factors reveals no significant collinearity concerns.

Results

Table 3 reports the key results of logistic regressions modeling the likelihood of genre *innovation* and genre *evolution* (fully specified models and results are in Tables S9 and S10 in the online supplement). We include both independent variables and their interaction in Model 24 and find that, in combination, the two measures of novelty are a powerful positive indicator of genre innovation. When a song is both semantically and aesthetically novel, it is more likely to signal the arrival of a new genre or genre combination. To better understand the interaction effect of the two novelty variables, we include a margins plot in Figure 6. The effect of a song's semantic novelty on the likelihood of that song signifying a new genre increases substantially when its aesthetic novelty is also high. This is reflected by the solid line in the figure. These results are supportive of Hypothesis 3a, which suggests innovation is the result of products that are highly novel

Table 3. Select Results from Logit Models for Genre Innovation and Genre Evolution in Study 3

Variable	Model 24 (H3a) (DV = Genre innovation)	Model 29 (H3b) (DV = Genre evolution)
Hypotheses 3a and 3b		
Aesthetic novelty	.197	.209***
	(.160)	(.051)
Semantic novelty	.143*	.116***
	(.072)	(.025)
Aesthetic novelty × Semantic novelty	.207***	−.002
	(.062)	(.022)
Musician-level controls		
First entry to Hot 100 dummy	.448*	.347***
	(.196)	(.059)
Gender dummy (female)	.366*	−.010
	(.184)	(.063)
Gender dummy (mixed)	.097	−.067
	(.309)	(.093)
Race dummy (non-White)	.601	−.161
	(.338)	(.102)
Race dummy (mixed)	.449*	−.174*
	(.195)	(.076)
Group dummy	−.341	.116*
	(.200)	(.058)
Artist tenure on Hot 100	−.019	−.043
	(.019)	(.006)
Constant	−4.571***	−3.396***
	(1.136)	(.398)
Year dummies	Yes	Yes
Sonic feature variables	Yes	Yes
Parent genre dummies	No	Yes
Pseudo R^2	.071	.112
AIC	2034.558	12762.248
BIC	2726.317	13629.432
Log-likelihood	−928.279	−6271.124
χ^2	242.965	1414.760
N of obs.	17,546	19,604

Note: Robust standard errors are in parentheses. Full results are in Tables S9 and S10 in the online supplement.
$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$ (two-tailed tests).

across both feature dimensions. Conversely, when semantic novelty is low (dashed line), the aesthetic novelty of a song has no influence on whether a song is likely to indicate the development of a new genre.

Results also suggest that "new" artists are more likely to be innovative: artists' first songs on the charts are more likely to be the progenitors of new genres than are songs from more established charting artists. As far as artist demographics, we find that female artists are more likely than male artists to produce songs that become genre innovations, as are racially mixed groups (compared to White groups and solo artists). Across all models in Table S9 in the online supplement, the effects for the musician-level covariates remain consistently significant; firm-level controls, however, appear to have little effect on new genres appearing. This may suggest that cultural innovation is more often due to small, cohesive project teams that are tightly

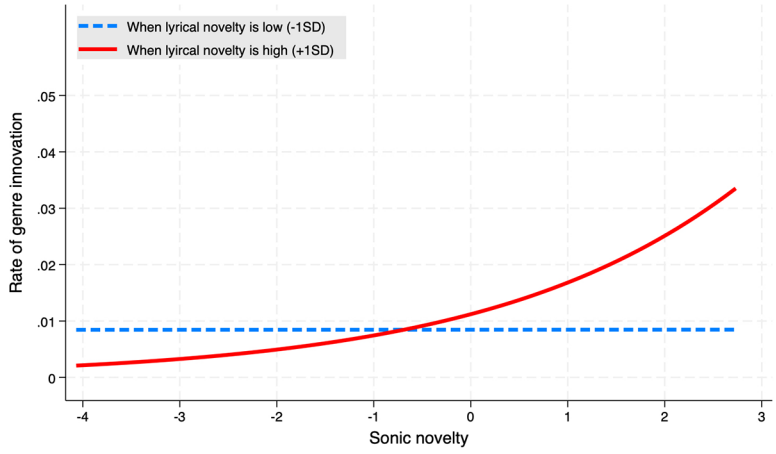

Figure 6. Predicted Likelihood of Genre Innovation as a Function of Semantic (Lyrical) Novelty and Aesthetic (Sonic) Novelty from Model 24 in Table 3

built around a focal musician or producer (Faulkner and Anderson 1987; Hsu 2006b), rather than the larger firms that help create and fund those teams.

In contrast to Hypothesis 3a, in Model 29 we are exploring the prediction that the main effects of semantic novelty and aesthetic novelty on genre *evolution* would be positive, and their interaction would have no effect. The results are consistent with this hypothesis: novelty in each dimension is associated with an increased likelihood of a song becoming the progenitor of a subgenre, but their interaction is not significant. As in the models for genre innovation, few of the firm-level variables appear to influence genre evolution.

However, in terms of musician-level controls, *first entry to Hot 100* is positive, again indicating a new-entrant effect on cultural evolution. Otherwise, the musician-level variables provide contrast between genre innovation and evolution. Female artists and racially mixed groups, each positively associated with genre *innovation*, appear to have a nonsignificant and negative relationship with genre *evolution*, respectively. Overall, gender and race dummies across the models for genre *innovation* and *evolution*, whether merely directional or statistically significant, suggest musicians from socially underrepresented backgrounds are more likely to innovate a new genre than are their overrepresented counterparts, while White and/or male musicians are more likely to evolve within an existing genre. Moreover, the coefficients on the group dummy across the models for genre *innovation* and *evolution* suggest solo artists disrupt and groups develop music genres, as in the field of science and technology (Wu, Wang, and Evans 2019).

Robustness Checks

To examine the robustness of our results for Hypotheses 3a and 3b, we ran additional logit models using the population-level data. Similar to the population dataset used for robustness checks in Study 2, we again do not have the same set of variables as in the Hot 100 models due to the size of the dataset—more than 400,000 observations—making it impossible to hand-code detailed information for covariates at the firm and musician level. However, the key dependent variables (i.e., *genre innovation in population* and *genre evolution in population*), the key independent variables (i.e., *aesthetic novelty*, *semantic novelty*, and their interaction), and all the other covariates we were able to include—individual *sonic feature*

variables, *year dummies*, and *parent genre dummies*—are measured in the same manner as the chart-level models in Tables S9 and S10 in the online supplement. The results of logit models using the population data are reported in Table S12 in the online supplement. Results of the population-level models further support Hypotheses 3a and 3b. As on the charts, songs high in both aesthetic and semantic novelty are more likely to become early songs in new genres (innovation), and high aesthetic or semantic novelty (but not both) is sufficient for a song to signal a genre evolution.

DISCUSSION

In this article, we add to an ongoing debate about the nature of market structure and its relationship to innovation (Alexander 1996; Bettis and Hitt 1995; de Laat 2014; Dowd 2004; Lopes 1992; Peterson and Berger 1975; Peterson and Kern 1996; Turner, Mitchell, and Bettis 2010), the persistence of which suggests that no one has satisfactorily resolved the question of which market conditions are most conducive to innovation. Focusing on the American popular music industry, we argue that the debate should include product features, both in aggregate and at the individual product level. Product features have been largely overlooked, as most research in this area focuses on the market actors that produce or consume the products, not the products themselves. We suggest the feature data across two key dimensions—aesthetic (sonic) and semantic (lyrical)—is captured by market information regimes that reveal consumers' tastes and producers' choices as well as opportunities. Drawing on intuition and conflicting expectations from institutional theory, resource partitioning theory, and work on innovation in creative industries, we theorize and find support for the claim that mismatched market configurations—characterized by feature diversity along one dimension and homogeneity along the other dimension—is most conducive to subsequent innovation.

To examine how product features structure a market and promote innovation, we conduct three studies, unpacking the potential mechanisms that connect the market-level feature configurations to producer output, which is the initial source of any innovative or evolutionary change in a field. We first analyze 113 market information regimes over 59 years, careful to include the factors that have thus far been regarded as key contributors to subsequent innovation: firm concentration in the market, the ratio of independent firms, the extent of organizational decentralization, and so on. In line with research suggesting that firm concentration and decentralized structure may have less to do with innovation than previously believed, or at least that the relationships vary over time (Burnett 1992b; de Laat 2014; Dowd 2004; Lopes 1992), we find little influence of each of these firm- and market-level factors in our primary analyses. However, when we consider the distribution of product features, we find they play a role in genre innovation: specifically, we discover a negative interaction between aesthetic and semantic homogeneity, suggesting a mismatched market configuration is most conducive to the emergence of innovation. Regarding the question of whether differentiation or conformity will drive subsequent innovation, the answer appears to be "both." Producers and consumers appear to be most inclined to welcome novelty in the form of new genres when the market provides a clear (likely aesthetic) dominant sound from which a new song can differentiate, as well as sufficient (likely semantic) diversity to signal openness to new themes. In addition to this novel finding, we also see the result as further evidence of the relevance of features to producers deciding what to create and put into the market (Rosa et al. 1999; White 1981).

Our second set of analyses refines our understanding of market-level feature configurations and their relationship to product novelty. We find that mismatched feature distributions in the market—specifically, diverse semantics coupled with homogeneous aesthetics—tend to precede periods characterized

by greater song novelty, more so than markets with aligned or inversely mismatched configurations (homogeneous semantics with diverse aesthetics). Interestingly, markets with dual diversity (diverse in both semantics and aesthetics) also appear to foster song novelty, but to a lesser extent. The fact that aesthetic diversity and homogeneity are both features of markets that precede greater song novelty suggests semantic diversity may be the more critical driver of song novelty.

This implies one of two phenomena about lyrics and lyrical themes. On one hand, it may be that when there is high lyrical diversity, we are picking up early signals of a fracturing of the existing genre structure of the market, just before it results in the appearance of an entirely new genre (see van Venrooij 2015). On the other hand, the relationship may be less about lyrical diversity leading to greater novelty and more about lyrical homogeneity acting as a barrier to subsequent novelty. Were this the case, lyrical homogeneity might be a sign of a conformity-inducing mechanism keeping more novel songs off the charts. In fact, this inference aligns with research showing that lyrical content often mirrors broader social dynamics: songs released during the COVID-19 pandemic possessed more negative emotional content (Putter, Krause, and North 2022); unemployment rates are correlated with heightened lyrical anger (Qiu et al. 2021); and in more challenging socioeconomic times, songs with profound, comforting, and romantic themes are more likely to gain prominence (Pettijohn and Sacco 2009) (see also Appendix D). If social dynamics drive similarity in lyrical themes, and songs that speak to those dynamics are rewarded, then subsequent (lyrical) novelty is likely to be diminished—a dynamic that would show up in a positive coefficient on semantic diversity. Exploring the details of the lyrical diversity–song novelty connection is a compelling area for future research. More generally, the Study 2 results suggest that an understanding of market conditions across both feature dimensions is important for artists and labels who seek popular success with novel songs.

Finally, in Study 3, we show that differing degrees of novelty along aesthetic and semantic features contribute to songs' likelihood of signaling genre innovation or evolution. Novelty in both feature dimensions signals a song's higher chance of being among the first in a new genre or genre combination, whereas novelty in one dimension is sufficient for a new subgenre or subgenre combination. That these results hold using a broader set of songs from the wider music industry suggests our variables are capturing something meaningful about the production and consumption of music, and not just at the level of the Hot 100 chart. They also suggest the social implications of cultural evolution (i.e., new genres, social categories, and market configurations) are reflected in the products themselves. While the emergence and diffusion of genres have historically been deeply intertwined with racial, gender, and organizational dynamics, as well as the strategic efforts of record labels, our findings add a layer to this understanding. They reveal that, alongside these social and industrial influences, the features of the products themselves also play a pivotal role in the genesis and shaping of new musical genres. Although likely unsurprising to anyone who creates or consumes music, our results underscore the need for a more multifaceted academic view of genre innovation and evolution.

In addition to the exploration of mechanisms potentially responsible for our findings, our study makes several other contributions to the literature. First, although we are not the first to suggest that product features play a role in the structuring and evolution of a cultural market, the addition of the fine-grained data we use represents a theoretical and an empirical contribution to the ongoing conversation. This type of data adds another dimension, alongside organizational concentration and the degree of decentralization, requiring consideration when conceptualizing market structure. Many markets will have more than two relevant features dimensions (more on this below), but any product-driven market can benefit from the inclusion of feature

data in their analyses. Cultural markets (e.g., wine, cuisine, television, fashion) may be the most obvious for collecting and including this kind of data, but the relevant features of technological products (see Khessina and Carroll 2008) and other consumer products (see Rosa et al. 1999) are easily found and have already been shown to be relevant to organizational research. More broadly, our study highlights the value of incorporating product features and their evolution into the organizational theorizing around innovation. Our findings underscore the idea that structural *and* feature-based forces should not be fully decoupled from the creative producers themselves, nor from their market.

Our study also contributes to the examination of gender, race, and power in the music industry. Taking a high-level view of our findings reveals that groups of underrepresented minorities (female artists and mixed-race artists/groups) are more likely to introduce new genres and genre combinations (Study 3), and that new genres are more likely to appear following periods of racial homogeneity (i.e., when the non-White ratio is lower; Study 1). Taken together, these results paint a picture of chart dominance by White male artists setting the stage for subsequent genre innovation, often coming from non-White-male artists and groups (cf. Dowd and Blyler 2002; Dowd, Liddle, and Blyler 2005). Yet mixed-race groups are less likely to introduce genre evolutions (cf. van Venrooij et al. 2022), perhaps reflecting more powerful record labels' moving in once a genre is established and keeping mixed-race groups from pushing internal boundaries (see Phillips and Owens 2004). Furthermore, the songs that become progenitors of new genres and subgenres tend to come from artists who have not previously appeared on the charts (cf. de Laat 2014; Lopes 1992; Peterson and Berger 1975). Despite findings that speak to these earlier explorations of who is "allowed" to innovate and under what conditions, we do not find evidence of indie labels being more likely to be sources of innovation, nor is the degree of organizational decentralization in the industry a contributor to greater or lower levels of innovation. Taken together, these results suggest innovation may have more to do with performers and their products than with the structure of the organizations behind them.

Our study also incorporates a variety of computational techniques for content analysis. In addition to the features themselves, which were calculated using audio analysis algorithms, we use a neural network text-encoding algorithm, enabling us to separately analyze the core feature dimensions for each product, and to do so at scale. Others have used NLP techniques on song lyrics (Berger and Packard 2018; Nie 2021), but to the best of our knowledge, we are the first to do so at this scale and scope, with the intent of better understanding how lyrical conformity and differentiation contribute to genre formation and evolution. The kind of data and analyses we use contribute to a growing trend in the social sciences, where the use of "digital trace" data—data created and made available as a function of interacting with digital tools and platforms—is being utilized to examine culture, broadly defined (e.g., Mohr et al. 2020), as well as music more specifically (Askin and Mauskapf 2017; Negro, Kovács, and Carroll 2022; Nie 2021). The explosion in available data capturing extensive metadata and algorithmically captured feature data, plus improvements in the accuracy of capturing lyrics and tools for analyzing them, means it is now much easier to understand trends and dynamics across cultural industries at scale, over time. Such techniques and computational advances allow us to propose new explanations that were previously not testable; this sets the stage for future research to dig deeper into the content and structure, and causes and consequences, of market dynamics.

As with all studies, ours is not without its limitations and boundary conditions. The first relates to the bias inherent in the use of *Billboard*'s Hot 100 chart. A focus on only the Hot 100 and the labels that put songs on this mainstream chart surely means we do not account for many genre and subgenre

innovations within the market for music—including many that alter the market's categorical structuring. However, we maintain that a new genre's appearance on the charts is a clear signal of an innovation, and we add support for our findings by showing similar patterns in a much broader set of songs. Thus, although we cannot wholly exclude the likelihood that we neglect earlier progenitors of new genres because of their relative lack of popularity or the new genre's comparatively niche appeal, we believe our findings are (1) robust beyond the limited chart and (2) can be tested in future research that can cover an even more comprehensive range of songs and producers that could not make it to the chart (cf. Negro et al. 2022).

A second limitation pertains to the feature dimensions we use and the generalizability of our findings. First, we do not include anything beyond basic demographic information in terms of the artists themselves. Musicians can play with novelty and creativity in ways we are unable to account for: personal image, visual presentation, and performance style, to name a few. There are currently no easy means for quantifying these, nor are they directly related to the features of the songs themselves. Qualitative ethnographies are necessary to test how these kinds of attributes contribute to assessments of novelty, but they fall outside the scope of this study.

Second, although the identity cues and themes in lyrics and the stylistic characteristics contained in songs' sonic fingerprints cover the cultural content contained by a piece of music, the same typology may not neatly apply to products outside cultural markets. Even within other cultural markets, when it comes to distinguishing feature dimensions, the field of commercial music is arguably among the simplest to analyze because the core features of popular songs can be easily divided into lyrics and sounds. However, in some markets (e.g., film), many more features exist, and in other markets (e.g., literature), perhaps one feature dimension may be enough. In others still, features can be very hard to measure due to intrinsic ambiguity (e.g., painting). Yet despite these concerns, advances in computational techniques are making data generated by the digital decomposition of cultural objects like paintings and images (e.g., Banerjee, Cole, and Ingram 2023), fashion (e.g., Godart and Galunic 2019), and even films (e.g., Harrison, Carlsen, and Škerlavaj 2019) into their constituent parts more accessible. Armed with these advances and the ability to access this kind of data for vast quantities of cultural objects, we believe future research can further test the role of feature dimensions on opportunities for innovating. Our working hypothesis is that some kind of mismatch between diversity and homogeneity across dimensions—regardless of how many there are—will be the most conducive to producers generating novel products and to markets welcoming more innovation.

A third limitation of our study is related to concerns about the dimensionality of the data needed to best capture the products in a given industry. Our own data are neither as symmetrical as we would like them to be, nor as nuanced. Our sonic feature data are sophisticated—they power the recommendation engines used by Spotify—but they still significantly flatten a song's complexity. Eleven summary features are inherently reductive, simplifying music in such a way that it is not possible to take the features we use in our analyses for a given song and reverse engineer that song. Moreover, although BERT provides a more detailed analysis of semantic content, it is similarly not feasible to take the lyrical vectors for a song and reconstruct the actual words sung. But what is lost in depth and precision is made up in scale: the inherent reductiveness of computational techniques used to capture cultural features allows us to compare tens of thousands (or more) of these objects. We believe the tradeoffs are worthwhile.

Finally, while our analytic and theoretical approach suggest a causal relationship, we do not test our results in a way that can definitively identify causality. Many factors beyond market structure surely play a role

in the rise of innovation, and we cannot categorically say we have ruled them out in favor of the role played by product features. Along these lines, our use of different ways of accounting for temporal changes in the music industry, including shifts in the methodology used to create the Hot 100, reveals varying results. The implications of these differing results as a function of how time is captured in the industry are outside the scope of this article, but we do believe the intersection of features and temporal dynamics in the music industry is compelling territory for future research. Furthermore, while we position the emergence of new genres as being situated between the decisions of artists and label executives and the preferences of consumers, we cannot speak specifically to the drivers of these phenomena—although we are sure both sides of the market are deeply involved. To that end, future research should further examine the questions we raise in two ways. First, we should qualitatively explore whether and how organizations, producers, and consumers process market information in the form of feature dimensions and incorporate it into their decisions about what to produce next or which songs they like. Second, we should take a more explicitly causal quantitative approach to see how producers respond when there are notable changes to the market information regime such that it is clear the configuration of features has changed.

Overall, our study makes important contributions to the theoretical and empirical analyses of the role of market structure in innovation. One of the early articles to explore these dynamics is titled "Cycles in Symbol Production: The Case of Popular Music" (Peterson and Berger 1975), and cycles remain an important aspect of this line of research (see Peterson 1997). We provided evidence that cycles include the features of the products in that market, as well. Innovation rates cycle alongside these feature distributions (see Figure S4 in the online supplement). We believe the framework we created invites scholars interested in the emergence of innovation to consider more seriously the product features of the markets and industries they study. We suspect that practitioners and consumers likely already do—now they have support for their intuition.

APPENDIX

Part A. Population-Level Data

We now offer an in-depth overview of the data collection processes for the population-level data used in our robustness checks. The data collection encompassed four steps. Initially, we developed our approach focusing on record labels with at least one song featured on the *Billboard* Hot 100 charts. Subsequently, we integrated their parent and subsidiary labels.

Step 1: Identifying commercial record labels. We began by compiling a list of record labels with at least one song on the *Billboard* Hot 100 charts and included any related labels affiliated with the initial set of labels. Concretely, in a manner similar to snowball sampling and using our primary dataset of Hot 100 songs from 1958 to 2016, which lists each track's label(s), we identified the parent and subsidiary labels via Discogs. For instance, Motown has approximately 40 sub-labels under its name, and Motown itself is under the Universal Music Group (UMG) umbrella. Thus, we incorporated all the Motown sub-labels, UMG, and all of UMG's U.S.-based subsidiaries. This iterative process continued until no further labels could be identified, resulting in a comprehensive list of 1,408 labels.

Step 2: Collecting releases by the identified labels. Subsequently, we gathered all releases (singles and albums) produced and distributed by the 1,408 labels within the U.S. market from 1958 to 2016 from Discogs. This process identified 345,462 releases (singles, EPs, and albums), and we subsequently collected the primary genre and subgenre information for each release.

Step 3: Collecting tracks by the identified releases and their audio features. We then searched each of the 345,462 releases on

Spotify via its API to identify tracks contained in each release. This search enabled us to collect data on over 1.53 million individual tracks. For every track found on Spotify, we gathered its distinct audio features.

Step 4: Collecting lyrics. In the next step, we searched for the lyrics for these 1.53 million tracks. Our search on Genius.com, one of the largest lyrics databases, yielded lyrics for 826,000 tracks.

From the data collected through these steps, we ultimately had complete information for approximately 374,000 tracks. Tracks were excluded due to reasons such as being derivative (e.g., remixes, live versions, instrumentals, radio edit); lacking lyrics or sonic features; falling under genres like non-music, children's, or military; being non-English; or discrepancies in release years between Discogs and Spotify (with a tolerance of +/– 1 year).

With this population-level dataset, we constructed several population-level variables, including *release density*, *genre innovation rate*, *aesthetic homogeneity*, and *semantic homogeneity* (for a comparison of our key variables in the Hot 100 and population-level datasets, see Figure S5 in the online supplement). We used these in the regressions for Study 1 to control for any influence of these population-level elements. We ran separate regressions using only these population-level variables to assess if a consistent pattern emerged beyond the Hot 100 samples (see Table S3 in the online supplement). To do so, we created a new dataset comprising 56 market information regimes (annually) and applied similar regression models. The count is 56 rather than the 113 observed in our primary analysis due to the population dataset's reliance on yearly timestamps, as opposed to weekly ones in our Hot 100 models.

Part B. Dummy Variables for Billboard Policy Change and Popular Release Formats

In Studies 1 and 2, we introduced *Billboard policy change* dummies and *popular release format* dummies as alternatives to our original decade dummies to test the robustness of our findings against different time demarcations. We provide descriptions of how we coded these two time-related dummy variables.

Billboard policy change dummies. Based on an official *Billboard* article (Trust 2019), we categorized the different eras of the *Billboard* charts into five distinct policy periods, each marked by significant shifts in the methodology for chart ranking. We coded these policy periods as time dummy variables to control for the effects of each era on the focal relationship of our interest in regression models. In the "Policy_1" era (1958 to 1991), *Billboard* relied on a formula that balanced airplay and singles sales. This period saw the dominance of vinyl records and cassette tapes. During this time, radio was the kingmaker of hits, and getting airtime was crucial for artists. Sales, predominantly in physical formats, coupled with radio airtime, were the most direct indicators of a song's popularity. During this time, the charts reflected commercial success and radio trends.

The "Policy_2" era (1991 to 1998) was a transformative time, as *Billboard* adopted SoundScan, which was more than just a technological shift, it represented a move toward greater transparency. Before digital collection, charts were based on manual sales reports from select stores, which could be inconsistent and biased. With SoundScan, *Billboard* could capture sales data from a broader range of retailers, leading to charts that more accurately represented consumer choices. "Policy_3" (1998 to 2005) was similarly transformative. No longer was a song's chart entry tied to its release format. With digital formats emerging, consumers began engaging with music through album deep cuts played in movies, television shows, and commercials. By dropping the single-release requirement, *Billboard* acknowledged various ways a track could gain popularity, making the charts more inclusive.

The "Policy_4" phase (2005 to 2012) was dominated by digital downloads. This

period saw the decline of CDs and the rise of platforms like iTunes, which revolutionized music ownership. Rather than buying a full album, consumers could now cherry-pick their favorite tracks. *Billboard*'s inclusion of these digital sales reflected this consumer behavior shift. Finally, "Policy_5" (2012 to 2016) reflects the streaming age. Music consumption was no longer just about ownership (physical or digital) but about access. With platforms like Spotify, listeners could access vast libraries of music for a monthly fee. The inclusion of video views recognized the visual medium's influence, especially with platforms like YouTube becoming vital for artists to showcase their work and connect with fans.

Popular release format dummies. Using the Recording Industry Association of America's (RIAA) official music revenue data (RIAA 2023), we classified the data into four distinctive periods based on the most dominant release format in the U.S. music industry: the LP/EP/vinyl period from the beginning of our observation through 1982, the cassette period from 1983 to 1990, the CD period from 1991 to 2010, and the digital period from 2011 onward. We determined the dominant release format for each period by the revenue generated by each format annually in USD, adjusted for inflation to 2022 U.S. Dollars.

Part C. Qualitative Examples of Our Study 1 Findings

In support of our findings, we provide two distinct historical examples of the kind of mismatched feature distributions and genre innovation that we discuss in the main text. First, our examination of the charts in 1969 to 1970 reveals substantial semantic diversity and aesthetic homogeneity, suggesting an opportunity conducive to genre innovation. Santana's "Black Magic Woman," released in late 1970, serves as an example of the beginning of a new genre combination that arose out of this specific structural configuration. The song is a cover, originally written and recorded by Peter Green and Fleetwood Mac, but the original is more of a standard blues song. It never charted in the United States. Santana's version took the original and blended blues, Latin, and rock, fusing genres in a way they had not previously been mixed. The song reached number 4 on the Hot 100. In so doing, the song and Santana created a path for artists such as Los Lobos, Miami Sound Machine, and Jon Secada in subsequent decades. Compared to other songs on the charts around that time, "Black Magic Woman" stands out for its high aesthetic and semantic novelty.

Second, in 1991, the Hot 100 could also be characterized as semantically diverse and aesthetically homogenous. In 1992, Prince's album "[Love Symbol]" placed four songs on the Hot 100 charts: "Sexy M.F.," "My Name Is Prince," "7," and "The Morning Papers." The first two, although not Prince's most popular songs, represented a different sound for his music, integrating electronic, hip-hop, and jazz, and challenging existing musical categories. The songs' distinct aesthetic and semantic novelty differentiated them from their contemporaries and set the stage for artists like US3 and Usher. Prince consistently blended genres throughout his career, and these tracks reflect how our identified "structures of opportunity" can facilitate transformations in the musical landscape.

Part D. Producer and Consumer Responses to Lyrics

We now present evidence of how producers and consumers may recognize and respond to semantic homogeneity or diversity, or the clustering of lyrical themes. First, we point to research that emphasizes how lyrics and their themes tend to homologize with concurrent socioeconomic sentiments. This evidence underscores the idea that musical production (especially lyrical crafting) and its reception transcend individual creativity and mirror broader societal dynamics. For example, Putter and colleagues (2022) find

that chart-topping songs from the UK and United States during the COVID-19 pandemic contained more negative emotional content, mirroring the global crisis's emotional toll. Similarly, Qiu and colleagues (2021) link rising unemployment rates with increased anger in song lyrics from the United States and Germany, suggesting that popular lyrics can reflect societal sentiments toward economic conditions. Zullow (1991) suggests that pessimistic undertones in lyrics could foreshadow shifts in media portrayals of global events and economic perceptions, even hinting at looming recessions. Further reinforcing this concept, Pettijohn and Sacco (2009) argue that during challenging times, songs with deep, comforting, and romantic themes become more popular, highlighting the connection between lyrics and broader societal sentiments.

Second, we searched for cases where musicians expressed awareness of lyrical trends and themes, especially those dominating the mainstream market. We found several interviews that reveal musicians' perceptions of popular lyrical themes. For example, Robert Plant, the lead vocalist and lyricist for Led Zeppelin, recalled writing "Achilles' Last Stand" in 1976: "There's something really splendid and otherworldly about trying to even touch those bigger ideas as a British rock group, to go past the idiom of singing about bars and chicks and all that crap, which unfortunately is the lingua franca of popular song" (Rogers 2017). ("Achilles' Last Stand" did not chart on the Hot 100.)

On the other end of the musical spectrum, Master P—an influential rapper, entrepreneur, and the founder of the label No Limit Records—showed a similar awareness of dominant lyrical themes during the 1990s to 2000s, largely influenced by bounce music. Bounce music was known for its lively and festive party anthems, often emphasizing themes of celebration, dance, and joyous escapism. Master P, however, recognized this pattern and felt the absence of any musical narrative: "You know, hearing the bounce music, it was a lot of party music, which was a good sound for New Orleans, but I felt like we were missing something, you know? Seeing so many people die young, seeing the crime, seeing the poverty, it just poured out of me. I was able to tell my story" (Wheeler and Bascuñán 2020).

More recently, Noel Gallagher, former lead guitarist and primary songwriter for Oasis, offered a critique on the modern music scene: "The charts for the first time in history are completely dominated by major label acts who have teams of songwriters and employ major-label thinking and it all sounds the same" (Khomami 2015). Gallagher's comment indicates a perceived lack of lyrical (and sonic) diversity in chart-topping songs, signaling his understanding of semantic patterns.

Although anecdotal, these interviews and comments provide some evidence in support of our assumption that cultural producers are not merely passive spectators of the lyrical landscape but engage in active observation and interpretation. Each of these could be viewed as musicians describing a clichéd caricature of the mainstream charts, but they speak to perceptions and a desire to differentiate. Furthermore, these musicians are acknowledging homogeneity more than diversity, perhaps suggestive of the cycle of feature diversity and innovation in the industry: sensing homogeneity leads artists to be more novel, which shows up in future charts, and paves the way for new genres to arise.

Acknowledgments

For their valuable ideas, comments, and support, we are grateful to Matt Bothner, Charlie Galunic, Martin Gargiulo, Frédéric Godart, Henrich Greve, Steve Mezias, Damon Phillips, and Henning Piezunka. Thanks are also due to the attendees of the 2022 Creative Industries Conference in Amsterdam. Finally, we are grateful to editor Lynn Selhat for her constructive feedback and editing prowess. The current and previous editorial teams at *ASR* provided valuable guidance and support, and three anonymous reviewers were among the most helpful we have ever encountered. We extend our sincere thanks to them all.

ORCID iDs

Khwan Kim https://orcid.org/0000-0003-3987-2829
Noah Askin https://orcid.org/0000-0003-3935-1484

Notes

1. The term "producer" has specific meaning in the context of music. The job requirements vary by individual and by project, but "record producers" are ultimately responsible, based on some combination of technical and creative duties, for making a recording project, typically an album, happen. For our purposes here, we use "producer" more generally to refer to the people who generate cultural output. In the context of the music industry, this means the record labels/executives, record producers, musicians, songwriters, and so on. We do our best to be explicit when referring to these specific roles versus the general production side of the music creation market.
2. Innovation and the appearance of new categories are mutually constitutive (Hargadon and Douglas 2001; Jones et al. 2012; Rao et al. 2003): innovations spawn new categories, which, in turn, help these products "latch on" in the marketplace.
3. The strategy and innovation literatures often distinguish between "radical" and "incremental" innovation (e.g., Abernathy and Clark 1985; Christensen and Rosenbloom 1995; Tushman and Anderson 1986). However, for our purposes, "incremental" feels inadequate for what amounts to a changing of the categorical structure of a market, even if such change may not be "radical." To that end, while we distinguish between degrees of innovation, we do not want to minimize the non-radical changes and instead refer to new subgenres and subgenre combinations as genre *evolution*.
4. Such a limited chart also means most produced and published music (to say nothing of music that was written but never published) is not included in our primary analyses. This motivated our collection of additional population-level data that captures a wider swath of the U.S. recording industry's output over the past 70 years. However, regarding genre emergence and the Hot 100, Lena and Peterson (2008:701) note that while most genres originate in what they call the "avant-garde" form, these are "quite small, having no more than a dozen participants who meet informally or regularly." That is, most of the recorded music that launches new genres is not heard by most people, nor picked up by the broader market. Yet well over half of these avant-garde genres ultimately gain the "hundreds of thousands" of fans needed to become the kind of "industry-based genres" we are counting as innovations. It is through this growth that real market shifts ultimately occur. Of the genres Lena and Peterson (2008:711) examine, 15 percent *originate* as industry-based genres, suggesting some genre-based innovation takes place directly in the realm of the Hot 100 chart. Either way, the appearance on the Hot 100 of a new genre or subgenre suggests the market has shifted in a meaningful way.
5. Whereas lyrics are semantic features, vocals would be considered aesthetic features. However, as our data do not capture vocal melodies, we label lyrics as exclusively semantic features.
6. We wish to thank an anonymous reviewer for highlighting the critical elements of this example.
7. These hypotheses may appear definitional (i.e., more novel songs are the progenitor of novel genres), but they serve a greater purpose than simply demonstrating that our feature data are capturing something substantively meaningful. First, these hypotheses form a critical piece of our Coleman boat model: they connect the micro-level phenomenon (more novel songs are produced in mismatched feature distribution markets) to a micro-level outcome (more novel songs are more likely to signal genre innovation and evolution). Second, as our population-level results support, the hypothesized dynamics are not limited to the Hot 100 charts but are reflective of more general, industry-wide dynamics.
8. Of the 25,417 unique songs that appeared on the charts during the period covered by our data, we were unable to find complete data (typically that meant missing lyrics) for 3,098 songs.
9. Of the 22,319 songs in our data, 71 (0.3 percent) are instrumental and therefore have no lyrics. These are primarily film soundtracks, such as the main theme from *Star Wars* (1977). We still calculate their lyrical novelty (i.e., they feature no semantics), but we also create a dummy variable "instrumental song," which for these songs is set to 1.
10. More information about the process of genre attribution and approval can be found on the Discogs site (https://support.discogs.com/hc/en-us/articles/360005055213-Database-Guidelines-9-Genres-Styles). Genre and subgenre attributions, especially for older music/releases, are made years or even decades after the songs were released, but the process required for submission and approval of these attributions likely keeps them generally accurate. Discogs was originally started as a means of cataloguing and selling electronic and (later) other obscure records. Because of the user community created by targeting this niche, the commitment to accuracy is high.
11. In Study 3 we differentiate between innovation and evolution by looking at new genres and new subgenres, respectively. Here in Study 1, we focus only on genre-level innovation. This is due largely to the fact that the mechanisms and market dynamics at work are likely different for these two types of innovation, and the article is already sufficiently complicated. See Figure S4 in the online supplement for levels of market homogeneity across our two feature dimensions and their alignment with counts of genre innovation and evolution.
12. Our results are robust to different observation periods for our variable construction (26 weeks with

a 13-week sliding window and 104 weeks with a 52-week sliding window). We ultimately settled on a one-year (52-week) observation period with 26-week overlap between periods because some time is required to get a sense of "what is popular." We also wanted to acknowledge the fact that in cultural markets there are rarely clear demarcations between trend periods. Furthermore, for much of the period we are covering, songs took some time to be written, recorded, released, and picked up by radio. Although this has shortened recently, based on our knowledge of the industry, we believe our formulation is sufficient to capture trends. Results of our robustness checks are available upon request.

13. For robustness checks, we applied more conservative classification standards, and they return substantively similar results. Using ±1 standard deviation as our criteria, we categorized the four market conditions based on whether the market is homogeneous (i.e., its semantic or aesthetic homogeneity is higher than +1 standard deviation) and whether it is diverse (i.e., its semantic or aesthetic homogeneity is lower than –1 standard deviation). We used quartiles in a similar manner. Both show results consistent with what we present in this article.
14. Parent genre dummies are used only in models for genre evolution.

References

Abernathy, William J., and Kim B. Clark. 1985. "Innovation: Mapping the Winds of Creative Destruction." *Research Policy* 14(1):3–22.

Adegbesan, J. Adetunji, and Matthew J. Higgins. 2011. "The Intra-alliance Division of Value Created through Collaboration." *Strategic Management Journal* 32(2):187–211.

Alexander, Peter J. 1996. "Entropy and Popular Culture: Product Diversity in the Popular Music Recording Industry." *American Sociological Review* 61(1):171–74.

Alridge, Derrick P. 2005. "From Civil Rights to Hip Hop: Toward a Nexus of Ideas." *Journal of African American History* 90(3):226–52.

Alridge, Derrick P., and James B. Stewart. 2005. "Introduction: Hip Hop in History: Past, Present, and Future." *Journal of African American History* 90(3):190–95.

Amabile, Teresa M. 1988. "A Model of Creativity and Innovation in Organizations." *Research in Organizational Behavior* 10(1):123–67.

Anand, Narasimhan, and Grégoire Croidieu. 2015. "Niches, Genres, and Classifications in the Creative Industries." Pp. 327–47 in *The Oxford Handbook of Creative Industries*, edited by C. Jones, M. Lorenzen, and J. Sapsed. Oxford, UK: Oxford Academic Press.

Anand, N., and Richard A. Peterson. 2000. "When Market Information Constitutes Fields: Sensemaking of Markets in the Commercial Music Industry." *Organization Science* 11(3):270–84.

Anderson, Neil, Kristina Potočnik, and Jing Zhou. 2014. "Innovation and Creativity in Organizations." *Journal of Management* 40(5):1297–333.

Askin, Noah, and Michael Mauskapf. 2017. "What Makes Popular Culture Popular? Product Features and Optimal Differentiation in Music." *American Sociological Review* 82(5):910–44.

Askin, Noah, and Joeri Mol. 2018. "Institutionalizing Authenticity in the Digitized World of Music." Pp. 159–202 in *Frontiers of Creative Industries: Exploring Structural and Categorical Dynamics*, Vol. 55, *Research in the Sociology of Organizations*, editedy by C. Jones and M. Maoret. Bingley, UK: Emerald Publishing Limited.

Banerjee, Mitali, Benjamin M. Cole, and Paul Ingram. 2023. "'Distinctive from What? And for Whom?' Deep Learning-Based Product Distinctiveness, Social Structure, and Third-Party Certifications." *Academy of Management Journal* 66(4):1016–41.

Baum, Joel A. C., and Stephen J. Mezias. 1992. "Localized Competition and Organizational Failure in the Manhattan Hotel Industry, 1898–1990." *Administrative Science Quarterly* 37(4):580–604.

Becker, Howard Saul. 1982. *Art Worlds*. Berkeley: University of California Press.

Berg, Justin M. 2022. "One-Hit Wonders versus Hit Makers: Sustaining Success in Creative Industries." *Administrative Science Quarterly* 67(3):630–73.

Berger, Jonah, and Grant Packard. 2018. "Are Atypical Things More Popular?" *Psychological Science* 29(7):1178–84.

Bettis, Richard A., and Michael A. Hitt. 1995. "The New Competitive Landscape." *Strategic Management Journal* 16(S1):7–19.

Beverland, Michael. 2006. "The 'Real Thing': Branding Authenticity in the Luxury Wine Trade." *Journal of Business Research* 59(2):251–58.

Bielby, William T., and Denise D. Bielby. 1994. "'All Hits Are Flukes': Institutionalized Decision Making and the Rhetoric of Network Prime-Time Program Development." *American Journal of Sociology* 99(5):1287–313.

Boone, Christophe, Carolyn H. Declerck, Hayagreeva Rao, and Kristin Van den Buys. 2012. "Out of Tune: The Rise and Fall of Modernistic Music in Brussels, 1919–1939." *Poetics* 40(1):44–66.

Bourdieu, Pierre. 1983. "The Field of Cultural Production, or: The Economic World Reversed." *Poetics* 12(4):311–56.

Bourdieu, Pierre. 1984. *Distinction: A Social Critique of the Judgement of Taste*. Cambridge, MA: Harvard University Press.

Burnett, Robert. 1992a. "Dressed for Success: Sweden from Abba to Roxette." *Popular Music* 11(2):141–50.

Burnett, Robert. 1992b. "The Implications of Ownership Changes on Concentration and Diversity in the Phonogram Industry." *Communication Research* 19(6):749–69.

Burnett, Robert. 1993. "The Popular Music Industry in Transition." *Popular Music & Society* 17(1):87–114.

Burnett, Robert. 1996. *The Global Jukebox: The International Music Industry*. London, UK: Psychology Press.

Cao, Likun, Ziwen Chen, and James Evans. 2022. "Destructive Creation, Creative Destruction, and the Paradox of Innovation Science." *Sociology Compass* 16(11) (https://doi.org/10.1111/soc4.13043).

Carey, James T. 1969. "Changing Courtship Patterns in the Popular Song." *American Journal of Sociology* 74(6):720–31.

Carroll, Glenn R. 1985. "Concentration and Specialization: Dynamics of Niche Width in Populations of Organizations." *American Journal of Sociology* 90(6):1262–83.

Carroll, Glenn R., and Anand Swaminathan. 2000. "Why the Microbrewery Movement? Organizational Dynamics of Resource Partitioning in the US Brewing Industry." *American Journal of Sociology* 106(3):715–62.

Carroll, Glenn R., and Dennis Ray Wheaton. 2009. "The Organizational Construction of Authenticity: An Examination of Contemporary Food and Dining in the U.S." *Research in Organizational Behavior* 29:255–82.

Castañer, Xavier, and Lorenzo Campos. 2002. "The Determinants of Artistic Innovation: Bringing in the Role of Organizations." *Journal of Cultural Economics* 26(1):29–52.

Caves, Richard E. 2000. *Creative Industries: Contracts between Art and Commerce*. Cambridge, MA: Harvard University Press.

Chang, Yu-Yu, and Ming-Huei Chen. 2020. "Creative Entrepreneurs' Creativity, Opportunity Recognition, and Career Success: Is Resource Availability a Double-Edged Sword?" *European Management Journal* 38(5):750–62.

Christensen, Clayton M., and Richard S. Rosenbloom. 1995. "Explaining the Attacker's Advantage: Technological Paradigms, Organizational Dynamics, and the Value Network." *Research Policy* 24(2):233–57.

Coase, Ronald Harry. 1979. "Payola in Radio and Television Broadcasting." *Journal of Law and Economics* 22(2):269–328.

Coleman, James S. 1986. "Social Theory, Social Research, and a Theory of Action." *American Journal of Sociology* 91(6):1309–35.

Coleman, James S. 1994. *Foundations of Social Theory*. Cambridge, MA: Harvard University Press.

Conti, Raffaele, Olenka Kacperczyk, and Giovanni Valentini. 2021. "Institutional Protection of Minority Employees and Entrepreneurship: Evidence from the LGBT Employment Non-discrimination Acts." *Strategic Management Journal* 43(4):758–91 (https://doi.org/10.1002/smj.3340).

Croidieu, Grégoire, Charles-Clemens Rüling, and Amélie Boutinot. 2016. "How Do Creative Genres Emerge? The Case of the Australian Wine Industry." *Journal of Business Research* 69(7):2334–42.

Dalal, Dev K., and Michael J. Zickar. 2012. "Some Common Myths about Centering Predictor Variables in Moderated Multiple Regression and Polynomial Regression." *Organizational Research Methods* 15(3):339–62.

Deephouse, David L. 1999. "To Be Different, or to Be the Same? It's a Question (and Theory) of Strategic Balance." *Strategic Management Journal* 20(2):147–66.

de Laat, Kim. 2014. "Innovation and Diversity Redux: Analyzing Musical Form and Content in the American Recording Industry, 1990–2009." *Sociological Forum* 29(3):673–97.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." ArXiv Preprint (https://doi.org/10.48550/arXiv.1810.04805).

DiMaggio, Paul. 1977. "Market Structure, the Creative Process, and Popular Culture: Toward an Organizational Reinterpretation of Mass Culture Theory." *Journal of Popular Culture* 11(2):436–52.

DiMaggio, Paul. 1987. "Classification in Art." *American Sociological Review* 52(4):440–55.

DiMaggio, Paul J., and Walter W. Powell. 1983. "The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields." *American Sociological Review* 48(2):147–60.

Dowd, Timothy J. 2000. "Musical Diversity and the US Mainstream Recording Market, 1955–1990." *Rassegna Italiana Di Sociologia* 41(2):223–63.

Dowd, Timothy J. 2002. "The Musical Structure and Social Context of Number One Songs, 1955 to 1988: An Exploratory Analysis." In *Vocabularies of Public Life: Empirical Essays in Symbolic Structure*, edited by R. Wuthnow. London, UK: Routledge.

Dowd, Timothy J. 2003. "Structural Power and the Construction of Markets: The Case of Rhythm and Blues." Pp. 147–201 in *Comparative Studies of Culture and Power*, Vol. 21, *Comparative Social Research*, edited by F. Engelstad. Bingley, UK: Emerald Group Publishing.

Dowd, Timothy J. 2004. "Concentration and Diversity Revisited: Production Logics and the U.S. Mainstream Recording Market, 1940–1990." *Social Forces* 82(4):1411–55.

Dowd, Timothy J., and Maureen Blyler. 2002. "Charting Race: The Success of Black Performers in the Mainstream Recording Market, 1940 to 1990." *Poetics* 30(1–2):87–110.

Dowd, Timothy J., Kathleen Liddle, and Maureen Blyler. 2005. "Charting Gender: The Success of Female Acts in the US Mainstream Recording Market, 1940–1990." Pp. 81–123 in *Transformation in Cultural Industries*. Bingley, UK: Emerald Group Publishing Limited.

Edelman, Linda, and Helena Yli-Renko. 2010. "The Impact of Environment and Entrepreneurial Perceptions on Venture-Creation Efforts: Bridging the Discovery and Creation Views of Entrepreneurship." *Entrepreneurship Theory and Practice* 34(5):833–56.

Ennis, Philip H. 1992. *The Seventh Stream: The Emergence of Rocknroll in American Popular Music*. Middletown, CT: Wesleyan University Press.

Faulkner, Robert R., and Andy B. Anderson. 1987. "Short-Term Projects and Emergent Careers: Evidence from Hollywood." *American Journal of Sociology* 92(4):879–909.

Fleming, Lee. 2001. "Recombinant Uncertainty in Technological Search." *Management Science* 47(1):117–32.

Frith, Simon. 1996. *Performing Rites: On the Value of Popular Music*. Cambridge, MA: Harvard University Press.

Godart, Frédéric C., and Charles Galunic. 2019. "Explaining the Popularity of Cultural Elements: Networks, Culture, and the Structural Embeddedness of High Fashion Trends." *Organization Science* 30(1):151–68.

Godart, Frédéric, Sorah Seong, and Damon J. Phillips. 2020. "The Sociology of Creativity: Elements, Structures, and Audiences." *Annual Review of Sociology* 46:489–510.

Greenwood, Royston, and Roy Suddaby. 2006. "Institutional Entrepreneurship in Mature Fields: The Big Five Accounting Firms." *Academy of Management Journal* 49(1):27–48.

Guimerà, Roger, Brian Uzzi, Jarrett Spiro, and Luís A. Nunes Amaral. 2005. "Team Assembly Mechanisms Determine Collaboration Network Structure and Team Performance." *Science* 308(5722):697–702.

Hargadon, Andrew B., and Yellowlees Douglas. 2001. "When Innovations Meet Institutions: Edison and the Design of the Electric Light." *Administrative Science Quarterly* 46(3):476–501.

Hargadon, Andrew, and Robert I. Sutton. 1997. "Technology Brokering and Innovation in a Product Development Firm." *Administrative Science Quarterly* 42(4):716–49.

Harrison, David A., and Katherine J. Klein. 2007. "What's the Difference? Diversity Constructs as Separation, Variety, or Disparity in Organizations." *Academy of Management Review* 32(4):1199–228.

Harrison, Spencer, Arne Carlsen, and Miha Škerlavaj. 2019. "Marvel's Blockbuster Machine." *Harvard Business Review* 97(4):136–45.

Hirsch, Paul M. 1971. "Sociological Approaches to the Pop Music Phenomenon." *American Behavioral Scientist* 14(3):371–88.

Hirsch, Paul M. 1972. "Processing Fads and Fashions: An Organization-Set Analysis of Cultural Industry Systems." *American Journal of Sociology* 77(4):639–59.

Hsu, Greta. 2006a. "Evaluative Schemas and the Attention of Critics in the US Film Industry." *Industrial and Corporate Change* 15(3):467–96.

Hsu, Greta. 2006b. "Jacks of All Trades and Masters of None: Audiences' Reactions to Spanning Genres in Feature Film Production." *Administrative Science Quarterly* 51(3):420–50.

Hsu, Greta, and Michael T. Hannan. 2005. "Identities, Genres, and Organizational Forms." *Organization Science* 16(5):474–90.

Inglehart, Ronald, and Wayne E. Baker. 2000. "Modernization, Cultural Change, and the Persistence of Traditional Values." *American Sociological Review* 65(1):19–51.

Jones, Candace, Massimo Maoret, Felipe G. Massa, and Silviya Svejenova. 2012. "Rebels with a Cause: Formation, Contestation, and Expansion of the de Novo Category 'Modern Architecture,' 1870–1975." *Organization Science* 23(6):1523–45.

Jones, Candace, Silviya Svejenova, Jesper Strandgaard Pedersen, and Barbara Townley. 2016. "Misfits, Mavericks and Mainstreams: Drivers of Innovation in the Creative Industries." *Organization Studies* 37(6):751–68.

Khessina, Olga M., and Glenn R. Carroll. 2008. "Product Demography of de Novo and de Alio Firms in the Optical Disk Drive Industry, 1983–1999." *Organization Science* 19(1):25–38.

Khomami, Nadia. 2015. "Noel Gallagher: 'I've given My Daughter a List of People to Insult at Ed Sheeran's Wembley Gig.'" *Www.Nme.Com*, February 4.

Lamont, Michèle, and Virág Molnár. 2002. "The Study of Boundaries in the Social Sciences." *Annual Review of Sociology* 28(1):167–95.

Lampel, Joseph, Theresa Lant, and Jamal Shamsie. 2000. "Balancing Act: Learning from Organizing Practices in Cultural Industries." *Organization Science* 11(3):263–69.

Laybourn, Wendy M. 2018. "The Cost of Being 'Real': Black Authenticity, Colourism, and Billboard Rap Chart Rankings." *Ethnic and Racial Studies* 41(11):2085–2103.

Leahey, Erin, Christine M. Beckman, and Taryn L. Stanko. 2017. "Prominent but Less Productive: The Impact of Interdisciplinarity on Scientists' Research." *Administrative Science Quarterly* 62(1):105–39.

Lee, Steve S. 2004. "Predicting Cultural Output Diversity in the Radio Industry, 1989–2002." *Poetics* 32(3–4):325–42.

Lena, Jennifer C. 2006. "Social Context and Musical Content of Rap Music, 1979–1995." *Social Forces* 85(1):479–95.

Lena, Jennifer C. 2012. *Banding Together*. Princeton, NJ: Princeton University Press.

Lena, Jennifer C., and Richard A. Peterson. 2008. "Classification as Culture: Types and Trajectories of Music Genres." *American Sociological Review* 73(5):697–718.

Lopes, Paul D. 1992. "Innovation and Diversity in the Popular Music Industry, 1969 to 1990." *American Sociological Review* 57(1):56–71.

Matusik, Sharon F., and Markus A. Fitza. 2012. "Diversification in the Venture Capital Industry: Leveraging Knowledge under Uncertainty." *Strategic Management Journal* 33(4):407–26.

McLeod, Kembrew. 2001. "Genres, Subgenres, Sub-subgenres and More: Musical and Social Differentiation within Electronic/Dance Music Communities." *Journal of Popular Music Studies* 13(1):59–75.

Meyer, John W., and Brian Rowan. 1977. "Institutionalized Organizations: Formal Structure as Myth and Ceremony." *American Journal of Sociology* 83(2):340–63.

Mezias, John M., and Stephen J. Mezias. 2000. "Resource Partitioning, the Founding of Specialist Firms, and Innovation: The American Feature Film Industry, 1912–1929." *Organization Science* 11(3):306–22.

Mohr, John W., Christopher A. Bail, Margaret Frye, Jennifer C. Lena, Omar Lizardo, Terence E. McDonnell, Ann Mische, Iddo Tavory, and Frederick F. Wherry. 2020. *Measuring Culture*. New York: Columbia University Press.

Negro, Giacomo, Balázs Kovács, and Glenn R. Carroll. 2022. "What's Next? Artists' Music after Grammy Awards." *American Sociological Review* 87(4):644–74.

Negus, Keith. 1999. "The Music Business and Rap: Between the Street and the Executive Suite." *Cultural Studies* 13(3):488–508.

Negus, Keith. 2011. *Producing Pop: Culture and Conflict in the Popular Music Industry*. London, UK: Edward Arnold.

Nie, Ke. 2021. "Disperse and Preserve the Perverse: Computing How Hip-Hop Censorship Changed Popular Music Genres in China." *Poetics* 88 (https://doi.org/10.1016/j.poetic.2021.101590).

Papke, Leslie E., and Jeffrey M. Wooldridge. 1996. "Econometric Methods for Fractional Response Variables with an Application to 401(k) Plan Participation Rates." *Journal of Applied Econometrics* 11(6):619–32.

Peterson, Richard A. 1990. "Why 1955? Explaining the Advent of Rock Music." *Popular Music* 9(1):97–116.

Peterson, Richard A. 1997. *Creating Country Music: Fabricating Authenticity*. Chicago: University of Chicago Press.

Peterson, Richard A. 2005. "In Search of Authenticity." *Journal of Management Studies* 42(5):1083–98.

Peterson, Richard A., and N. Anand. 2004. "The Production of Culture Perspective." *Annual Review of Sociology* 30(1):311–34.

Peterson, Richard A., and Bruce A. Beal. 2001. "Alternative Country: Origins, Music, World-View, Fans, and Taste in Genre Formation." *Popular Music and Society* 25(1/2):233–49.

Peterson, Richard A., and David G. Berger. 1975. "Cycles in Symbol Production: The Case of Popular Music." *American Sociological Review* 40(2):158–73.

Peterson, Richard A., and David G. Berger. 1996. "Measuring Industry Concentration, Diversity, and Innovation in Popular Music." *American Sociological Review* 61(1):175–78.

Peterson, Richard A., and Roger M. Kern. 1996. "Changing Highbrow Taste: From Snob to Omnivore." *American Sociological Review* 61(5):900–907.

Pettijohn, Terry F., and Donald F. Sacco. 2009. "The Language of Lyrics: An Analysis of Popular Billboard Songs across Conditions of Social and Economic Threat." *Journal of Language and Social Psychology* 28(3):297–311.

Phillips, Damon J. 2013. *Shaping Jazz*. Princeton, NJ: Princeton University Press.

Phillips, Damon J., and Young-Kyu Kim. 2009. "Why Pseudonyms? Deception as Identity Preservation among Jazz Record Companies, 1920–1929." *Organization Science* 20(3):481–99.

Phillips, Damon J., and David A. Owens. 2004. "Incumbents, Innovation, and Competence: The Emergence of Recorded Jazz, 1920 to 1929." *Poetics* 32(3–4):281–95.

Phillips, Damon J., and Ezra W. Zuckerman. 2001. "Middle-Status Conformity: Theoretical Restatement and Empirical Demonstration in Two Markets." *American Journal of Sociology* 107(2):379–429.

Putter, Kaila C., Amanda E. Krause, and Adrian C. North. 2022. "Popular Music Lyrics and the COVID-19 Pandemic." *Psychology of Music* 50(4):1280–95.

Qiu, Lin, Sarah Hian May Chan, Kenichi Ito, and Joyce Yan Ting Sam. 2021. "Unemployment Rate Predicts Anger in Popular Music Lyrics: Evidence from Top 10 Songs in the United States and Germany from 1980 to 2017." *Psychology of Popular Media* 10(2):256–66.

Rao, Hayagreeva, Philippe Monin, and Rodolphe Durand. 2003. "Institutional Change in Toque Ville: Nouvelle Cuisine as an Identity Movement in French Gastronomy." *American Journal of Sociology* 108(4):795–843.

RIAA. 2023. "U.S. Recorded Music Revenue by Format." The Recording Industry Association of America (https://www.riaa.com/u-s-sales-database/).

Rietveld, Joost, and Joe N. Ploog. 2022. "On Top of the Game? The Double-Edged Sword of Incorporating Social Features into Freemium Products." *Strategic Management Journal* 43(6):1182–207.

Rietveld, Joost, Melissa A. Schilling, and Cristiano Bellavitis. 2019. "Platform Strategy: Managing Ecosystem Value through Selective Promotion of Complements." *Organization Science* 30(6):1232–51.

Rogers, Jude. 2017. "Robert Plant on Led Zeppelin, Alison Krauss and His Endless Wanderlust." *The Guardian*, November 2.

Rosa, Jose Antonio, Joseph F. Porac, Jelena Runser-Spanjol, and Michael S. Saxon. 1999. "Sociocognitive Dynamics in a Product Market." *Journal of Marketing* 63(4_suppl1):64–77.

Rosch, Eleanor. 1978. "Principles of Categorization." In *Cognition and Categorization*, edited by E. Rosch and B. L. Lloyd. Hillsdale, NJ: Lawrence.

Rosen, Sherwin. 1981. "The Economics of Superstars." *American Economic Review* 71(5):845–58.

Roy, William G. 2004. "'Race Records' and 'Hillbilly Music': Institutional Origins of Racial Categories in the American Commercial Recording Industry." *Poetics* 32(3):265–79.

Roy, William G., and Timothy J. Dowd. 2010. "What Is Sociological about Music?" *Annual Review of Sociology* 36(1):183–203.

Salancik, Gerald R., and Jeffrey Pfeffer. 1978. "A Social Information Processing Approach to Job Attitudes

and Task Design." *Administrative Science Quarterly* 23(2):224–53.

Schmutz, Vaughn. 2009. "Social and Symbolic Boundaries in Newspaper Coverage of Music, 1955–2005: Gender and Genre in the US, France, Germany, and the Netherlands." *Poetics* 37(4):298–314.

Schumpeter, Joseph A. 1934. *The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest, and the Business Cycle*. Cambridge, MA: Harvard University Press.

Schumpeter, Joseph A. 1939. *Business Cycles: A Theoretical, Historical, and Statistical Analysis of the Capitalist Process*. New York: McGraw-Hill.

Schumpeter, Joseph A. 1950. *Capitalism, Socialism and Democracy*. New York: Harper & Row.

Scott, W. Richard. 2003. "Institutional Carriers: Reviewing Modes of Transporting Ideas over Time and Space and Considering Their Consequences." *Industrial and Corporate Change* 12(4):879–94.

Shannon, Claude Elwood. 1948. "A Mathematical Theory of Communication." *The Bell System Technical Journal* 27(3):379–423.

Shi, Yuan. 2022. "A Change of Tune: The Democratization of Market Mediation and Crossover Production in the U.S. Commercial Music Industry." *Administrative Science Quarterly* 68(2):319–54.

Singerman, Howard. 1999. *Art Subjects: Making Artists in the American University*. Berkeley: University of California Press.

Storper, Michael, and Allen J. Scott. 2008. "Rethinking Human Capital, Creativity and Urban Growth." *Journal of Economic Geography* 9(2):147–67.

Swaminathan, Anand. 1995. "The Proliferation of Specialist Organizations in the American Wine Industry, 1941–1990." *Administrative Science Quarterly* 40(4):653–80.

Tavassoli, Sam, and Charlie Karlsson. 2015. "Persistence of Various Types of Innovation Analyzed and Explained." *Research Policy* 44(10):1887–901.

Teachman, Jay D. 1980. "Analysis of Population Diversity: Measures of Qualitative Variation." *Sociological Methods & Research* 8(3):341–62.

Trust, Gary. 2019. "How the Charts Have Evolved across *Billboard*'s History: The Hot 100, *Billboard* 200 & More" (https://www.billboard.com/pro/billboard-chart-history-evolution-milestones/).

Turner, Scott F., Will Mitchell, and Richard A. Bettis. 2010. "Responding to Rivals and Complements: How Market Concentration Shapes Generational Product Innovation Strategy." *Organization Science* 21(4):854–72.

Tushman, Michael L., and Philip Anderson. 1986. "Technological Discontinuities and Organizational Environments." *Administrative Science Quarterly* 31(3):439–65.

Tversky, Amos. 1977. "Features of Similarity." *Psychological Review* 84(4):327–52.

Uzzi, Brian, Satyam Mukherjee, Michael Stringer, and Ben Jones. 2013. "Atypical Combinations and Scientific Impact." *Science* 342(6157):468–72.

van Venrooij, Alex. 2015. "A Community Ecology of Genres: Explaining the Emergence of New Genres in the UK Field of Electronic/Dance Music, 1985–1999." *Poetics* 52:104–23.

van Venrooij, Alex, Candace Miller, and Vaughn Schmutz. 2022. "Race and Genre Ambiguity in the Critical Reception of Popular Music." *Sociological Inquiry* 92(2):568–96.

van Venrooij, Alex, and Vaughn Schmutz. 2018. "Categorical Ambiguity in Cultural Fields: The Effects of Genre Fuzziness in Popular Music." *Poetics* 66:1–18.

Walter, Achim, Michael Auer, and Thomas Ritter. 2006. "The Impact of Network Capabilities and Entrepreneurial Orientation on University Spin-off Performance." *Journal of Business Venturing* 21(4):541–67.

Watson, Mary R., and N. Anand. 2006. "Award Ceremony as an Arbiter of Commerce and Canon in the Popular Music Industry." *Popular Music* 25(1):41–56.

Wheeler, Darby, and Rodrigo Bascuñán, directors. 2020. "Bounce to This." *Hip-Hop Evolution*, season 4, episode 1.

Whitburn, Joel. 2008. *Joel Whitburn Presents Across the Charts: The 1960s*. Milwaukee, WI: Hal Leonard Corporation.

White, Harrison C. 1981. "Where Do Markets Come From?" *American Journal of Sociology* 87(3):517–47.

Wohl, Hannah. 2022. "Innovation and Creativity in Creative Industries." *Sociology Compass* 16(2) (https://doi.org/10.1111/soc4.12956).

Wu, Lingfei, Dashun Wang, and James A. Evans. 2019. "Large Teams Develop and Small Teams Disrupt Science and Technology." *Nature* 566(7744):378–82.

Xiao, Han. 2018. "BERT-as-Service" (https://github.com/hanxiao/bert-as-service).

Yoshimura, Ryoma, Hiroki Shimanaka, Yukio Matsumura, Hayahide Yamagishi, and Mamoru Komachi. 2019. "Filtering Pseudo-References by Paraphrasing for Automatic Evaluation of Machine Translation." Pp. 521–25 in *Proceedings of the Fourth Conference on Machine Translation*, Vol. 2, *Shared Task Papers, Day 1*, Florence, Italy. Association for Computational Linguistics.

Young, Tom, Devamanyu Hazarika, Soujanya Poria, and Erik Cambria. 2018. "Recent Trends in Deep Learning Based Natural Language Processing." *IEEE Computational Intelligence Magazine* 13(3):55–75.

Zanella, Paola, Paola Cillo, and Gianmario Verona. 2022. "Whatever You Want, Whatever You Like: How Incumbents Respond to Changes in Market Information Regimes." *Strategic Management Journal* 43(7):1258–86.

Zhang, Tianyi, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. 2019. "BERTscore: Evaluating Text Generation with BERT." ArXiv Preprint (https://doi.org/10.48550/arXiv.1904.09675).

Zhou, Changhui, and Jing Li. 2008. "Product Innovation in Emerging Market-Based International Joint Ventures: An Organizational Ecology Perspective." *Journal of International Business Studies* 39(7):1114–32.

Zuckerman, Ezra W. 1999. "The Categorical Imperative: Securities Analysts and the Illegitimacy Discount." *American Journal of Sociology* 104(5):1398–438.

Zuckerman, Ezra W., and Tai-young Kim. 2003. "The Critical Trade-off: Identity Assignment and Box-Office Success in the Feature Film Industry." *Industrial and Corporate Change* 12(1):27–67.

Zullow, Harold M. 1991. "Pessimistic Rumination in Popular Songs and Newsmagazines Predict Economic Recession via Decreased Consumer Optimism and Spending." *Journal of Economic Psychology* 12(3):501–26.

Khwan Kim is a PhD candidate in Management at INSEAD. His research draws on organization theories and computational methods to understand the antecedents and consequences of novelty and innovation in creative industries at both the individual and organizational levels.

Noah Askin is an Assistant Professor of Teaching Organization and Management at The Paul Merage School of Business at the University of California-Irvine. His research examines the antecedents and consequences of creativity and innovation (particularly in the creative industries), the production and consumption of culture, and the dynamics of organizational and individual status. He holds a joint PhD from the Booth School of Business and the Department of Sociology at the University of Chicago.