Running Head: ATYPICALITY AND CULTURAL SUCCESS

Atypicality and Cultural Success

Jonah Berger

University of Pennsylvania

Grant Packard

Wilfrid Laurier University

The authors thank Shawn Zamechek for assistance with the LDA and Amir Goldberg for analysis suggestions. The authors contributed equally and are listed alphabetically. Address correspondence to Jonah Berger: jberger@wharton.upenn.edu.

Abstract

Why do some cultural items become popular? While some have argued that success is random, we suggest that how similar items are to their peers plays an important role. Natural language processing of thousands of songs examines the relationship between lyrical differentiation (i.e., atypicality) and song popularity. Results indicate that the more different a song's lyrics are from its genre, the more popular it becomes. This relationship is weaker in genres where lyrics matter less (i.e., dance) or where differentiation matters less (i.e., pop) and occurs for lyrical topics but not style. The results shed light on cultural dynamics, why things become popular, and the psychological foundations of culture more broadly.

Keywords: popularity, natural language processing, cultural success, psychological foundations of culture, music

Why do some things catch on? Academics and popular writers alike have long been interested in cultural dynamics, or why some songs, movies, and other cultural items become popular (Kashima 2014; Lieberson 2000; Simonton 1980). Some songs become hits while others fail and some movies become blockbusters while others don't. What makes some cultural items successful?

One possibility is that popularity is random. Even domain experts have difficulty predicting success (Bielby & Bielby, 1994), and researchers have argued that popularity is driven by chance patterns of social influence (Salganik, Dodds, & Watts, 2006). These perspectives suggest that success has little to do with features of the cultural item themselves.

Building on research on the psychological foundations of culture, however, we suggest that the similarity between items may help shape success. Research on cross-cultural psychology has demonstrated the influence of culture on individual-level psychological processes (Markus & Kitiyama, 1991). But the reverse is also true; psychological processes shape the norms, practices, and items that make up culture (Norenzayan et al., 2006; Kashima, 2008; Schaller & Crandall, 2004). In particular, people have a drive for stimulation (Zuckerman, 1979), and from an early age, children are attracted to novel stimuli, or those different from what they have experienced (Flavell, Miller, & Miller 2001). This suggests that similarity between concepts should shape cultural success. The things people have experienced should determine how novel a given new cultural item seems. Taken to a collective level, cultural items that are more atypical, or differentiated from their peers, may be liked more, and, consequently, become more popular.

Unfortunately, empirically testing such propositions has been constrained by the ability to quantify differences between cultural items at scale. To address this issue, we use textual analysis to measure lyrical differentiation across thousands of songs. Importantly, the differentiation we examine is bounded, not infinite. A Country song whose lyrics sound like Death Metal would be different from most Country songs, but would also be unlikely to be classified as Country in the first place. Thus, we examine whether *among* songs classified as belonging to a given cultural category (i.e., genre), those whose lyrics are more atypical are more successful. Similar to Boyer's (1994) notion of minimally counterintuitive concepts, such songs are different from the prototype but not so different as to be outside the genre.

Method

First, we collected data on song popularity. To focus on individuals' preferences, we used Billboard's digital download rankings (www.billboard.com/biz) which capture over 90% of major paid song services (e.g., Apple iTunes and Google Play). We focused on this measure of popularity, rather than say, radio airplay, because it is more likely to be driven by individual preferences rather than a small number of institutionalized actors (e.g., DJs). We sampled the ranking data once every three months over a three-year period (2014-16) for each of seven major genres (Christian, Country, Dance, Pop, Rap, Rock and R&B). Comprehensive data was unavailable for the Alternative genre. We obtained all songs that appeared in each genre ranking and their position in that genre's chart (1-50). We reverse code song ranks so that positive coefficients describe a positive relationship with song success. This resulted in a dataset of 4,200 song rankings and 1,879 unique songs. We captured artist name and whether the song appeared on the Billboard radio airplay lists for the same periods as covariates.

Second, we acquired the complete lyrics for each of these songs at SongLyrics.com.

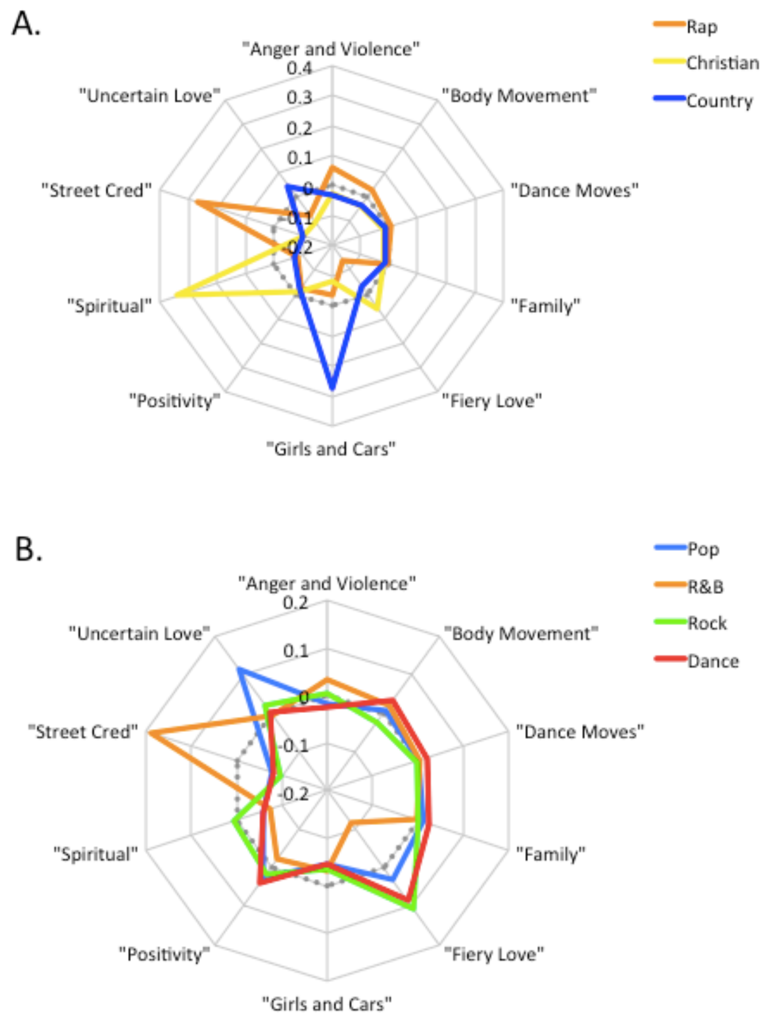
Third, we used latent Dirichlet allocation (LDA; Blei, 2012) to determine the main themes discussed across songs. This approach takes texts (e.g., song lyrics), and by measuring word co-occurrence within and across texts, determines the key latent topics or themes that make up those texts, and the words that make up each topic (Table 1; see Supplemental Materials for more detail on the methodological approach and potential shortcomings). Aggregating across all songs within a genre provides that genre's average topic composition (Figure 1). Country songs, for example, sing a lot about "girls and cars" (39%) and less about "body movement" (2%).

Table 1. Topics and Demonstrative Topic Words

Note: Longer lists of high probability words by topic are presented in table S2.

"Anger and Violence" bad, dead, hate, kill, slay	**"Body Movement"** body, bounce, clap, jump, shake	**"Dance Moves"** bop, dab, mash, nae, twerk	**"Family"** american, boy, daddy, mamma, whoa	**"Fiery Love"** burn, feel, fire, heart, love
"Girls and Cars" car, drive, girl, kiss, road	**"Positivity"** feel, like, mmm, oh, yeah	**"Spiritual"** believe, grace, lord, one, soul	**"Street Cred"** ass, b*tch, dope, rich, street	**"Uncertain Love"** aint, cant, love, need, never

Fig. 1. Relative Use of Lyrical Topics by Genre



Note: Values indicate the difference between a genre's topic use and the average topic use across all songs, where zero (grey dotted line) indicates no difference. Country songs, for example, sing 27% more about "girls and cars" than the average song. Panel **A** depicts genres with more extreme variation in topic use, Panel **B** depicts genres with less extreme variation in topic use.

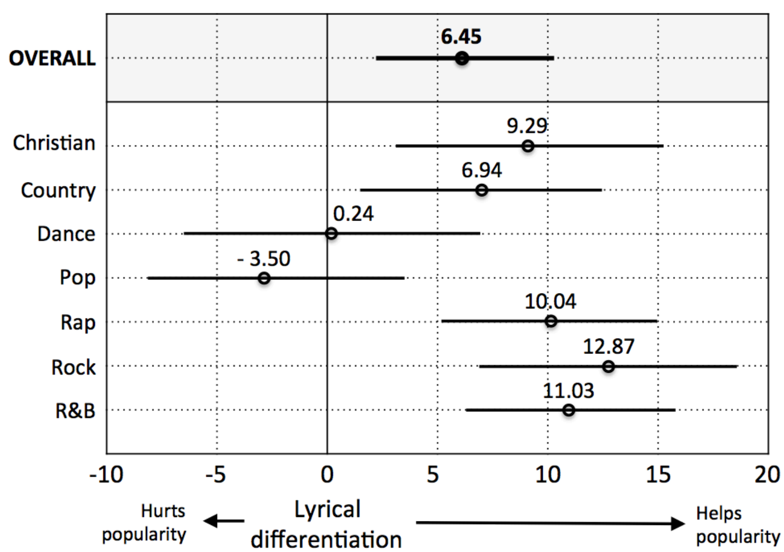
Fourth, we calculate how lyrically different each song is from its genre. For each topic, we take the absolute value difference between that song's lyrical topic composition and the genre mean. Then, we aggregate these differences across topics using Ireland and Pennebaker's (2010) language style matching equation. We invert the resultant value to describe differentiation (rather than matching).

Finally, OLS regression examines the relationship between lyrical differentiation and song performance. Analysis treating rank dependent measures as continuous is appropriate given the large number and fixed range of ranks.

Results

Results indicate that the more differentiated a song's lyrics are from its genre, the more popular that song is (B = 6.45, $t = 3.23$, $p = .001$; Figure 2 and Table 2, model 1). A 16% increase in lyrical differentiation, for example, is associated with a one-position improvement in chart ranking. Results are the same using an ordinal logistic specification (proportional odds; estimate = 0.77, $t = 3.22$, $p = .001$) or log transformation of the rank dependent measure (B = 0.18, $t = 3.42$, $p < .001$). Results are also the same using alternative methods of calculating lyrical differentiation, such as squared (rather than absolute value) differences (B = 43.08, $t = 2.83$, $p = .005$) and Jensen-Shannon divergence (B = 155.14, $t = 3.10$, $p = .002$).

Fig. 2. Impact of Lyrical Differentiation across Genres.



Note: Coefficient estimates and 95% confidence intervals for all songs (first row) and partitioned across genres (subsequent rows) for the main OLS regression. The effect is significant for a genre if the confidence interval does not intersect with zero. In some specifications, Pop has a significantly negative coefficient, suggesting that lyrical differentiation may sometimes *hurt* Pop music success

Table 2. Song Popularity as a Function of Lyrical Differentiation.

	Model 1, Simple effect (1)		*Model 2, artist, song, time controls* (2)		*Model 3, language controls* (3)		*Model 4, top 100 words* (4)	
Lyrical differentiation	**6.45**	**	**8.01**	**	**8.38**	***	**7.84**	**
	(1.99)		(2.45)		(2.48)		(2.66)	
Times charted			0.83	***	0.69	***	0.75	***
			(0.10)		(0.11)		(0.11)	
Multi-genre count			4.79	***	4.58	***	5.24	***
			(0.80)		(0.69)		(0.77)	
Radio airplay			11.17	***	11.10	***	11.09	***
			(0.55)		(0.55)		(0.56)	
LIWC Dictionaries								
Word count					0.00		0.00	
					(0.00)		(0.00)	
Six letter					0.02		-0.02	
					(0.07)		(0.07)	
Cognitive words					-0.06		0.03	
					(0.07)		(0.08)	
Affect words					0.00		0.05	
					(0.08)		(0.08)	
Social words					-0.07		-0.01	
					(0.05)		(0.06)	
Perceptual words					0.02		0.18	
					(0.08)		(0.12)	
Motivation words					-0.02		-0.03	
					(0.06)		(0.06)	
Temporal words					-0.05		-0.06	
					(0.05)		(0.06)	
Relativity words					0.00		-0.05	
					(0.05)		(0.05)	
Swear words					0.15		0.06	
					(0.20)		(0.22)	
Controls								
Artist/song	No		Yes		Yes		Yes	
Topic	No		Yes		Yes		Yes	
Time	No		Yes		Yes		Yes	
Top 100 words	No		No		No		Yes	
Intercept	23.34	***	38.25	***	39.95	***	33.03	***
	(0.70)		(1.75)		(2.48)		(6.40)	
Adjusted R^2	0.023							
Marginal R^2			0.142		0.146		0.175	
Conditional R^2			0.344		0.347		0.367	

**** $p < .001$, ** $p < .01$, * $p < .05$; Coefficients predict reverse-coded song ranking.*

Robustness Checks. We included numerous covariates in the model to assess the stability of the main results and rule out alternative explanations. Even controlling for a range of factors including radio airplay, artist, time, and the topics themselves, the effect of lyrical differentiation remains significant (B = 8.01, $t = 3.27$, $p = .001$, Table 2, model 2). We also considered other factors pertaining to lyrical content. One might wonder whether things like number of words, language complexity, or other major linguistic features not captured by our LDA approach could explain the results. To account for these, we ran a model adding word count, six letter words (a proxy for language complexity), and baskets of words empirically linked to social or psychological constructs (cognitive processing, emotion, sociality, perception, motivation, time, relativity, and formality) from the Linguistic Inquiry and Word Count (LIWC) dictionaries (Pennebaker, Boyd, Jordan, & Blackburn, 2015). Even after including these factors, however, lyrical differentiation remained significant (B = 8.38, $t = 3.38$, $p < .001$, Table 2, model 3).

Finally, we considered whether individual words, rather than the bundles of words used in our LDA-based approach or the LIWC dictionaries might explain the results. Even after including dummies for the presence of each of the 100 most frequent words used across all songs, however, the effect for lyrical differentiation remained significant (B = 8.08, $t = 2.52$, $p = .01$, Table 2, model 4). In sum, across a variety of specifications, lyrical differentiated songs were more popular.

To best identify lyrical differentiation's effect, one would ideally keep all other song aspects the same, vary lyrical differentiation, and examine its influence on popularity. To approximate this, we examine songs that chart in two different genres at the same time. While artist, lyrics, and all other textual features are identical (it is the same song), lyrical differentiation will be greater in one genre than the other, providing a stricter test of lyrical differentiation's impact. If the same song by the same artist is more successful in the genre in which it is more lyrically differentiated (versus one in which it is less differentiated), this would support the notion that lyrical differentiation, rather than some other factor, is driving popularity.

To test this possibility, we analyze the 410 songs that appear on two different genre charts at the same time using an analysis approach similar to difference-in-differences. We calculate the lyrical differentiation of each song from each of its two genres and difference these values.

Results underscore the prior findings, indicating that songs are more popular in genres where they are more lyrically differentiated (B = 34.64, $t = 3.00$, $p = .003$). This, combined with additional analyses (see Supplemental Materials), including using a matched comparison group of less popular (i.e., non-ranked) songs by the same artist from the same album, a model accounting for right truncation, and a two-stage Heckman selection model all cast doubt on the notion that selection can explain the results.

Results also remain the same (see Supplementary Materials) controlling for topical diversity (entropy) and alternative approaches to calculating topics (e.g., within genre) or differentiation (i.e., across time).

Variation by Genre. While lyrically differentiated songs are more popular, might this relationship vary by genre? Lyrics may matter less in Dance music, for instance, where attributes that drive movement (e.g., the beat) may be more important than lyrics. Results are consistent with this prediction (Figure 2). A partitioned regression model shows that while lyrical differentiation is linked to popularity in most genres, the relationship is weaker in Dance, as well as Pop music, which, almost by definition, is more about mainstreaming than differentiation (Frith, 1986).

Type of Atypicality. Further analyses shed light on the type of differentiation linked to success. While Country songs with more differentiated lyrics are more successful, for example, is this because they include *more* Country-associated content (i.e., more "girls and cars" and other genre-typical lyrics than other Country songs) or because they include *less* Country-associated content?

Topics were ordered based on the degree to which their use within a genre deviated from their use across genres. Then, for the five most and least typical topics for each genre we separately calculated and aggregated directional (rather than absolute) lyrical differentiation, Including these two variables as predictors in the base specification indicates that successful songs use *less* genre-typical topics ($B_{\text{Typical topics}} = -3.60$, $t = 3.28$, $p = .001$). In terms of what takes their place, use of less typical topics like "uncertain love" ($t = 2.78$, $p = .005$) and "dance movement" ($t = 2.69$, $p = .007$) are generally linked to greater popularity. Less-used topics linked to success also vary by genre. For example, "street cred" represents only 2% of Rock lyrics, yet Rock songs that sing more about it are more popular ($B = 35.65$, $t = 2.30$, $p = .02$). Additional results are reported in Supplemental Materials.

Topical vs. Style Differentiation. While topically differentiated songs are more popular, the same does not hold for stylistic differentiation. Linguistic topics (or content) refers to *what* someone is discussing (e.g., cars, love, or money), but linguistic style refers to the small subset of words (e.g., prepositions and conjunctions) that relate to *how* a person writes or speaks (Ireland & Pennebaker, 2010). Stylistic differentiation, however, does not predict success ($B = 0.11$, $t = 0.05$, $p = .96$; see Supplemental Materials). Further, the relationship between topical differentiation and popularity persists even controlling for stylistic differentiation. This suggests that successful songs tend to sing about different topics, but not necessarily in a different style (though specific stylistic word features, on their own, may be linked to popularity).

Discussion

Though some have argued that cultural success is impossible to predict, textual analysis of thousands of songs suggests that those whose lyrics are more differentiated from their genres are more popular. This dovetails with recent perspectives on the psychological foundations of culture (Kashima, 2008; Schaller & Crandall, 2004). When shared across individuals, psychological processes shape the practices and items that constitute culture. In this case, value for novelty or difference may underlie the link between lyrical differentiation and cultural success.

Might different aspects (e.g., melody vs. lyrics or cinematography vs. script) play different roles in cultural adoption? In songs, for example, would melodic differentiation also be beneficial? While extrapolation from our findings might suggest yes, melodic elements may be particularly important in determining how people classify a song (e.g., banjo signals Country). If so, lyrics may be freer to deviate from genre norms.

Different aspects of cultural products (e.g., music and lyrics) may also combine to shape differentiation, and thus popularity. Though novelty can be good, there are also benefits to familiarity (Kunst-Wilson & Zajonc, 1980). Rather than achieving the right balance on one dimension alone, successful items may mix similarity and differentiation across dimensions. Similar enough to evoke the warm glow of familiarity but differentiated enough to feel new and exciting. Remixes, for example, add new lyrics to an old tune or vice versa. Differentiation on one dimension may be balanced by similarity on another.

As with any investigation, boundaries apply. While we believe this illustrates a broader pattern, we cannot speak outside of the years we examined. Future work might examine whether these effects generalize across cultures. East-Asians prefer less social differentiation, for example, which might extend to preferences for novelty versus typicality. Similarly, other genres (e.g., opera) may show different patterns based on norms associated with them.

Finally, our findings highlight the value of natural language processing to study cultural dynamics. From digitized books and social media, to news and the congressional record, more and more texts are available. Advances in computational social science provide a rich set of opportunities to extract cultural and behavioral insight from text. Hopefully, this emerging toolkit will help provide deeper insight into why things catch on.

Author Contributions

All authors contributed to the study design and data collection. G. Packard performed the data analysis and interpretation with help from J. Berger. J. Berger drafted the manuscript and G. Packard provided critical revisions. All authors approved the final version of the manuscript for submission.

References

Bielby, W. T., & Bielby, D. D. (1994). All hits are flukes: Institutionalized decision making and the rhetoric of network prime-time program development. *American Journal of Sociology*, 99, 1287–1313.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55, 77-84.

Boyer, P. (1994). *The Naturalness of Religious Ideas. A Cognitive Theory of Religion*, Berkeley, CA: University of California Press.

Flavell, J. H, Miller, P. H. & Miller, S. A. (2001). *Cognitive development*, Upper Saddle River, NJ: Prentice Hall.

Frith, S. (1986), Why do songs have words? *The Sociological Review*, 34, 77-106.

Ireland, M. E., & Pennebaker, J. W. (2010). Language style matching in writing: Synchrony in essays, correspondence and poetry. *Journal of Personality and Social Psychology*, 99, 549-571.

Kashima, Y. (2008). A social psychology of cultural dynamics: Examining how cultures are formed, maintained, and transformed. *Social and Personality Psychology Compass*, 2, 107-120,

Kashima, Y. (2014). How can you capture cultural dynamics? *Frontiers in Psychology*, 5, 3-16.

Kunst-Wilson, W. R. & Zajonc, R. B. (1980). Affective discrimination of stimuli that cannot be recognized. *Science*, 207, 557-588.

Lieberson, S. (2000). *A matter of taste: How names, fashions, and culture change.* New Haven, CT: Yale University Press.

Markus, H. & Kitayama, S. (1991). Culture and the Self: Implications for Cognition, Emotion, and Motivation. *Psychological Review*, 98, 224-253.

Norenzayan, A., Atran, S., Faulkner, J., & Schaller, M. (2006). Memory and mystery: The cultural selection of minimally counterintuitive narratives. *Cognitive Science*, 30, 531-553.

Pennebaker, J.W., Boyd, R.L., Jordan, K., & Blackburn, K. (2015). The development and psychometric properties of LIWC2015. Austin, TX: University of Texas at Austin.

Salganik, M. J., Dodds, P. S., &Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *Science* 311, 854–856.

Schaller, M., & Crandall, C. S. (2004). *The Psychological Foundations of Culture*. Hillsdale, NJ: Erlbaum.

Simonton, D. K. (1980). Thematic fame, melodic originality, and musical zeitgeist: A biographical and transhistorical content analysis. *Journal of Personality and Social Psychology* 38, 972-983.

Zuckerman, M. (1979). *Sensation seeking: Beyond the optimum level of arousal.* Hillsdale, NJ: Erlbaum.

SUPPLEMENTAL ONLINE MATERIALS

Additional Details Regarding Topic Modeling

We used latent Dirichlet allocation (LDA) to determine the main themes discussed across songs. LDA takes a number of texts (e.g., songs), and by measuring word co-occurrence within and across texts, determines the key latent topics or themes that make up those texts, and the words that make up each topic. Each word has a probability of appearing in each topic, with words that are more relevant to a given topic having higher probability.

We pre-processed the data using common techniques (Hopkins & King 2010) including removing punctuation, numbers, and stemming related words (e.g., forest, forests and forested all appear as "forest"). Results were robust to the inclusion/exclusion of both infrequent words and stop words (Lewis, Yang, Rose, & Li, 2004). Both were included in the final analysis given that common stop words (e.g., personal pronouns like "I" and "you") occurred frequently in the data and offer interpretive value (e.g. self- vs. other-involvement; Tausczik & Pennebaker, 2010). All LDA models used Gibbs sampling at 5,000 iterations with a random seed starting point.

We followed prior research (DiMaggio, Nag, & Blei, 2013; Hansen, McMahon, & Prat, 2014) in determining the number of topics. The number of topics (K) to use generally depends on a combination of predictive model fit and face validity or interpretability of topics that emerge at different K. Research incorporating semantic interpretation of topic models frequently settle on a range of seven to 13 topics given diminishing interpretive value of topics after this point with minimal loss in predictive model fit for the focal relationship (DiMaggio, Nag, & Blei, 2013; Chang, Boyd-Graber, Gerrish, Wang, & Blei, 2009).

We performed LDA analysis at levels of K from five to 15, and chose the number that maximized statistical reliability, interpretability, and parsimony. Our main results are robust to a wide range of values of K. The relationship between lyrical differentiation and song popularity is significant at 99% confidence or better at K greater than six. Goodness of fit statistics suggest little benefit as K increases (adjusted R-squared range = 0.10 - 0.11 and conditional R-squared range [Nakagawa & Schielzeth, 2013] = 0.30 - 0.31 for K from seven to 15). We observed modest peaks in model fit at 10 topics. Researcher interpretation of the topics and predictive fit for individual topics also suggested diminishing benefit of reporting more than 10 topics. Thus, in the main text we report results at K = 10. While the perplexity fit measure for topic models reveals a stable improvement in fit (lower perplexity) as K increases, this statistic holds little relationship with interpretability or predictive model fit (Chang, Boyd-Graber, Gerrish, Wang, & Blei, 2009) and so was not used in selecting K.

Representative high probability words for each of the ten topics appear with interpretative labels for these topics in table S1. The topic with words like “car,” “drive,” and “girl,” for example, may be labeled as “girls and cars” and the topic with words like “shake,” “bounce,” and “clap” can be labeled as “body movement.” We use these latent topics to calculate each genre’s average lyric topic composition.

The LDA model outputs the proportion of each song that belongs to each latent topic. A given song, for example, might be 25% “girls and cars,” 10% “body movement,” and so on, with the 10 topics summing to 100%. Averaging each topic’s proportion across all songs in a genre provides that genre’s mean topic composition. Figure 1 (in the main text) depicts the prevalence of each topic in each genre relative to the mean incidence of these topics across all songs (i.e., the 0 point). Christian songs are more likely to sing about “spiritual” topics (words like believe, soul, and will), for example, while Rock songs are more likely to contain lyrics about “fiery love” (words like burn, heart and fire).

Note that while LDA is a central method in natural language processing (Cambria and White 2014), it is not without shortcomings. For example, first, as a “bag of words” model, it does not account for clause, sentence, or larger linguistic structures (e.g., paragraphs) that may moderate the meaning of the words observed in each document (song). Second, the LDA approach we apply is unsupervised and permits correlated topics. While some LDA model variations account for these issues, they are commonly applied where there is conceptual or substantive support for a-priori topics and strongly isolated topics (Andrzejewski & Zhu 2009; Tan & Ou 2010). For the present analysis, we did not wish to impose a-priori restrictions on what culture-makers sing about or whether they sometimes sing about inter-related topics. Third, while we followed common LDA procedures, this method incorporates a degree of subjectivity in the number of topics to report and their interpretation (DiMaggio, Nag, & Blei, 2013).

Robustness Checks

First, one could argue that rather than being driven by consumer demand, the results are driven by promotional activity, or in this case, radio airplay. Differentiated songs may be played more on the radio, and that repeated exposure, rather than lyrical differentiation itself, is what drives success. To test this possibility, we measured promotional activity by capturing whether a song appeared on Billboard’s radio airplay charts the week prior to when it appeared on the download list (*Radio airplay;* $M = 0.21$, SD = 0.40).

Second, we included random effects for artist. Songs by more popular artists may get more attention (Adler, 1988; Rosen, 1981), for example, and if popular artists happen to use more differentiated lyrics (or unpopular artists tend to use less differentiated lyrics) one could argue this is driving the observed relationship between lyrics and success.

Third, because a variety of unobserved song-level factors such as producer or instrumentation could differentially impact song popularity, we nest a random effect for song within the artist effect.

Fourth, rather than a song's differentiation from other genre songs in how much it sings about "uncertain love," for example, it could be that what matters is just how much "uncertain love" a song sings about. We incorporate fixed effects for each of the 10 topics to account for this possibility.

Fifth, we control for the number of genres with which a song was associated (*Multi-genre count; M* = 1.32, SD = 0.48). Multi-genre songs may differentially benefit from broader market presence or, alternatively, experience muted effects of lyrical differentiation given their genre-bending status because the genre norms against which the song may deviate are less defined.

Sixth, we account for time by including fixed effects for each of the 12 quarterly song success measures. There may be time-specific factors driving song success such as seasonality or historical events. This factor also accounts for the fact that songs released in early chart periods have more opportunity to be ranked over multiple periods than songs released later in time.

Seventh, we control for a song's prior success by measuring the number of times the song had appeared among the top 50 songs in prior quarters (*Times charted*; *M* = 4.04, SD = 3.55).

Our focal effect of lyrical differentiation remains significant after including all of these factors in the model (B = 8.01, $t = 3.27$, $p = .001$; Table 2, model 2).

Addressing Selection Bias. While our results demonstrate that more popular songs have more differentiated lyrics, one could argue that the relationship is driven by the particular sample used (i.e., selection). Maybe highly popular songs have more differentiated lyrics than somewhat popular songs, for example, but relatively unpopular songs are also less differentiated.

We tried to address this with the breadth of data analyzed (i.e., thousands of songs across multiple years and a range of genres), but we also address it four additional ways. First, we collected an alternative sample of less popular songs. The number of unpopular songs is effectively unbounded, making random sampling challenging, so we created a matched comparison group of less popular songs by the same artists. For three of our ranking periods (i.e., 1,050 ranked songs), we took each artist with a song on the chart in that period and randomly selected another song from that artists' album that did not make the top 50 in any period.[1] We then recalculated genre topic means and lyrical differentiation for all songs including these additional less popular songs. This comparison allows us to test whether among songs by the *same* artists, released at the *same* time, appearing on the *same* albums, those that achieve greater popularity have more differentiated lyrics. Even using this stricter within-artist comparison, the relationship between differentiation and popularity persists. Compared to less popular songs by the same artist, more popular (i.e., top 50) songs were more lyrically differentiated from their genres (M = .25 vs. M = .27; B = 0.02, t = 6.24, p < .001). Using the main model and treating less popular songs as rank 51 produces the same result (B = 29.59, t = 6.14, p < .001).

Second, results persist accounting for right truncation in the song rank dependent measure (truncated Gaussian, Heckman, 1976; B = 14.41, t = 30.07, p < .001).

Third, we ran a two-stage Heckman selection model (Heckman, 1979) using the matched set of more and less popular songs from the within-artist analysis. Consistent with those results, the first stage probit model confirms that increased lyrical differentiation indeed predicts

---

[1] In a small minority of cases, the top 50 song had been released as a single or on a compilation album of varied artists. In these cases, we selected either the most recent prior single or a random song from the most recent prior album by the top 50 song artist. Effects sustained after including dummy variables controlling for these songs.

membership in the top 50 (B = 2.32, $t = 6.16$, $p < .001$). Importantly, however, a non-significant inverse Mills ratio ($\lambda = -106.93$, $t = 0.6$, $p = .55$) in the second stage (outcome) model indicates that selection does not affect the relationship between lyrical differentiation and song performance.

Fourth, as reported in the main text, we find the same result using songs that simultaneously charted in multiple genres.

Lyrical Topic Diversity. One might wonder whether rather than lyrical differentiation, the results are driven by entropy. Songs that are more differentiated may also sing about more topics, which may add richness or nuance. To test this possibility, we used the two most common measures of entropy (Zhang & Grabchak, 2016). Although greater topical diversity is weakly linked to popularity (Shannon entropy B = 4.32, $t = 2.24$, p =.03; Renyi entropy $\alpha = 2$, B = 2.93, $t = 1.94$, p =.05), lyrical differentiation remained significant when including either form of entropy in the model (Bs > 12.32, *t*s > 2.93, *p*s < .001).

Lyrical Differentiation by Genre and Time. Results are also the same when we calculate topics separately by genre. We performed the same LDA analysis (i.e., 10 topics) within each genre and then calculated lyrical differentiation. As in the main model, songs that were more lyrically differentiated from their genre were more successful (B = 10.11, $t = 4.01$, $p < .001$).

Results are also the same when we calculate lyrical differentiation for each song compared to the smaller set of peer songs that were ranked in the same genre at the same time (B = 6.38, $t = 3.14$, $p = .002$).

Finally, results are the same when we consider multi-genre and single-genre songs independently. Lyrical differentiation predicts song success in analyses limited to songs that appear in two or more genres (N = 1,320; B = 11.08, $t = 3.45$, $p < .001$) or only one genre (N = 2,880; B = 9.94, $t = 3.90$ $p < .001$).

Ancillary Analyses

We conducted several ancillary analyses to provide additional insight into the nature of the observed effects.

More or Less Typical? As discussed in the main paper, to examine whether popular songs include more or less genre typical content, we isolated topics that were more or less typical for each genre. Topics were ranked based on the extent to which their proportional use within a genre deviated from the mean use of that topic across genres. For the Rap genre, for example, "street cred" was the highest ranked *more* genre-typical topic, while "fiery love" was the highest ranked *less* genre-typical topic (confer fig. 1). We then calculated directional (rather than absolute) lyrical differentiation, aggregated separately across the five most typical topics and five least typical topics for each genre. We included these two variables as simultaneous predictors in the base specification.

As reported in the main text, results indicate that popular songs use *less* of the genre's typical topics ($B_{\text{typical topics}} = -3.60$, $t = 3.28$, $p = .001$). Breaking topics down further into typicality quintiles produces similar results, indicating that more popular songs use *less* use of the top two (most genre typical) lyric topic quintiles (1^{st} quintile B = -1.86, $t = 2.57$, $p = .01$; 2^{nd} quintile B = -1.56, $t = 2.84$, $p = .005$).

These results suggest that popular songs tend to avoid genre-typical topics, but is there any pattern as to which less typical topics take their place? Songs that include more language from a genre's least typical topic are less popular (5th quintile B = -1.16, $t = 2.19$, $p = .03$), but given there is no aggregate relationship between popularity and slightly less typical topics (i.e., 3rd and 4th quintile *t*s < 0.5, *p*s > .8), we turn to individual topic differentiation to shed some light on this question.

Topic-Specific Differentiation? We explored whether individual topics may be more or less important in achieving lyrical differentiation both across and within genres. Regression models including the 10 topics as simultaneous predictors of song popularity suggest that less use of the topics labeled "street cred" and "anger and violence" (*t*s > 1.84, *p*s < .05) and more use of the topics labeled "uncertain love" (t = 2.78, p = .005) and "dance movement" (t = 2.69, p = .007) are linked to greater popularity.

Exploratory genre analyses suggest that while "street cred" is negatively related to popularity overall, it may be underutilized in Rock music. This topic represents only 2% of Rock lyrics, yet Rock songs that sing more about it are more popular (B = 35.65, t = 2.30, p = .02). Rock songs may also benefit from singing less about "girls and cars" (B = -13.37, t = 2.47, p = .01), and more about "family" (B = 16.94, t = 2.69, p < .01). Overall, the less-used topics linked to success seem to vary by genre.

Content vs. Style Differentiation. While linguistic content refers to *what* someone is discussing (e.g., cars, love, or money), linguistic style (Ireland and Pennebaker, 2010) refers to the use of a small subset of words (e.g., prepositions and conjunctions) that relate to *how* a person writes or speaks. To test the relationship between stylistic differentiation and popularity, we followed the procedure outlined in Ireland and Pennebaker (2010) to calculate language style matching, transposing the final value to represent differentiation rather than matching. As discussed in the main text, results indicate that stylistic differentiation alone does not predict success (B = 0.11, t = 0.05, p = .96). Further, the relationship between content differentiation and popularity persists even controlling for stylistic differentiation (B = 8.17, $t = 2.28$, $p = .02$). This suggests that successful songs tend to sing about different topics or content, but not necessarily in a different style.

Non-Linear Effects of Differentiation. One might wonder whether the relationship between lyrical differentiation and popularity is non-linear. When squared or cubed terms were incorporated in the main model, however, they were not significant (ts < 1, ps > .3), though our main linear effect still persists. This suggests that the relationship between lyrical differentiation and popularity is linear rather than curvilinear.

Differentiation from Other Genres. We also examined whether song success is driven by similarity to other genres. Rather than being driven by difference from its own genre, for example, one could argue that atypical songs are more popular because they are more similar to other genres and thus appeal to a broader set of people. We tested this alternative two ways. First, we created a variable capturing a song's average lyrical similarity (i.e., the inverse of differentiation) from the genres to which it was *not* assigned. When this new variable was included as a predictor of song success in the main model, it was non-significant (B = 1.30, $t = .31$, $p = .76$) while lyrical differentiation from a song's own genre remained significant (B = 7.42, $t = 1.99$, $p < .05$). Second, we created separate variables capturing a song's similarity to each of the six genres in which it was *not* assigned, and included these as simultaneous predictors of song success alongside lyrical differentiation (from a song's own genre). Lyrical differentiation again predicted success, while a song's similarity to other genres was non-significant (table S1). These analyses cast doubt on the possibility that it is similarity to other genres that is driving success.

Table S1: Assessing the impact of similarity to other genres.

	B	SE	t	p	
Lyrical differentiation	9.82	4.08	2.41	0.02	*
Lyrical similarity (other genre)					
Christian	-2.10	1.97	1.07	0.29	
Country	-0.28	1.85	0.15	0.88	
Dance	3.33	2.16	1.55	0.12	
Pop	1.09	1.98	0.55	0.58	
Rap	1.21	1.91	0.64	0.52	
R&B	1.19	1.89	0.63	0.53	
Rock	-0.67	1.84	0.37	0.72	
Intercept	23.11	0.74	31.17	0.00	***

*** $p < .001$, ** $p < .01$, * $p < .05$

Differentiation from Music More Generally. When studying cultural success, it's important to consider the exposure set. Most people primarily listen to specific genres, and lyrics that feel novel to a Country listener may not feel novel to someone who usually listens to Rap. When measuring aspects like similarity or differentiation, considering its effect across all music (rather than specific genres) may overstate people's actual exposure to music.

Indeed, when we measure differentiation from the entire set of songs across all genres, we find similar, albeit weaker results (B = 4, t = 1.66, p = .10) than the genre-specific differentiation we use in our analyses. How different a song is from its genre should be a better proxy of whether listeners will like it than how different that song is from all music.

Additional References Not in Main Text

Adler, M. (1988). Stardom and talent. *American Economic Review*, 745, 208-212.

Andrzejewski, D. & Zhu, X. (2009). Latent Dirichlet Allocation with Topic-in-Set Knowledge. *Proceedings of the NAACL HLT Workshop on Semi-supervised Learning for Natural Language Processing*, 43-48.

Cambria, E., & White, B. (2014). Jumping NLP Curves: A Review of Natural Language Processing Research. *IEEE Computational Intelligence*, May, 49-57.

Chang, J., Boyd-Graber, J., Gerrish, S., Wang, C., & Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. *Neural Information Processing Systems*, 22, 1-9.

DiMaggio P., Nag M., & Blei D. M. (2013). Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding. *Poetics*, 41, 570-606.

Hansen S., McMahon M., & Prat A. (2014). Transparency and deliberation within the FOMC: a computational linguistics approach. *CEP Discussion Papers*, CFM-DP2014-11.

Heckman, J. (1976). The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for such Models. *Annals of Economic and Social Measurement*, 5, 476-492.

Heckman, J. (1979). Sample selection bias as a specification error. Econometrica, 47, 153-161.

Hopkins, D., & King, G. (2010). A method of automated nonparametric content analysis for social science. *American Journal of Political Science*, 54, 229-247.

Ireland M. E., Pennebaker J. W. (2010). Language style matching in writing: Synchrony in essays, correspondence and poetry. *Journal of Personality and Social Psychology*, 99, 549-571.

Lewis, D. D., Yang, Y., Rose, T. G., Li, F., & RCV1 (2004). A new benchmark collection for text categorization research. *Journal of Machine Learning Research*, 5, 361-397.

Lin, J. (1991). Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory*, 37, 145-151.

Nakagawa S., & Schielzeth H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4, 133-142.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. Retrieved from https://repositories.lib.utexas.edu/handle/2152/31333).

Rosen S., (1981). The economics of superstars. *American Economic Review*, 71, 845-858.

Tan, Y., & Ou, Z. (2010). Topic-weak-correlated Latent Dirichlet Allocation. *IEEE International Symposium on Chinese Spoken Language Processing*, 224-228.

Tausczik Y. R., & Pennebaker J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29, 24-5.

Zhang Z., & Grabchak M. (2016). Entropic representation and estimation of diversity indices. *Journal of Nonparametric Statistics*, 28, 563-575.

"Anger and Violence"	"Body Movement"	"Dance Moves"	"Family"	"Fiery Love"
bad	body	bop	american	alive
dead	bounce	break	boy	burn
get	clap	dab	daddy	feel
hate	feel	funk	dance	fire
kill	hand	mash	diamond	heart
man	high	nae	girl	high
people	jump	nasty	head	hold
real	low	twerk	mama	inside
sh*t	shake	walk	mirror	light
slay	turn	watch	momma	love

"Girls and Cars"	"Positivity"	"Spiritual"	"Street Cred"	"Uncertain Love"
babi	dream	believe	bag	aint
car	feel	god	b*tch	cant
drive	give	grace	blow	dont
eye	green	heart	dope	keep
girl	like	holy	f*ck	love
kiss	mmm	lord	hoe	need
parti	oh	one	money	never
road	please	power	rich	nobody
roll	whoa	soul	street	take
she	yeah	will	tryna	try

Table S2: Representative High Probability Words by Topic