sempre:
Society for Education, Music and Psychology Research

Original Empirical Investigations

Burning out, fading away, and the sophomore slump: Critics' versus fans' ratings of music artists' album quality over time

Psychology of Music
1–18

DOI: 10.1177/03057356241289076
journals.sagepub.com/home/pom
Sage

Gregory D. Webster[1] and Lysann Zander[2]

Abstract

Folk psychology posits that music artists' first albums are considered their best, whereas later albums draw fewer accolades, and that artists' second albums are considered worse than their first—a phenomenon called the "sophomore slump." This work is the first large-scale multi-study attempt to test changes in album quality over time and whether a sophomore slump bias exists. Study 1 examined music critics, sampling all A, B, and C entries from *The New Rolling Stone Record Guide* (2,078 album reviews, 387 artists, 38 critics). Study 2 examined music fans, sampling crowdsourced Rate Your Music ratings of artists with at least one *Rolling Stone* top 500 album (4,030 album reviews, 254 artists). Using multilevel models, both studies showed significant linear declines in ratings of artists' album quality over artists' careers; however, the linear effects were qualified by significantly positive quadratic effects, suggesting slightly convex patterns where declines were steeper among earlier (vs later) albums. Controlling for these trends, a significant and substantial sophomore slump bias was observed for critics' ratings, but not for fans' ratings. We discuss theoretical perspectives that may contribute to the observed effects, including regression to the mean, cognitive biases and heuristics, and social psychological accounts.

Keywords

archival data, sophomore slump, regression to the mean, creativity, music, artistic output, multilevel modeling

Rock and roll is here to stay
It's better to burn out than to fade away

—Neil Young, *My My, Hey Hey (Out of the Blue)*

[1]Department of Psychology, University of Florida, Gainesville, FL, USA
[2]Department of Educational Science, Leibniz University Hannover, Hannover, Germany

Corresponding author:
Gregory D. Webster, Department of Psychology, University of Florida, P.O. Box 112250, Gainesville, FL 32611-2250, USA.
Email: gdwebs@ufl.edu

Creative output changes over careers and lifespans. For example, academic psychologists often publish at higher rates before tenure than afterward (Duffy et al., 2011). Further still, although the quantity of books written, modern paintings created, and jazz albums recorded peak at slightly different ages (mid 40s, mid 30s, and early 30s, respectively), all show a fairly steep rise in output in people's 20s, steep declines in the decade or two after peaking, and a more gradual decline during their sunset years (Miller, 1999).

In addition to quantity, artistic quality also varies over the course of a career. Artists who achieve some recognition for the quality of their debut work (a first novel, solo art show, debut album) and would like to continue their career are faced with the daunting task of producing a second work that is as good as or better than their first (Deichmann & Baer, 2023). Many artists' follow-up efforts fall short of their debut work's quality, which has been described as the "sophomore slump." This term was originally coined to describe a decline in achievement, motivation, and goal focus during a student's second year in college (Baldwin, 1933; Roberts, 1933), but has since spread to not only athletics (e.g., baseball players; Taylor & Cuave, 1994), but also the music industry (Zackery, 2014) to describe artists' challenges of building upon initial success. Simply put, the prevailing folk psychology is that artists' second works are rarely as good as their first ones.

To date, the phenomenon has rarely been investigated empirically. One example is an unpublished thesis in which the author, analyzing the ratings of the first and second albums of 100 randomly selected bands, showed a clear difference: First albums received higher quality ratings that second albums—a sophomore slump effect (Zackery, 2014). But is this the case if we examine a larger sample of albums while accounting for declining trajectories in music artists' album quality over their careers?

One way to address these questions is to examine changes in music artists' album quality over time, because most moderately successful artists—those with recording contracts—release multiple albums over the courses of their careers.[1] This work examines two archival, publicly accessible data sources on album quality ratings—one from music critics and one from music fans. There are multiple theoretical approaches and possible explanations that can help us understand changes in artists' album quality over their careers (or people's perceptions of those changes). These include theories of creativity and genius, statistical artifacts, cognitive heuristics and biases, and other social psychological effects.

Creativity in music

Most prior research on creative music production quantity and quality has focused on Western classical music composers. For example, artistic expertise may guide creativity and productivity across composers' careers, as suggested in a study of works by composers Cole Porter and Irving Berlin, which showed an initial lull in activity followed by a flourish of mid-career hits (Hass & Weisberg, 2009). Related research on five American composers (including Porter and Berlin) showed some support for a 10-year rule of development before truly great works emerge (Ericsson et al., 2009), but there was heterogeneity; some composers required more or less time to develop great works (Hass & Weisberg, 2015).

Other research on classical music composers focusing on age or career stage has shown that creative productivity rapidly increases, peaks around age 40, and then gradually declines (Kozbelt, 2008a, 2014; Simonton, 1977, 1997). In addition, quantity and quality of creative works tend to correlate positively (Simonton, 1997). One-hit (vs multi-hit) wonders were more likely to occur at earlier ages (Kozbelt, 2008b). Moreover, late-peaking (vs early-peaking) composers tended to create more high-quality works later in their careers (Kozbelt, 2008a). Use of

multilevel modeling found no clear relationship between age and melodic originality among 173 composers (Kozbelt & Meredith, 2011), but an inverted-U-shaped function between age and aesthetic significance peaking at 56 years was found among 91 composers (Kozbelt, 2011). Among film composers, creative output correlated negatively with age at first hit or critical accolade, but positively with age at last hit or critical accolade (Simonton, 2007).

Because popular music, such as rock and roll, is more subject to ever-changing cultural fads, dance crazes, market demands, and even new media (e.g., online streaming) than classical music (Hass & Weisberg, 2015; Wald, 2009), other forces may influence how rock artists' albums are rated over the courses of their careers. In addition, because the longitudinal research on creativity described above examines different units of analysis (e.g., individual songs or compositions, works aggregated across years or decades), it remains unclear whether it would apply to music albums as a unit of creative output, which are themselves a collection of songs and often a collective enterprise (e.g., songwriters, artists or performers, producers, sound engineers).

Regression to the mean

As a statistical artifact, regression to the mean involves having at least two sampling time points. At Time 1, we observe the highest- and lowest-scoring cases in a normal distribution. At Time 2, we observe that the highest-scoring cases from Time 1 now have lesser scores, and similarly, the lowest-scoring cases from Time 1 now have higher scores. Both extremes have thus regressed to the mean over time. Inversely, scores near the mean at Time 1 are more likely to be more extreme at Time 2. Regarding music, artists typically get signed to a major recording label only after they have surpassed a high-quality or novelty threshold (Time 1), meaning that their debut album will likely reflect this high quality. Producing a second album (Time 2) that is just as high in quality may be unlikely, given that chance, luck, or years of hard work (e.g., extensive touring, songwriting, and revising) may have played a role in earning an initial recording contract (Time 1). Thus, regression to the mean may play a role in folk psychology's expectations of the sophomore slump effect. Nevertheless, regression to the mean would also predict that debut albums of middling quality would be followed by less-average or more-extreme second efforts, a trend which is not included in folk psychological expectations for a sophomore slump effect.

Anchoring and adjustment

Anchoring and adjustment is a broad cognitive bias that applies to making estimates or judgments under uncertainty (Tversky & Kahneman, 1974)—such as when critics are among the first to rate the follow-up album to a successful debut album. Regarding temporal anchoring and adjustment, people will often use an indicator of the past to make judgments about the present or future (Givi & Galak, 2019). The prior indicator or evaluation serves as the anchor and the new one represents an adjustment away from that anchor. For our purposes, a fan or critic might rate an artist's new album by first recalling their rating for that artist's prior album (4 stars); their new-album rating could then be biased in the direction of their prior rating (e.g., "This album isn't as good as their last one, so I'll give it 3 stars"). Regarding comparative anchoring and adjustment, if a fan or critic believes in the folk psychology of the sophomore slump, then they may view all artists' sophomore albums with skepticism and rate them lower than normal. In this case, the expectation of a below-average sophomore album for any artist in general serves as the anchor for fans' or critics' rating of a specific artist's sophomore album

(e.g., "This album is okay, but because I know that sophomore albums aren't usually as good as debuts, I'll give it 2 stars."). In this example, people's ratings of album quality may be adjusted based simply on their awareness of or belief in the sophomore slump effect. To be sure, such assessments need not be conscious; theoretically, knowledge of the sophomore slump may be sufficient to affect one's implicit or subconscious attitudes (Gawronski, 2019) about second albums.

Conformity and need to belong

In Asch's (1955, 1956) classic studies on group conformity, people routinely gave incorrect answers to simple perceptual tasks (e.g., judging which of three lines was the longest), but only when other people in the group uniformly gave the same incorrect answers prior to the participant's turn to provide a judgment. Participant conformity rates reduced dramatically when only a single other group member announced a correct—but socially dissenting—judgment. More recent replications of Asch's studies suggest that group conformity effects remain persistent and generalize beyond their original US samples (e.g., Portugal; Neto, 1995). Regarding album quality ratings, neither critics nor fans live in a social vacuum, and there is some evidence of social conformity among professional music critics (Lundy, 2010). One study, which compared music critics' ratings of 50 randomly selected albums with those of non-experts found considerable consensus among critics' album ratings ($r = .61$), but little if any consensus among non-experts ($r = .08$; Lundy & Smith, 2017). Conformity, however, is merely one behavioral aspect of a broader psychological desire to feel socially included—the need to belong (Baumeister & Leary, 1995; Leary, 2010)—by which quality ratings may also be influenced. If fans or critics believe or note that valued or popular others are giving lower ratings to artists' later albums, higher ratings to artists' earlier albums, or exceptionally low ratings to their second albums (i.e., a sophomore slump bias), then they may be more likely to do so themselves to gain social acceptance, status (Anderson et al., 2015; Field et al., 2024), or perceived competence (Kervyn et al., 2009) among their peers. Indeed, people's preferences for certain artists or music genres can be strong enough to permeate their social identities (Hargreaves & North, 1999; Lonsdale, 2021), from which they can enhance their self-esteem and gain a sense of community and group membership (e.g., deadheads—fans of the Grateful Dead), by aligning their responses with those of a valued reference group (Field et al., 2024).

This research and predictions

This work used two studies—one on music critics and one on music fans—to examine change over time (i.e., music careers) in the quality ratings of thousands of albums by hundreds of music artists. Based on some of the literature and theoretical perspectives reviewed above, we developed three predictions for this study:

1. Ratings of album quality will decrease over time (career course) for the average music artist (negative linear effect).[2] On average, artists' debut albums will be rated as their best works; their last or latest albums will be rated as their worst.
2. The negative linear effect of time (career course) will be moderated by a positive quadratic effect of time such that the overall temporal effect will be slightly convex. In other words, ratings of the average artist's album quality will see steeper drops over their first few albums, but more gradual decreases over their later albums (i.e., the negative linear effect becomes less negative over time).[3]

3. Ratings of artists' second albums will be lower than what would be expected, even given the above-mentioned linear and quadratic temporal effects (Predictions 1 and 2). Thus, the average artist will experience a "sophomore slump" in their album quality, over and above the expected post-debut-album drop in quality that a quadratic temporal model implies.

Study 1: Album ratings from music critics

Method

Sample and procedure. All A, B, and C artist entries (i.e., ABBA to Sonny Curtis) from *The New Rolling Stone Record Guide* (Marsh & Swenson, 1983) were sampled, yielding 2,545 album ratings for over 400 artists from 38 critics. Although this sampling range or stopping rule—just the ABCs—was arbitrary, we ran no analyses before completing data collection (Simmons et al., 2011; Wicherts et al., 2016). Although determining precise statistical power for three-level multilevel models is as much an art as a science (De Jong et al., 2010; Lee & Hong, 2021), the number of artists sampled—the focal unit of analysis—was sufficient for detecting the median effect size in social psychology ($r = .18$; Richard et al., 2003) at .80 power (assuming $\alpha = .05$, two-tailed). We recorded critics' initials, their album ratings, and each album's release year; albums without release-year data ("NA" date entries) were excluded from the sample. Because testing temporal trends and the sophomore slump required multiple albums with release-year data per artist, the analyzed sample had 2,078 album reviews from 387 artists ($M = 5.37$ albums per artist). Analyzing only studio albums may provide a purer test of album-rating trends and effects because compilation albums are often an artist's hits aggregated over several studio albums recorded years apart, thus making temporal trends difficult to discern. Thus, we further refined the sample for some analyses to exclude all obvious live and compilations albums (titles containing, e.g., "live," "concert," "best," "hits"), reducing the sample to 1,909 albums from 377 artists ($M = 5.06$ albums per artist; 92% and 97% of the overall sample, respectively).

Measures. Music critics used a six-point, star-based rating scale: 0 (*Worthless*), 1 (*Poor*), 2 (*Mediocre*), 3 (*Good*), 4 (*Excellent*), and 5 (*Indispensable*). Critics used the full scale range (0–5; $M = 2.36$, $SD = 1.19$). Album release years ranged from 1954 to 1982 (Mode = 1978, $M = 1975.13$, $SD = 4.82$). Although infrequent, when two albums by the same artist were released in the same year—say, 1970—we coded them as 1970.0 and 1970.5; we coded three as 1970.0, 1970.33, and 1970.67; four as 1970.00, 1970.25, 1970.50, and 1970.75; and so on. We also recorded the number of albums each critic rated, ranging from 2 to 637 ($Mdn = 25$, $M = 54.68$, $SD = 111.20$). Because these data were positively skewed (skewness = 4.49, kurtosis = 21.81), we log transformed them to produce a normal distribution ($Mdn = 3.22$, $M = 3.24$, $SD = 1.12$, skewness = 0.48, kurtosis = 1.35). If we consider this variable as a measure of reviewer experience, then this log transformation is reasonable because the psychological difference in experience between rating 10 and 11 albums is larger than that between rating 100 and 101 albums, and so on.

Data analysis. Album ratings were hierarchically nested across three levels: albums (Level 1) nested within artists (Level 2) nested within critics (Level 3). We ran a series of multilevel polynomial regression models (Raudenbush & Bryk, 2002) using the program HLM 6 (Raudenbush et al., 2004);[4] see online supplemental materials (OSM) for details. OSM, data, and HLM outputs for both studies are available for inspection via OSF: https://osf.io/nzxwm.

Table 1. Study 1: Multilevel Model Results for Music Critics' Ratings.

Predictor	All albums[a]			Studio albums[b]		
	Coefficient	*t*	$p \leqslant$	Coefficient	*t*	$p \leqslant$
Prediction 1						
Intercept	2.496	27.85	.001	2.489	26.51	.001
Debut Year	–0.033	–4.94	.001	–0.039	–4.99	.001
Prediction 2						
Intercept	2.536	26.64	.001	2.526	25.68	.001
Debut Year	–0.063	–4.07	.001	–0.064	–4.29	.001
Debut Year^2	0.0028	2.60	.014	0.0025	2.80	.008
Prediction 3						
Intercept	2.570	26.28	.001	2.568	25.04	.001
Debut Year	–0.059	–4.25	.001	–0.061	–4.42	.001
Debut Year^2	0.0023	2.50	.017	0.0020	2.60	.014
Sophomore	–0.112	–2.52	.016	–0.123	–3.04	.005

Note. Outcome: 0–5-star rating scale. Debut Year is the release year of an album centered around the release year of each artist's first album (linear time); Debut Year^2 is its squared term (quadratic time). Sophomore is a binary variable testing the sophomore slump, whereby an artist's second album is coded 1 and their other albums are coded 0.
[a]2,078 albums from 387 artists rated by 38 music critics.
[b]1,909 albums from 377 artists rated by 38 music critics.

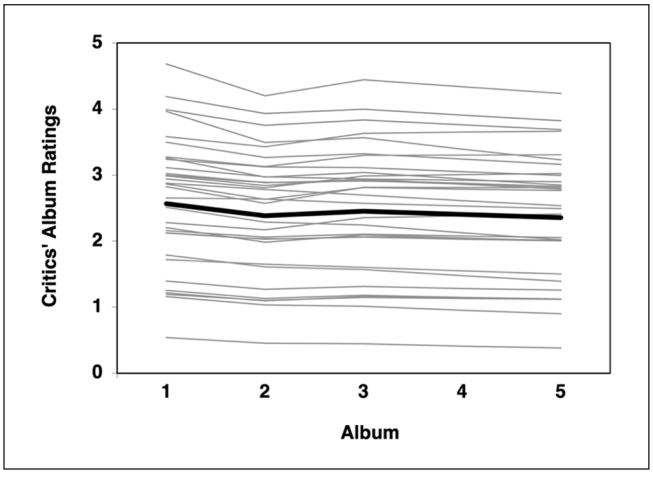

Figure 1. A Sample of 32 Artists' Album Ratings (Thin Gray Lines) and the Average Artist for the Entire Sample (Thick Black Line).
Note the "sophomore slump" decrease between the first and second album for the average artist.

Results and discussion

We present three sets of results below. The first examined all albums, including live and compilation albums (see Table 1, left). The second excluded these to focus on studio albums, thus providing more robust tests of temporal trends and the sophomore slump effect (see Table 1,

Table 2. Study 1: Studio Album Effects Moderated by Critics' Reviewer Experience.

Predictor	Coefficient	*t*	$p \leqslant$
Intercept	2.656	25.71	.001
Experience	–0.153	–2.97	.006
Less	2.831	21.29	.001
More	2.480	24.01	.001
Debut Year	–0.083	–3.38	.002
Experience	0.021	2.24	.031
Less	–0.107	–3.13	.004
More	–0.058	–3.57	.001
Debut Year^2	0.0033	1.82	.076
Experience	–0.0011	–1.65	.106
Less	0.0046	1.80	.079
More	0.0020	1.75	.088
Sophomore	–0.167	–2.52	.017
Experience	0.054	2.20	.034
Less	–0.229	–2.54	.016
More	–0.106	–2.21	.033

Note. Outcome: 0–5-star rating scale. Debut Year is the release year of an album centered around the release year of each artist's first album (linear time); Debut Year^2 is its squared term (quadratic time). Sophomore is a binary variable testing the sophomore slump, whereby an artist's second album is coded 1 and their other albums are coded 0. Experience is the log number of albums reviewed and rated by each critic in the sample. Less and more are the simple effects of experience at 1 *SD* below and above the mean experience level. Sample: 1,909 albums from 377 artists rated by 38 music critics.

right; Figure 1). The third examined the moderating effects of critics' experience on their ratings of artists' studio albums (see Table 2).

Preliminary, intercept-only models at all levels (i.e., null models) revealed that 43.9% (all) and 42.5% (studio) of the variance in album ratings was within artists (variance across an artist's albums over time), 51.1% and 51.8% was between artists (some artists are rated differently on average than others), and 5.1% and 5.7% was between critics. There was sufficient variance to proceed with multilevel models at both Levels 2 (all: $\chi^2_{349} = 2{,}470.75$, $p < .001$; studio: $\chi^2_{339} = 2{,}430.84$, $p < .001$) and 3 (all: $\chi^2_{37} = 55.73$, $p = .025$; studio: $\chi^2_{37} = 56.29$, $p = .022$). Album ratings data were sufficiently normal; residuals for both models—all and studio—at all three levels produced Gaussian distributions with normal quantile–quantile (Q–Q) plots and trivial skewness statistics (< 1 in absolute value). The overall, grand-mean ratings were 2.399 ($SE = 0.081$) and 2.379 ($SE = 0.083$) for all and studio albums, respectively; neither differed significantly from the scale midpoint of 2.5 stars ($t_{37} = -1.25$, $p = .220$; $t_{37} = -1.46$, $p = .153$), suggesting that music critics were collectively well-calibrated to their six-point scale.

We tested three main models: linear, quadratic, and sophomore slump (see Table 1). Supporting Prediction 1 (linear time), for both all-album and studio-album samples, critics' ratings of album quality decreased linearly over the course of the average artist's career, suggesting that their best works were often their debut albums, and late-stage-career albums were among their worst. Supporting Prediction 2 (quadratic time), for both samples, the above-mentioned linear effect was qualified by its quadratic effect: Critics' ratings of artists' album quality

showed a slightly convex function over their artistic careers; the drop in quality became less steep with each ensuing album released. Although statistically significant, the quadratic effect was comparatively weaker than the overall linear effect. Prediction 3 was also supported for both samples: Critics' ratings of artists' second albums fell reliably below the expected trajectory given the quadratic function described above (see Figure 1). Even after accounting for the expected decline in album quality over career time, artists' second albums were rated especially low by critics, showing substantial support for a sophomore slump bias. Although the sophomore slump effect shown in Figure 1 may appear small, its corresponding effect size based on $t(37) = -3.04$ for studio albums is actually substantial—Cohen's $d = -0.50$—suggesting that the average critic devalues second albums by half a standard deviation even after accounting for the observed linear and quadratic trends.

Table 2 shows the results of testing critics' experience—log number of albums reviewed (grand-mean-centered)—as a moderator of the overall intercepts, linear and quadratic temporal career effects, and the sophomore slump effect. Critics' experience related negatively to overall ratings, suggesting that critics who had rated more albums were more critical in their ratings. Critics with less experience (1 *SD* below the mean) gave mean ratings of 2.831, substantially above the scale midpoint of 2.5, whereas those with more experience (1 *SD* above the mean) averaged very near the scale midpoint: 2.480. Reviewer experience did not moderate the quadratic effect of debut year ($p = .106$), and its moderation of its linear effect ($p = .031$) only reflects the line tangent to the quadratic curve at that specific time point (debut album year)—a simple effect in the context of this model, which is not especially meaningful. Reviewer experience related positively to the sophomore slump effect ($p = .034$). Although both less- and more-experienced critics showed a sophomore slump effect, less-experienced critics showed a stronger (more negative) bias (coefficient = –0.229, $p = .016$) than more-experienced ones (coefficient = –0.106, $p = .033$).

The results supported all three main predictions. Critics' ratings of album quality decreased over the temporal course of artists' careers on average (linear effect), and the rate of this decline was greater for their first few albums than later ones (quadratic effect). Adjusting for these temporal career effects, critics gave unusually negative ratings to artists' second album, showing evidence consistent with a sophomore slump bias. Perhaps paradoxically, less-experienced critics—those who had rated and reviewed fewer albums—showed comparatively more of a sophomore slump bias than more-experienced critics, though critics showed a significant amount of bias regardless of experience.

Study 1 had multiple limitations. First, although results were strikingly similar for both all-album and studio-album samples, and although the studio-album sample provided a more robust test of our main predictions, there may have been some live and compilation albums that remained in the studio-album sample despite a fairly exhaustive culling based on keywords (i.e., not all live and compilation albums contain "live," "concert," "best," or "hits" in their titles). Second, in part because the sample was drawn from a book published in 1983, the mean number of albums per artist was only 5.06; a contemporary sample from 2023 would yield 30 more years of data and likely increase the average number of albums per artist. Third, drawing on the wisdom of crowds (or crowdsourcing; Surowiecki, 2004), a better measure of album quality might be an aggregate score across hundreds of fans rather than relying on a single rating from a single critic. To these ends, in Study 2, we sampled aggregated album ratings from thousands of fans via a crowdsourced album-rating website.

Study 2: Album ratings from music fans

Method

Study 2 attempted to replicate the findings of Study 1 using a larger sample of albums while extending it using album ratings from music fans (vs professional music critics).

Sample and procedure. Because there are millions of music albums (Stein, 2020), we chose to circumscribe a sample that would produce adequate variance in ratings of artists' albums over time. To this end, we chose *Rolling Stone* magazine's initial December 2003 listing of the top 500 albums of all time (hereafter *RS* 500; Rolling Stone, 2003). By definition, artists featured in this list have at least one outstanding album, and most have produced several albums to allow for adequate estimation of temporal and sophomore-slump effects. Several artists had multiple albums in the *RS* 500, including The Beatles (11), Bob Dylan (10), The Rolling Stones (10), Bruce Springsteen (8), and The Who (7). In contrast, some *RS* 500 entries were various-artist compilations (e.g., *Nuggets: Original Artifacts from the First Psychedelic Era, 1965–1968*) or by artists with only a single studio album (e.g., *Layla* by Derek and the Dominos); these were excluded, yielding an analyzed sample of 254 artists from the *RS* 500.

Using data from Rate Your Music (RYM; https://rateyourmusic.com), which aggregates music fans' ratings of albums, we recorded by hand the average album ratings from all studio albums released by the abovementioned 254 artists, yielding a sample of 4,030 albums, their ratings, and years of release. Rather than rely on a 0–5-star rating from a single professional music critic, RYM simply aggregates hundreds or even thousands of fans' ratings using a 0.5–5.0-star rating system (with 0.5-star increments). Fans can provide only one rating per album. Thus, instead of relying on professional critics, RYM harnesses the hivemind of crowdsourcing (Surowiecki, 2004).

Measures. People who choose to rate albums on RYM use a 10-point, five-star rating system from 0.5 stars (*lowest*) to 5.0 stars (*highest*). RYM shows users aggregated average scores rounded to three significant figures (i.e., two decimal places). In the present sample of 4,030 albums, the range was from 1.04 (two different albums) to 4.32 (Pink Floyd's *Wish You Were Here*) with a grand mean of 3.31 stars ($SD = 0.42$). As of January 2023, RYM claimed to have logged over 110 million ratings of over 5 million releases, suggesting that the average album's mean rating is based on over 20 ratings. Because our Study 2 sample featured more popular artists (those with $\geqslant 1$ *RS* 500 album), most album averages were based on far more ratings, often in the hundreds or thousands. Because RYM features separate categories for studio albums versus other types of releases (e.g., live albums, compilation albums, EPs, singles), there was no need to exclude non-studio albums based on their title-word searches (e.g., "live," "greatest hits"; cf. Study 1). Only studio albums that RYM labeled as "archival"—often older material released years after it was originally recorded—were omitted from analysis. Album release years ranged from 1946 to 2023 (Mode = 1973, $M = 1984.63$, $SD = 16.53$). The analyzed sample included 4,030 albums from 254 artists ($M = 15.87$ albums per artist).

Data analysis. Because albums (Level 1) were nested within artists (Level 2), we examined a series of multilevel models using the program HLM 6 (Raudenbush et al., 2004), which mirrored those presented in Study 1, but without a Level 3 (critics); see OSM for details.

Table 3. Study 2: Multilevel Model Results for Music Fans' Ratings.

Predictor	Coefficient	t	$p \leqslant$
Prediction 1			
Intercept	3.524	251.98	.001
Debut Year	–0.017	–15.24	.001
Prediction 2			
Intercept	3.595	210.99	.001
Debut Year	–0.034	–11.94	.001
Debut Year^2	0.00048	7.65	.001
Prediction 3			
Intercept	3.595	212.84	.001
Debut Year	–0.034	–12.21	.001
Debut Year^2	0.00048	7.86	.001
Sophomore	0.00072	0.04	.968

Note. Outcome: 0.5–5.0-star rating scale. Debut Year is the release year of an album centered around the release year of each artist's first album (linear time); Debut Year^2 is its squared term (quadratic time). Sophomore is a binary variable testing the sophomore slump, whereby an artist's second album is coded 1 and their other albums are coded 0. Sample: 4,030 albums from 254 artists rated by tens of thousands of music fans.

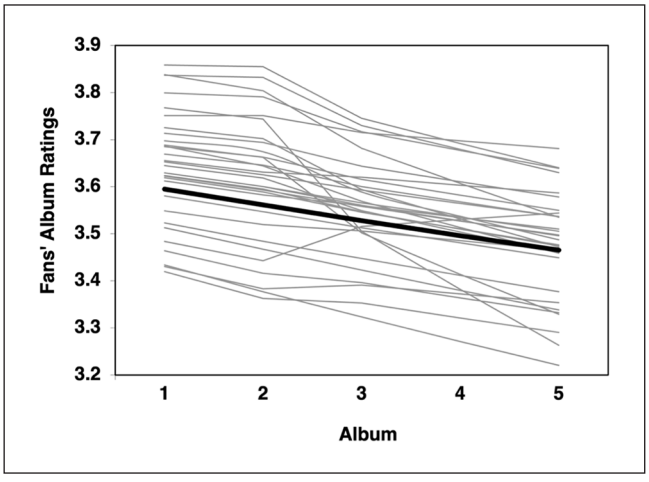

Figure 2. A Sample of 32 Artists' Album Ratings (Thin Gray Lines) and the Average Artist for the Entire Sample (Thick Black Line).
Note the absence of a "sophomore slump" decrease between the first and second album for the average artist.

Results and discussion

A null model (no predictors) revealed that 79.6% of the variance in album rating was within artists (variance across an artist's albums over time) and 20.4% was between artists (some artists are rated differently on average than others), showing there was sufficient variance to proceed with a multilevel model ($\chi^2_{253} = 1{,}136.51$, $p < .001$). Album ratings data were sufficiently normal; residuals at both levels produced Gaussian distributions with normal Q–Q plots and trivial skewness statistics (< 1 in absolute value). The overall, grand-mean album rating was 3.339 ($SE = 0.014$), which was significantly higher than the scale midpoint of 2.75 stars

($t_{253} = 42.07$, $p < .001$), suggesting that fans were generous in their use of RYM's 0.5- to - 5-star scale, or that artists with at least one album in the *RS* 500 simply earn higher ratings. This 0.589 mean difference may also indicate some selection bias—fans are more likely to rate albums by artists they like and to do so positively. The standard errors in the fans' data ($SE = 0.014$) are notably smaller than those in the critics' data (SEs ≈ 0.082)—by 83%—because the former are based on averages of aggregated ratings from hundreds of fans versus the latter, which are based on single ratings from single critics.

We again tested three main models: linear, quadratic, and sophomore slump (see Table 3; Figure 2). Supporting Prediction 1 (linear time), fans' ratings of album quality decreased linearly over the course of the average artist's career; their best works were often their debut albums, their worst were often their latest. Supporting Prediction 2 (quadratic time), the abovementioned linear effect was qualified by its quadratic effect: Fans' ratings of artists' album quality showed a slightly convex function over their artistic careers; the decrease in quality was less steep with each additional album released. Although significant, the quadratic effect was again comparatively weaker than the overall linear effect. Prediction 3 was not supported: Fans' ratings of artists' second albums fell right along the predicted quadratic function described above ($p = .968$). Thus, in stark contrast to the results for critics, there was no evidence of even a hint of a sophomore slump effect or bias in fans' album quality ratings for the average artist.

General discussion

Critics' and fans' ratings were consistent regarding the diminishing quality of music artists' albums over the courses of their careers; however, only critics appeared susceptible to a sophomore slump bias, in which artists' second albums were given especially low ratings. Specifically, both critics' and fans' ratings showed negative linear effects over time (i.e., the courses of music artists' multi-album careers) that were qualified by positive quadratic effects, thus yielding slightly convex functions for the average artist (see Figures 1 and 2). Thus, to address Neil Young's adage, it appears that most music artists show a pattern of "fading away" rather than "burning out." In other words, album quality tends to diminish over time or with each successive album; it is exceedingly rare that music artists burn out in a blaze of glory, achieving their greatest accolade with their final album. That both critics and fans showed markedly similar patterns in their ratings across artists and their albums suggests that these gradual decreases in album quality over artists' careers is a robust phenomenon. This pattern of slowly decaying rock album quality over time contrasts with findings on the quality and quantity of output over the careers of classical music composers, who often show an inverted-U-shaped trajectory, typically peaking around midlife (Kozbelt, 2008a, 2014; Simonton, 1977a, 1997). Although speculative, this difference may suggest that popular (vs classical) music is more likely to be influenced by several factors that are beyond the artist's control, such as mercurial fads and market forces (Wald, 2009). In contrast, the sophomore slump effect appears to be an expectation bias solely among music critics, but not among music fans. Although our research cannot speak to critics' motivations, if there is a culture of consensus or conformity (Asch, 1955, 1956; Neto, 1995) among critics of believing that artists' second albums are often substantially worse than their debut albums, then such a sophomore slump bias would likely be persistent and pervasive. Recall that critics' album ratings showed substantial consensus ($r = .61$), whereas non-experts showed next to none ($r = .08$; Lundy & Smith, 2017). As non-experts, fans who rate artists' albums online are presumably unburdened by the same social norms as critics because fans need not conform to the same consensus pressures (i.e., fans have no skin in the

game). From this social psychological perspective, fans' lack of a sophomore slump rating bias may be unsurprising.

Implications for theory

In the introduction section, we outlined multiple statistical and theoretical perspectives that offer possible explanations for the observed effects, including regression to the mean, anchoring and adjustment bias, and social conformity. The overall pattern of artists producing higher-quality albums earlier in their careers was partly consistent with a regression-to-the-mean account, whereby artists must exceed a certain quality threshold to have an album produced and distributed in the first place. Thus, producing a second album of equal or greater quality may be difficult for artists, especially if they benefited initially from luck, chance, or a large catalog of well-honed songs to choose from for their debut album. Consequently, a high rating for an artist's first album can be expected to naturally regress to a more middling rating for their second album, at least partly due to regression to the mean following some initial success. Nevertheless, recall that regression to the mean also implies that debut albums with average ratings should be followed by sophomore albums with more extreme ratings, which is a trend we did not find in our analyses. Thus, a regression-to-the-mean explanation is an incomplete one that the data can only partly support.

Our observation of a sophomore slump bias—at least among music critics—was consistent with an anchoring-and-adjustment heuristic (Tversky & Kahneman, 1974). Specifically, critics may show a local anchoring-and-adjustment bias when rating multiple albums by the same artist, if they believe that an artist's debut album is especially high in quality, and they adjust away from that initial anchor by too much when rating their second album. Alternatively, critics may show a global anchoring-and-adjustment bias toward an artist's second album if there is a cultural expectation among music critics that sophomore albums are of lower quality, and they then fail to adjust enough away from this initial anchor to give an artist's second album an unbiased—or less-biased—rating.

Assuming that professional music critics often read each other's album reviews and ratings, they are likely susceptible to social norms including conformity effects. And if most music critics are simply aware of the sophomore slump concept, then they may act on it, seeking social approval from other critics by unduly panning artists' second albums. To be a critic requires some calibration; one's album ratings cannot all be zero-star or five-star reviews. Instead, to show one belongs to the community of critics, one should calibrate to the social norms, which might include denigrating music artists' second albums. If such a sophomore slump bias exists among critics, as the findings of Study 1 suggest, then it need not even require conscious effort. Although our data cannot speak to it directly, we speculate that simply knowing that the sophomore slump effect exists may be enough for critics to exhibit implicit or unconscious bias in their album ratings (see Gawronski, 2019).

Limitations and constraints on generality

Although we have highlighted multiple theoretical perspectives, because our findings were largely descriptive, we cannot say if any one account is the "right" or "best" one. In all likelihood, each of these perspectives—creativity, biases and heuristics, social psychology, and even regression to the mean—are contributing in complementary ways to both the overall pattern of declining album quality and the sophomore slump observed among music critics (but not fans).

A potential confound across studies is that critics often rate and review albums as they are released, whereas fans, especially new fans, may do so years later and out of chronological order, which may diminish our ability to detect a sophomore slump in the RYM data. If fans do indeed rate albums in a haphazard, non-chronologic order, then they might operate like a moving average, thus washing out small, specific fluctuations such as a sophomore slump effect. Another limitation is that we did not have access to critics' ages at the time they rated various albums. This is important because people tend to rate the music they experienced during their 20s as the best (Holbrook & Schindler, 1989), which likely relates to nostalgia for the intense and formative feelings of their youth (Holbrook & Schindler, 2003). A controlled experiment in which (a) temporal album order is manipulated and (b) raters' ages are assessed may be necessary to address both of these potential limitations.

The generalizability of our findings is constrained in multiple ways. First, most of the artists sampled in both studies were rock and roll musicians, largely because of *Rolling Stone* magazine's long history of covering this genre. Although some country and western, rhythm and blues, and hip-hop musicians were featured in both samples, only those that have experienced cross-over success in rock and roll were included. Second, because British and American straight cisgendered White men have long dominated rock and roll bands and solo artists (despite the genre's roots in African American blues), this dominance was reflected in both samples (Consequence of Sound, 2020). Third, nearly all artists and albums feature English-language lyrics. Neither of our samples can speak to possible album quality trends in other widely spoken languages (e.g., Mandarin Chinese, Hindi, French, Spanish, Portuguese). Fourth, because the critics sampled in Study 1 were also largely Anglo-American White men, it remains unclear how a more diverse set of music critics would rate the same albums. Although Study 2 addressed this limitation by taking a crowdsourcing approach—often featuring hundreds of fans rating each album—we know next to nothing about the diversity of RYM's thousands of raters other than that they were almost certainly more diverse than our sample of 38 music critics. Fifth, we did not account for concomitant changes in American popular music over time (e.g., slower tempos, more minor modes; Schellenberg & von Scheve, 2012), which could affect how fans and critics rate albums. Sixth, because our studies of critics (1954–1982) and fans (1946–2023) covered partly overlapping time periods, it is unclear how concomitant temporal changes in social or cultural norms may have influenced the strength of the sophomore slump effect.

This work also cannot speak to precisely why people like or dislike certain music or artists or albums (Ladinig & Schellenberg, 2012; Schäfer & Sedlmeier, 2010). For example, we did not examine the likely crucial role of emotion—especially feelings of happiness and sadness—that music may directly or indirectly induce (Brattico & Pearce, 2013; Hunter et al., 2010; Konečni, 2008; Vuoskoski & Eerola, 2012; Webster & Weir, 2005).

Future directions

Future research should attempt an experimental manipulation where album raters are randomly assigned to be told that a given album by a fictional band is either their first or second album, and to see if their ratings differ. Information on raters' self-reported expertise or interest in recorded music could be examined as a moderator to see whether any evidence of a sophomore slump effect is stronger among people who believe that they have more expertise in rating albums.

In addition, future research could consider many environmental factors that can influence musicians' creativity and productivity. For example, a biographical analysis of 10 classical music composers identified not only age but also physical illness and peer competition as key

indicators of productivity (Simonton, 1977). Perhaps, The Beatles, The Rolling Stones, The Beach Boys, and Bob Dylan each benefited from some friendly mutual competition that helped fuel further inspiration.

Finally, people's personalities shape the music they choose to listen to (Rentfrow & Gosling, 2003), how they perceive its emotionality (Chamorro-Premuzic et al., 2009, 2010), and thus, presumably, how they rate its quality. For example, people scoring higher on the Big Five trait of openness to experience—but not the other four traits—were more likely to experience emotional awe in response to an awe-inspiring song (i.e., "Hoppipolla" by Sigur Rós; Silvia et al., 2015). Indeed, given that people appear to project their personality traits onto fictional characters (Webster & Campbell, 2023), it is likely that people also give higher quality ratings to albums whose aesthetics match their personality traits (e.g., extraverts might prefer more energetic, up-tempo music and rate it more highly than introverts). Openness to experience likely influences creative scientific output by reducing behavioral thresholds for creative activity (Grosul & Feist, 2014); the same may hold true for artistic creativity.

Conclusion

Across two studies, we found convincing evidence that people—both music critics and fans—rated music artists' debut albums higher in quality than their last or most recent albums. Although this effect showed a mostly linear decline over the course of the average artist's multi-album career, it was also slightly convex, suggesting that the largest declines in quality occurred between the first and second albums, with smaller declines for each successive album. Even after controlling for these linear and quadratic effects, critics—but not fans—showed a pronounced sophomore slump bias, whereby they punished artists' second albums with unusually low ratings. Although no single theoretical perspective can definitively explain these findings, it appears that regression to the mean, the anchoring-and-adjustment heuristic, and social conformity effects, may independently or jointly contribute to the observed effects. Future research should continue to examine the quality and quantity of artistic and creative output in other domains (e.g., film, literature) and using other methods (e.g., experiments with random assignment). With all apologies to Neil Young, we found that album quality over artists' careers tends to fade away (vs burn out). We also found that the sophomore slump bias may emerge as a conformity norm among less-experienced music critics, but not at all among music fans.

Acknowledgements

Greg Webster thanks Sam Gosling for the generous use of his Austin, Texas apartment where the first draft of this article was completed during the 2023 Austin Psych Fest.

Data availablity

All data and HLM outputs for both studies are available for inspection via OSF: https://osf.io/nzxwm.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Gregory D. Webster https://orcid.org/0000-0001-7342-8444

Supplemental material

Supplemental material for this article is available online.

Notes

1. We acknowledge that there are multiple definitions of artistic quality and success, and that these are often subjective and arbitrary. For simplicity, we have presumed that critics' and fans' observed ratings of album quality reflect some form of true-score quality plus error, even though there is no true, objective measure of album quality. Similarly, we have assumed success is a multi-album recording contract.
2. Because our predictions are nomothetic (vs idiographic), we use the term "average music artist" to describe the statistical mean effect averaged across hundreds of rock musicians using multilevel models.
3. Similar quadratic effects of career stage have been observed for measures of quality (i.e., hit ratios and citation indices) among five composers from the Great American Songbook (Hass & Weisberg, 2015).
4. We ran polynomial models (linear and quadratic effects) because we have not seen exponential decay functions estimated in the context of three-level models, let alone implemented using the HLM program.

References

Anderson, C., Hildreth, J. A. D., & Howland, L. (2015). Is the desire for status a fundamental human motive? A review of the empirical literature. *Psychological Bulletin*, *141*(3), 574–601. https://doi.org/10.1037/a0038781

Asch, S. E. (1955). Opinions and social pressure. *Scientific American*, *193*(5), 31–35. https://www.jstor.org/stable/24943779

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs: General and Applied*, *70*(9), 1–70. https://doi.org/10.1037/h0093718

Baldwin, O. B. (1933). The maturation of the college student as evidenced by retests with the national council tests. *Psychological Monographs*, *44*(1), 233–262. https://doi.org/10.1037/h0093323

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, *117*(3), 497–529. https://doi.org/10.1037/0033-2909.117.3.497

Brattico, E., & Pearce, M. (2013). The neuroaesthetics of music. *Psychology of Aesthetics, Creativity, and the Arts*, *7*(1), 48–61. https://doi.org/10.1037/a0031624

Chamorro-Premuzic, T., Fagan, P., & Furnham, A. (2010). Personality and uses of music as predictors of preferences for music consensually classified as happy, sad, complex, and social. *Psychology of Aesthetics, Creativity, and the Arts*, *4*(4), 205–213. https://doi.org/10.1037/a0019210

Chamorro-Premuzic, T., Gomà-i-Freixanet, M., Furnham, A., & Muro, A. (2009). Personality, self-estimated intelligence, and uses of music: A Spanish replication and extension using structural equation modeling. *Psychology of Aesthetics, Creativity, and the Arts*, *3*(3), 149–155. https://doi.org/10.1037/a0015342

Consequence of Sound. (2020, September). *Rolling Stone updates its top 500 albums of all time list so it's no longer just white dudes*. https://consequence.net/2020/09/rolling-stone-top-500-albums

Deichmann, D., & Baer, M. (2023). A recipe for success? Sustaining creativity among first-time creative producers. *Journal of Applied Psychology*, *108*(1), 100–113. https://doi.org/10.1037/apl0001019

De Jong, K., Moerbeek, M., & Van der Leeden, R. (2010). A priori power analysis in longitudinal three-level multilevel models: An example with therapist effects. *Psychotherapy Research*, *20*(3), 273–284. https://doi.org/10.1080/10503300903376320

Duffy, R. D., Jadidian, A., Webster, G. D., & Sandell, K. J. (2011). The research productivity of academic psychologists: Assessment, trends, and best practice recommendations. *Scientometrics*, *89*(1), 207–227. https://doi.org/10.1007/s11192-011-0452-4

Ericsson, K. A., Nandagopal, K., & Roring, R. W. (2009). Toward a science of exceptional achievement: Attaining superior performance through deliberate practice. *Annals of the New York Academy of Sciences*, 1172, 199–217. https://doi.org/10.1196/annals.1393.001

Field, N. H., Choukas-Bradley, S., Giletta, M., Telzer, E. H., Cohen, G. L., & Prinstein, M. J. (2024). Why adolescents conform to high-status peers: Associations among conformity, identity alignment, and self-esteem. *Child Development*, *95*, 879–894. https://doi.org/10.1111/cdev.14038

Gawronski, B. (2019). Six lessons for a cogent science of implicit bias and its criticism. *Perspectives on Psychological Science*, *14*(4), 574–595. https://doi.org/10.1177/1745691619826015

Givi, J., & Galak, J. (2019). The "future is now" bias: Anchoring and (insufficient) adjustment when predicting the future from the present. *Journal of Experimental Social Psychology*, *84*, 103830. https://doi.org/10.1016/j.jesp.2019.103830

Grosul, M., & Feist, G. J. (2014). The creative person in science. *Psychology of Aesthetics, Creativity, and the Arts*, *8*(1), 30–43. https://doi.org/10.1037/a0034828

Hargreaves, D. J., & North, A. C. (1999). The functions of music in everyday life: Redefining the social in music psychology. *Psychology of Music*, *27*(1), 71–83. https://doi.org/10.1177/0305735699271007

Hass, R. W., & Weisberg, R. W. (2009). Career development in two seminal American songwriters: A test of the equal odds rule. *Creativity Research Journal*, *21*(2–3), 183–190. https://doi.org/10.1080/10400410902855275

Hass, R. W., & Weisberg, R. W. (2015). Revisiting the 10-year rule for composers from the Great American Songbook: On the validity of two measures of creative production. *Psychology of Aesthetics, Creativity, and the Arts*, *9*(4), 471–479. https://doi.org/10.1037/aca0000021

Holbrook, M. B., & Schindler, R. M. (1989). Some exploratory findings on the development of musical tastes. *Journal of Consumer Research*, *16*(1), 119–124. https://doi.org/10.1086/209200

Holbrook, M. B., & Schindler, R. M. (2003). Nostalgic bonding: Exploring the role of nostalgia in the consumption experience. *Journal of Consumer Behaviour: An International Research Review*, *3*(2), 107–127. https://doi.org/10.1002/cb.127

Hunter, P. G., Schellenberg, E. G., & Schimmack, U. (2010). Feelings and perceptions of happiness and sadness induced by music: Similarities, differences, and mixed emotions. *Psychology of Aesthetics, Creativity, and the Arts*, *4*(1), 47–56. https://doi.org/10.1037/a0016873

Kervyn, N., Judd, C. M., & Yzerbyt, V. Y. (2009). You want to appear competent? Be mean! You want to appear sociable? Be lazy! Group differentiation and the compensation effect. *Journal of Experimental Social Psychology*, *45*(2), 363–367. https://doi.org/10.1016/j.jesp.2008.08.006

Konečni, V. J. (2008). Does music induce emotion? A theoretical and methodological analysis. *Psychology of Aesthetics, Creativity, and the Arts*, *2*(2), 115–129. https://doi.org/10.1037/1931-3896.2.2.115

Kozbelt, A. (2008a). Longitudinal hit ratios of classical composers: Reconciling "Darwinian" and expertise acquisition perspectives on lifespan creativity. *Psychology of Aesthetics, Creativity, and the Arts*, *2*(4), 221–235. https://doi.org/10.1037/a0012860

Kozbelt, A. (2008b). One-hit wonders in classical music: Evidence and (partial) explanations for an early career peak. *Creativity Research Journal*, *20*(2), 179–195. https://doi.org/10.1080/10400410802059952

Kozbelt, A. (2011). Age and aesthetic significance in classical music: A multi-level reanalysis of Halsey's (1976) ratings. *Empirical Studies of the Arts*, *29*(2), 129–148. https://doi.org/10.2190/EM.29.2.a

Kozbelt, A. (2014). Musical creativity over the lifespan. In D. K. Simonton (Ed.), *The Wiley handbook of genius* (pp. 451–472). Wiley.

Kozbelt, A., & Meredith, D. (2011). Lifespan melodic originality trajectories in classical composers: A hierarchical linear modeling approach. *The International Journal of Creativity & Problem Solving, 21*(2), 63–79.

Ladinig, O., & Schellenberg, E. G. (2012). Liking unfamiliar music: Effects of felt emotion and individual differences. *Psychology of Aesthetics, Creativity, and the Arts, 6*(2), 146–154. https://doi.org/10.1037/a0024671

Leary, M. R. (2010). Affiliation, acceptance, and belonging: The pursuit of interpersonal connection. In S. T. Fiske, D. T. Gilbert & G. Lindzey (Eds.), *Handbook of social psychology* (pp. 864–897). John Wiley & Sons, Inc. https://doi.org/10.1002/9780470561119.socpsy002024

Lee, E., & Hong, S. (2021). Adequate sample sizes for a three-level growth model. *Frontiers in Psychology, 12*, Article 685496. https://doi.org/10.3389/fpsyg.2021.685496

Lonsdale, A. J. (2021). Musical taste, in-group favoritism, and social identity theory: Re-testing the predictions of the self-esteem hypothesis. *Psychology of Music, 49*(4), 817–827. https://doi.org/10.1177/0305735619899158

Lundy, D. E. (2010). A test of consensus in aesthetic evaluation among professional critics of modern music. *Empirical Studies of the Arts, 28*(2), 243–258. https://doi.org/10.2190/EM.28.2.h

Lundy, D. E., & Smith, J. L. (2017). It's tough to be a critic: Professional versus nonprofessional music judgment. *Empirical Studies of the Arts, 35*(2), 139–168. https://doi.org/10.1177/0276237416661989

Marsh, D., & Swenson, J. (Eds.). (1983). *The new Rolling Stone record guide*. Random House / Rolling Stone Press.

Miller, G. F. (1999). Sexual selection for cultural displays. In R. Dunbar, C. Knight & C. Power (Eds.), *The evolution of culture* (pp. 71–91). Rutgers University Press.

Neto, F. (1995). Conformity and independence revisited. *Social Behavior and Personality: An International Journal, 23*(3), 217–222. https://doi.org/10.2224/sbp.1995.23.3.217

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Sage.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. T., Jr. (2004). *HLM 6: Hierarchical linear and nonlinear modeling*. Scientific Software International, Inc.

Rentfrow, P. J., & Gosling, S. D. (2003). The do re mi's of everyday life: The structure and personality correlates of music preferences. *Journal of Personality and Social Psychology, 84*(6), 1236–1256. https://doi.org/10.1037/0022-3514.84.6.1236

Richard, F. D., Bond, C. F., Jr., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology, 7*, 331–363. https://doi.org/10.1037/1089-2680.7.4.331

Roberts, R. L. (1933). The maturation of college students as evidenced by the eight semester average grade points. *Psychological Monographs, 44*(1), 263–281. https://doi.org/10.1037/h0093325

Rolling Stone. (2003, December 11). The 500 greatest albums of all time. *Rolling Stone, 937*, 83–187. https://www.rollingstone.com/music/music-lists/500-greatest-albums-of-all-time-156826

Schäfer, T., & Sedlmeier, P. (2010). What makes us like music? Determinants of music preference. *Psychology of Aesthetics, Creativity, and the Arts, 4*(4), 223–234. https://doi.org/10.1037/a0018374

Schellenberg, E. G., & von Scheve, C. (2012). Emotional cues in American popular music: Five decades of the Top 40. *Psychology of Aesthetics, Creativity, and the Arts, 6*(3), 196–203. https://doi.org/10.1037/a0028024

Silvia, P. J., Fayn, K., Nusbaum, E. C., & Beaty, R. E. (2015). Openness to experience and awe in response to nature and music: Personality and profound aesthetic experiences. *Psychology of Aesthetics, Creativity, and the Arts, 9*(4), 376–384. https://doi.org/10.1037/aca0000028

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allow presenting anything as significant. *Psychological Science, 22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Simonton, D. K. (1977). Creative productivity, age, and stress: A biographical time-series analysis of 10 classical composers. *Journal of Personality and Social Psychology, 35*(11), 791–804. https://doi.org/10.1037/0022-3514.35.11.791

Simonton, D. K. (1997). Creative productivity: A predictive and explanatory model of career trajectories and landmarks. *Psychological Review*, *104*(1), 66–89. https://doi.org/10.1037/0033-295X.104.1.66

Simonton, D. K. (2007). Cinema composers: Career trajectories for creative productivity in film music. *Psychology of Aesthetics, Creativity, and the Arts*, *1*(3), 160–169. https://doi.org/10.1037/1931-3896.1.3.160

Stein, G. G. (2020, July 13). How much music is really released per year? *Streaming Machinery*. https://streamingmachinery.wordpress.com/2020/07/13/how-much-music-is-really-released-per-year/

Surowiecki, J. (2004). *The wisdom of crowds*. Doubleday; Anchor.

Taylor, J., & Cuave, K. L. (1994). The sophomore slump among professional baseball players: Real or imagined? *International Journal of Sport Psychology*, *25*(2), 230–239.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.112

Vuoskoski, J. K., & Eerola, T. (2012). Can sad music really make you sad? Indirect measures of affective states induced by music and autobiographical memories. *Psychology of Aesthetics, Creativity, and the Arts*, *6*(3), 204–213. https://doi.org/10.1037/a0026937

Wald, E. (2009). *How the Beatles destroyed rock "n" roll*. Oxford.

Webster, G. D., & Campbell, J. T. (2023). Personality perception in Game of Thrones: Character consensus and assumed similarity. *Psychology of Popular Media*, *12*(2), 207–218. https://doi.org/10.1037/ppm0000398

Webster, G. D., & Weir, C. A. (2005). Emotional responses to music: Interactive effects of mode, texture, and tempo. *Motivation and Emotion*, *29*(1), 19–39. https://10.1007/s11031-005-4414-0

Wicherts, J. M., Veldkamp, C. L., Augusteijn, H. E., Bakker, M., Van Aert, R., & Van Assen, M. A. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking. *Frontiers in Psychology*, *7*, Article 1832. https://doi.org/10.3389/fpsyg.2016.01832

Zackery, S. M. (2014). *Musical missteps: The severity of the sophomore slump in the music industry* (Unpublished senior thesis). Scripps College. https://scholarship.claremont.edu/scripps_theses/335/