A quantitative analysis of Beethoven as self-critic: implications for psychological theories of musical creativity

Aaron Kozbelt

Psychology of Music 2007 35: 144
DOI: 10.1177/0305735607068892

ARTICLE

A quantitative analysis of Beethoven as self-critic: implications for psychological theories of musical creativity

Psychology of Music

Psychology of Music

vol 35(1):144–168 [0305-7356 (200701) 35:1; 144–168]
10.1177/0305735607068892
http://pom.sagepub.com

AARON KOZBELT
DEPARTMENT OF PSYCHOLOGY, BROOKLYN COLLEGE OF THE CITY UNIVERSITY OF NEW YORK, USA

ABSTRACT How well can composers judge their own ideas and works? This question remains largely unanswered, in part because qualitative data sources like quotations and anecdotes have not been systematically examined. This archival case study of Ludwig van Beethoven remedies this by comprehensively analyzing his self-critical statements. Explicit self-criticisms of 70 compositions were found in Beethoven's letters or conversations, spanning his whole career and most musical forms. Beethoven's positive or negative assessments are reliably associated with three citation measures of aesthetic success, and the likelihood of correct decisions strongly increased with age. His comments comparing several similar masterpieces are likewise largely consistent with expert ratings and recording counts. Finally, the ranking of works by listener-hours (number of complete recordings multiplied by performance duration) correspond closely to Beethoven's intra-genre preferences. The results suggest considerable self-critical acumen on Beethoven's part. The results support an expertise view of musical creativity in which knowledge and experience are likely to enable both progressively greater creative accomplishments and sounder self-criticism.

KEYWORDS: *aesthetic judgment, Beethoven, creativity, expertise, self-criticism*

Introduction

How well can composers judge their own ideas and works? Composers must make many decisions and evaluations regarding an emerging piece, such as its structure, scoring, duration, and key. When a work is provisionally finished, they must make decisions about publishing, performance, and possibly revision. The ability to make sound judgments throughout the creative process would be extraordinarily beneficial. One could then invest effort wisely into ideas or works with a higher aesthetic (or financial) payoff, rather than wasting it on futile projects.

However, evaluative processes in creativity are not well understood (Runco and Chand, 1994). A priori, it is unclear if a first-person perspective is an asset or a liability in critical evaluation. Since great creators make many judgments and evaluations in the course of composing, one might think that their evaluative skills are quite well calibrated, especially for their own ideas and works. Perkins (1981: 128) encapsulated this view: 'A fundamental fact – maybe even *the* fundamental fact – about making is that critical abilities are more advanced than productive abilities.' However, a composer's own perspective is highly subjective, containing idiosyncratic memories of a work's genesis, their goals and motivations and how these may have changed, what aspects of a work were satisfying or not, etc. This plus the difficulty of predicting which ideas are likely to be well received might mean that even 'the creative genius seems unable to determine which works will earn future applause' (Simonton, 1977: 803).

Great composers do sometimes appear to misjudge their compositions. For example, Brahms called his *String Quintet in F*, op. 88, 'one of [his] finest works', despite its current position among his least popular chamber compositions (Swafford, 1997: 473–4). Mahler once referred to his *Seventh Symphony* as his 'best work' (Martner, 1979: 312), a minority opinion. Likewise, Richard Strauss curiously commented that his penultimate opera, *Die Liebe der Danae*, contained some of his finest music (Osborne, 1988: 225).

The notion that creators are often poor judges of their own work has been incorporated into a Darwinian theory that is the most thoroughly developed psychological account of lifespan creativity (Simonton, 1977, 1984, 1988, 1997, 1999, 2003a). This model highlights the fundamental role of chance in creativity, both in idea generation and subsequent reception of finished works. In this view, creators can do little to learn to generate better ideas or to learn to judge their ideas better: feedback on new creative products is rarely unanimous; even if it were, which aspects of the work contributed to success would be unclear; even if these could be identified, creators who repeatedly used the same ideas would no longer be innovative (see Simonton, 1999: 188–97). Thus, the sheer complexity of creation and critical evaluation, in which numerous factors interact in multifaceted ways, precludes reliable evaluation and decision-making.

One point of support for this position is the claim that 'Beethoven's own favorites among his symphonies, sonatas, and quartets are not those most frequently performed and recorded by posterity' (Simonton, 1977: 802–3). Again:

> It is as if even the greatest musical creators were unable to separate the grain from the chaff . . . Because the contributors themselves remember their hardships and frustrations, they will have very different favorites among their conceptions in comparison with those of their contemporaries. Creators are thus not the best judges of their own work. (Simonton, 1988: 89–90)

More recent incarnations of this perspective (e.g. Simonton, 2003a) have perhaps softened some earlier claims: creators need not have zero reliability with posterity, just not a better critical sense than, e.g., contemporary audiences. However, the complexity and low reliability of feedback is still thought to preclude creators learning to better predict the reception of creative products. Also supporting this view is the empirical finding that the proportion of great music that composers write in a given period of time (called a 'hit ratio') does not systematically change with age (Simonton, 1977). In particular, the lack of an increase in hit ratio with age strongly suggests that even great composers do not develop the self-critical acumen that would enable them to allocate progressively more compositional resources to likely masterworks.

However, this null result has recently been questioned (Kozbelt, 2004). In an analysis of 18 great composers, about two-thirds did show a significant *increase* in hit ratio with age. If hit ratio at the initial ideational level cannot be improved (as forcefully argued by Simonton, 1999), then composers must learn to recognize, judge, and develop their ideas more wisely. In other words, self-critical ability must be quite strong and should often continue to develop with additional experience. Indeed, some self-critical remarks by composers are very astute. Midway through his career, Verdi named *La Traviata* and *Rigoletto* his own favorites (Budden, 1985: 231). These remain among his most recorded and acclaimed operas (Ford, 2003; Halsey, 1976). Mozart's widow said he esteemed *Don Giovanni* above all his other works (Hildesheimer, 1982: 218), a choice basically consistent with posterity's verdict (Moles, 1966[1958]: 30). Tchaikovsky considered his *Sixth Symphony* his best work (Garden, 1973: 141), a judgment echoed in later criticism: 'the slow finale . . . is a stroke of genius which solves all the artistic problems . . . most baffling to symphonic writers since Beethoven' (Tovey, 1935–9, Vol. 2: 84).

As can be seen, the extent of composers' self-critical ability remains unclear, and other approaches could be informative. Primary source quotations of eminent creators have been of great interest to researchers, providing information on creators' subjective experiences and hinting at underlying psychological processes (Campbell, 1960; Csikszentmihalyi, 1996; Perkins, 1981). However, quotations and other kinds of historical data have typically been treated in a qualitative, anecdotal, descriptive manner (Simonton, 2003b). Naturally researchers often make informal references to such data for illustrative purposes, citing examples as incidental support for a particular point of view. However, such illustrations are limited in their scope and almost inevitably biased. Isolated, handpicked quotations or anecdotes can be used to support virtually *any* view of creativity, as the composer comments cited earlier show. To be reasonably objective and useful for testing hypotheses, such data must be examined in a thorough, systematic, quantitative way.

The present study implements this methodology by comprehensively analyzing all instances of self-criticism by Ludwig van Beethoven (1770–

1827). Beethoven was chosen for several reasons. First, he is often explicitly invoked in discussions of self-critical ability (Perkins, 1981; Simonton, 1977, 1984, 1988, 1994). Beethoven is also very well documented: virtually his entire output has been recorded and can be accurately dated (Ford, 2003). Expert ratings, citation measures, and recording counts give a quantitative basis for posterity's judgment, against which Beethoven's opinions can be compared. Moreover, his high critical standards and famously effortful compositional process provide a good test of the value of intensive self-criticism. Beethoven routinely jotted down fleeting musical thoughts, and his sketchbooks show painstaking evaluation, variation, and revision of his ideas throughout the compositional process (Cooper, 1991: 172–4). Finally, while Beethoven claimed he had made it a principle never to write anything about himself (Letter 1542, see Anderson, 1961: 1322), many letters and remarks to contemporaries contain explicit evaluations of his own compositions.

While creators may or may not have difficulty assessing the quality of their work, posthumous measures of aesthetic success tend to be reliable (Galenson, 2001; Simonton, 1987, 1998). If Beethoven's self-critical ability was sound, there should be a strong association between his judgments and those of posterity. A coarse test examines Beethoven's positive or negative assessments compared to each work's posthumous anthologization in lists of musical masterworks (Ewen, 1965; Halsey, 1976) or thematic dictionaries (Barlow and Morgenstern, 1948, 1950). The impact of factors like age or characteristics of each work on the correctness of his assessments can also be tested. A finer measure compares his judgments of the relative quality of several works with the expert ratings, recording counts, or the number of 'listener-hours' for those compositions (Moles, 1966[1958]).

The distribution of works by listener-hours can also be compared with that of the positively skewed 'Price law' distribution (Price, 1963; see also Kozbelt, 2005; Moles, 1966[1958]; Simonton, 1999). Aesthetic success typically follows such a distribution, where a few works achieve great popularity but most remain unknown. The Price law was originally formulated to account for the distribution of achievement among persons, rather than works. However, it also characterizes the popularity distributions of individual works in a creator's *oeuvre*, including e.g., Mozart's compositions (Kozbelt, 2005; see also Moles, 1966[1958]: 29–30). In the present context, the Price law predicts that the square root of the total number of a composer's works will account for half of the total performance time. Top compositions show a clear rank order that progressively muddies lower in the distribution.

Finally, these data allow comparison of two potentially dissociable aspects of aesthetic success: general popularity with the musical public versus works of great significance appealing primarily to connoisseurs (Bamossy et al., 1985; Simonton, 1986). Halsey (1976: 43) claims that this dichotomy is a 'systemic, pervasive condition' found in composers like Bartók, Debussy, and

Schönberg, and he provides examples of Beethoven works in both categories: popular attractions include the *Moonlight*, *Appassionata*, *Pathétique*, and *Kreutzer* sonatas, *Fifth Symphony*, and *Emperor Concerto*; more esoteric masterworks include the last five quartets, late piano sonatas, *Missa Solemnis*, *Third Symphony*, and *Fidelio*. If this assertion is true, works in the latter category should be considerably less recorded and heard than those in the former.

Method

INSTANCES OF SELF-CRITICISM

Several sources were consulted to gather all of Beethoven's documented self-criticisms. Of primary interest were Beethoven's letters (Anderson, 1961; Kalischer, 1969[1909]) and contemporary accounts of anecdotes and conversations (Wegeler and Ries, 1987; Sonneck 1967[1926]). Thayer's (1967) monumental biography, revised and edited by Forbes, was checked for additional cases, as were Solomon (1977) and Cooper (1991).

Only explicit instances of self-criticism were analyzed (see the Appendix for dates and representative quotations for each work). In all, self-criticisms were found for 70 compositions, spanning Beethoven's whole career and most musical forms and accounting for 34 percent of his output by performance duration. Compositions were counted based on the standard opus and WoO cataloguing systems. Sonata form works sharing an opus number (e.g., three trios, op. 1) were counted separately.

AESTHETIC SUCCESS CRITERIA

Posthumous evaluations were gauged by several criteria. First is Halsey (1976), which purports to list and rate the several thousand classical music compositions with a secure place in the repertoire. Works are rated on a 5-point aesthetic significance scale, with '1' representing the best. These have often been used as a measure of aesthetic success or criterion for a musical masterpiece (Kozbelt, 2004, 2005; Simonton, 1986, 1987, 1991, 1995). Halsey lists 134 compositions by Beethoven, which account for 67 percent of his music in terms of performance duration. A similar measure of posthumous aesthetic success is inclusion in Ewen (1965), who, like Halsey, lists specific major works of great composers. Ewen lists 87 Beethoven compositions, accounting for 52 percent of his music by performance duration. Barlow and Morgenstern's (1948, 1950) thematic dictionaries represent another anthologization measure. The dictionaries include themes from 134 works by Beethoven, accounting for 59 percent of his music by performance duration.

Recording counts measure composition popularity, another index of aesthetic success. Counts were taken from the *RED Classical Catalog* (Ford, 2003), which lists all commercially available classical music recordings. Only recordings of complete works (not excerpts) were counted here.

LISTENER-HOURS

The number of 'listener-hours' for each work was computed (cf. Moles, 1966[1958]: 28–30). This measure multiplies the duration of each work by the number of complete recordings of the work listed by Ford (2003). It thus takes into account both popularity (number of recordings) and a rough index of compositional effort (performance duration), and therefore is more informative than raw recording counts. Listener-hours indicate how frequently each work is likely to be heard by the musical public, although works whose highlights are frequently excerpted will be somewhat underestimated. Listener-hours also reflect finer distinctions between works than binary anthologization measures. They are thus more useful in evaluating comments on the relative quality of two masterpieces. The distribution of works by listener-hours can also be compared to the Price law, which predicts a dominance of and strong rank ordering among a creator's most popular productions.

Results

SELF-CRITICISM AND MASTERPIECE STATUS

Beethoven's self-criticisms are usually clearly positive or negative. Posterity's verdicts are also generally straightforward. Chi-square tests of independence between each citation source (Barlow and Morgenstern, 1948, 1950; Ewen, 1965; Halsey, 1976) are uniformly very significant and show large effect sizes, indicating strong consistency across the three measures. For Halsey versus Ewen, χ^2 (1) = 192, $p < .0001$, $\phi = .67$; for Halsey versus Barlow and Morgenstern, χ^2 (1) = 215, $p < .0001$, $\phi = .71$; for Ewen versus Barlow and Morgenstern, χ^2 (1) = 169, $p < .0001$, $\phi = .63$. These three sources thus provide simple, reliable assessments of masterpiece status. If Beethoven was an astute self-critic, his assessments should show a reliable association with each citation measure.

Table 1 categorizes the 70 relevant compositions into positive and negative assessments by Beethoven and by Halsey (1976). A chi-square test of independence between the two was highly significant, χ^2 (1) = 15.1, $p < .001$, with a fairly large effect size, $\phi = .46$. Chi-squared tests comparing Beethoven's assessments with inclusion (or not) in Ewen (1965) and Barlow and Morgenstern (1948, 1950) were also performed. Each yielded significant results with medium effect sizes, χ^2 (1) = 7.2, $p < .01$, $\phi = .32$, and χ^2 (1) = 6.1, $p < .05$, $\phi = .29$, respectively. Tables 2 and 3 show the categorization of works by these criteria.

Because overall agreement is high, instances of discord are of special interest. By all three criteria, most false alarms are early works assessed by Beethoven near the time of composition, e.g., calling the op. 9 string trios his best works. The trios are not cited in any of the three anthologies, although Solomon (1977: 100) notes, 'more than one critic has agreed with

TABLE 1 *Contingency table of Beethoven's self-criticisms versus those of Halsey (1976)*

	Halsey: cited	Halsey: not cited	Marginals
Beethoven: positive assessment	38 hits: opp. 1.1, 1.2, 1.3, 46, 21, 22, 37, 28, 34, 35, 121a, 55, 57, 72, 58, 59.1, 59.2, 59.3, 60, 61, 62, 86, 78, 92, 93, 101, 106, 109, 110, 111, 123, 125, 126, 127, 132, 130, 133, 131	6 false alarms: WoO 65, opp. 9.1, 9.2, 9.3, WoO 74, op. 45	44
Beethoven: negative assessment	11 misses: opp. 44, 19, 15, 20, 27.2, 43, 85, WoO 80, opp. 138, 117, 91	15 correct rejections: WoO 38, 40, 41, 28, op. 71, WoO 136, 137, 18, 19, 20, op. 113, WoO 2, op. 115, WoO 24, 3	26
Marginals	49	21	70

Note: Works listed in chronological order within each cell by opus or WoO number. See Appendix for representative quotations.

TABLE 2 *Contingency table of Beethoven's self-criticisms versus those of Ewen (1965)*

	Ewen: cited	Ewen: not cited	Marginals
Beethoven: positive assessment	29 hits: opp. 46, 21, 37, 28, 55, 57, 72, 58, 59.1, 59.2, 59.3, 60, 61, 62, 86, 92, 93, 101, 106, 109, 110, 111, 123, 125, 127, 132, 130, 133, 131	15 false alarms: WoO 65, opp. 1.1, 1.2, 1.3, 9.1, 9.2, 9.3, WoO 74, opp. 22, 34, 35, 121a, 45, 78, 126	44
Beethoven: negative assessment	8 misses: opp. 19, 71, 15, 20, 27.2, 43, 85, 138	18 correct rejections: WoO 38, op. 44, WoO 40, 41, 28, 80, 136, 137, 18, 19, 20, 24, opp. 113, 117, WoO 2, opp. 91, 115, WoO 3	26
Marginals	37	33	70

Note: Works listed in chronological order within each cell by opus or WoO number. See Appendix for representative quotations.

TABLE 3 *Contingency table of Beethoven's self-criticisms versus citation in Barlow and Morgenstern (1948, 1950)*

	Barlow and Morgenstern: cited	Barlow and Morgenstern: not cited	Marginals
Beethoven: positive assessment	31 hits: opp. 1.3, 46, 21, 22, 37, 28, 55, 57, 72, 58, 59.1, 59.2, 59.3, 60, 61, 62, 78, 92, 93, 101, 106, 109, 110, 111, 123, 125, 127, 132, 130, 133, 131	13 false alarms: WoO 65, opp. 1.1, 1.2, 9.1, 9.2, 9.3, WoO 74, opp. 34, 35, 121a, 45, 86, 126	44
Beethoven: negative assessment	10 misses: opp. 19, 71, 15, 20, 27.2, 43, 85, WoO 80, opp. 138, 113	16 correct rejections: WoO 38, op. 44, WoO 40, 41, 28, 136, 137, 18, 19, 20, op. 117, WoO 2, op. 91, op. 115, WoO 24, 3	26
Marginals	41	29	70

Note: Works listed in chronological order within each cell by opus or WoO number. See Appendix for representative quotations.

[Beethoven's] judgment.' Other false alarms, by two criteria, comprise early chamber works, opp. 1.1, 1.2, and 121a, or piano variations or bagatelles, opp. 34, 35, and 126, which as a genre are less popular than sonatas (see Kozbelt, 2005).

Reasons for consistent misses are more varied. For instance, Beethoven negatively rated his first two piano concertos, especially op. 19, primarily by comparison to his third. The first two have fewer recordings, and Halsey (1976) gives op. 19 the lowest rating among the piano concerti, consistent with Cooper's (1991: 160) assertion that it 'is still considered his weakest'. Beethoven's evaluations of the *Septet*, op. 20, and *Moonlight Sonata*, op. 27.2, likewise seem motivated by his aversion to some early works' excessive popularity (Cooper, 1991: 158–9). His negative appraisal of *Die Geschöpfe des Prometheus*, op. 43, may be attributed to the context of a conversation with his former teacher, Haydn, and comparison with Haydn's oratorio, *The Creation*. Beethoven's distaste for his oratorio, *Christus am Ölberge*, op. 85, may be partly due to its poor text (Letter 323); Solomon (1977: 191–2) suggests it may also have painful associations with the near-contemporary Heiligenstadt Testament. Finally, Beethoven repeatedly disparaged *Wellingtons Sieg*, op. 91, perhaps his weakest mature work (Cooper, 1991: 220; Solomon, 1977: 222–4). Neither Ewen (1965) nor Barlow and Morgenstern (1948) cite it; Halsey rates it very low.

Note also that in each analysis, one ambiguous case, three marches, op. 45, was conservatively classified as a false alarm. Beethoven praised the marches in a letter to music publishers, but one was written quickly while Beethoven was giving a lesson, suggesting he did not take it too seriously. Beethoven typically judged hastily written works poorly, e.g., WoO 65, 73, 61, and 70 and op. 17 (for details see, respectively, Thayer, 1967: 125, 215, 776; Wegeler and Ries, 1987: 67, 71). Of these, only the Horn sonata, op. 17, is cited (Barlow and Morgenstern, 1948, Halsey, 1976). Along these lines, one might infer Beethoven's opinions from his choices of which works to publish (Thayer, 1967: 277), though naturally published works are much more widely available than non-published ones, an important confound.

MULTIVARIATE ANALYSES

These qualitative observations can be buttressed with multivariate analyses to identify which factors predict accurate self-criticisms. For instance, did Beethoven's self-critical acumen change with age? Was he a better self-critic for certain kinds of works (e.g. instrumental versus vocal works, or longer versus shorter works)? Does the interval between composition and self-criticism or the source of the self-criticism (e.g. Beethoven's own letters versus others' anecdotal reports) matter?

To investigate these questions, logistic regression analyses were performed to predict correct decisions (hits and correct rejections) versus incorrect decisions (false alarms and misses), using several predictor variables. First, the year of the earliest comment on each composition was noted. Since the dates of several comments could not be ascertained, six works (opp. 27.2, 28, 57, 78, 93, and WoO 80) were excluded. (Composition year could also be used. It is very strongly correlated with comment year, $r\,(62) = .86$, $p < .001$, and yielded very similar results.) The interval between composition and criticism, in years, was also recorded: longer intervals may have afforded Beethoven the opportunity to receive audience feedback. Unfortunately, a direct measure of the influence of public opinion on Beethoven's self-criticisms could not be derived, since precise first performance and publication dates of many of his works are unknown. The performance duration of each work was also recorded. Finally, two dummy variables coded whether each composition was an instrumental work (coded as '1') or a vocal work ('0'), and whether Beethoven himself made the self-critical comment in a letter ('1') or if it was an anecdote reported by someone else ('0').

Logistic regression results for each dependent measure are shown in Table 4. As can be seen, each analysis yielded highly significant age effects. No other variables were significant, suggesting that Beethoven's critical acumen was consistent across different kinds of music, and that his contemporaries' anecdotal reports are comparable to Beethoven's own written statements. In sum, not only are Beethoven's self-criticisms generally quite accurate, but this ability seems to have improved considerably throughout his career.

TABLE 4 *Results of logistic regression analyses predicting Beethoven's accurate self-criticisms according to each aesthetic success measure*

	B	SE	Wald χ^2	*df*	*p*	Odds ratio
Halsey						
Comment year	.121	.043	8.034	1	.005	1.128
Interval	–.049	.076	0.404	1	.525	0.953
Duration	–.003	.018	0.023	1	.879	0.997
Instrumental?	.315	.951	0.110	1	.740	1.371
Quotation source	.993	1.008	0.970	1	.325	2.699
Constant	–206.369	77.588	7.075	1	.008	0.000
Cox and Snell R^2 = .172						
Nagelkerke R^2 = .264						
Ewen						
Comment Year	.162	.046	12.513	1	<.001	1.176
Interval	–.089	.068	1.738	1	.187	0.914
Duration	–.011	.018	0.344	1	.558	0.989
Instrumental?	–.436	.959	0.206	1	.650	0.647
Quotation source	.355	1.123	0.100	1	.752	1.427
Constant	–291.797	82.798	12.420	1	<.001	0.000
Cox and Snell R^2 = .293						
Nagelkerke R^2 = .408						
Barlow and Morgenstern						
Comment Year	.125	.039	10.497	1	.001	1.137
Interval	–.078	.063	1.544	1	.218	0.925
Duration	–.008	.016	0.223	1	.637	0.992
Instrumental?	.282	.871	0.105	1	.746	1.325
Quotation source	.387	.994	0.152	1	.697	1.473
Constant	–225.577	69.828	10.436	1	.001	0.000
Cox and Snell R^2 = .209						
Nagelkerke R^2 = .291						

PERIOD-BY-PERIOD ANALYSES OF SELF-CRITICAL ABILITY

While these results suggest that Beethoven was a good self-critic, the overall effects are perhaps no greater than the agreement between contemporary works' reception and later aesthetic success measures. For instance, the initial success of 496 operas correlated quite well with various archival measures of each opera's ultimate aesthetic success, average $r = .48$ (Simonton, 1998), comparable to the effects reported for Beethoven above (ϕ is mathematically equivalent to the absolute value of the Pearson correlation coefficient between two dichotomous dummy-coded variables; thus ϕ^2, like r^2, also yields the proportion of variance the two variables share). This might imply that Beethoven was not a better self-critic than contemporary audiences, if Simonton's (1998) estimate of contemporary reception is typical. However, since age is so strongly associated with Beethoven's self-critical accuracy, one can refine the effect size estimates by calculating them separately for different periods of Beethoven's career.

Beethoven's career is often divided into three stylistic 'periods,' but standard cut-points are not agreed upon (Cooper, 1991: 198–200). Beethoven's early period was defined as 1791 to 1802, his middle period as 1803 to 1814, and his late period as 1815 and later. This captures several inaugurating landmarks, e.g. the *Third Symphony* launching the middle period and the last piano sonatas (beginning with op. 101), the late period. Alternatively, Beethoven's career was divided into only two periods, corresponding to the first half (1791 to 1808) and second half (1809 to 1826) of the range of dates which his comments span. This improves statistical power and avoids violating assumptions about minimum cell counts in chi-square tests (Delucchi, 1983). Finally, period categorization can be based on dates of composition or self-criticism: in the present study, both were used.

Several undatable self-criticisms could be assigned to career period based on other information. For instance, the *Eighth Symphony* was written in 1811–12, so any comments must have been made after 1809. This allowed the *Eighth* to be classified in the two-period analysis, but not the three-period analysis. For the three-period analysis, reasonable estimates of comment dates for six works (opp. 27.2, 28, 57, 78, 93, and WoO 80) could not be established; they were thus excluded. For the two-period analysis, three works (op. 28, op. 57, and WoO 80) could not be classified; these were likewise dropped.

Separate chi-square analyses were calculated on each dependent measure, for each career period, for both composition and comment dates. Results are shown in Table 5. They are consistent with the strong age effect found in the logistic regression analyses. In Beethoven's early period, his assessments showed a uniformly negligible association with those of posterity. However, in his last period, his assessments were uniformly very significantly associated with posterity's. Indeed, effect sizes here were extremely high, mostly .8 or

TABLE 5 *Results of chi-square analyses of Beethoven's self-criticisms, divided into different periods of his career*

Category (dates)	*N*	Halsey χ^2	Halsey ϕ	Ewen χ^2	Ewen ϕ	Barlow and Morgenstern χ^2	Barlow and Morgenstern ϕ
Compositions sorted by date							
Early (1791–1802)	26	0.4	.12	2.1	.28	0.5	.14
Middle (1803–14)	28	9.1	.58	11.5	.65	6.6	.49
Late (1815–26)	16	16.0	1.00	11.1	.83	11.1	.83
First half (1791–1808)	41	0.9	.14	0.1	.05	0.3	.08
Second half (1809–26)	29	21.6	.86	21.9	.87	21.3	.86
Self-criticisms sorted by date							
Early (1791–1802)	20	0.3	.12	2.3	.34	0.6	.18
Middle (1803–14)	19	1.6	.29	3.4	.42	1.1	.24
Late (1815–26)	25	22.0	.94	19.0	.87	16.0	.80
First half (1791–1808)	34	0.4	.11	0.1	.06	0.0	.02
Second half (1809–26)	33	22.4	.82	15.9	.69	15.6	.69

Note: All $df = 1$. $\chi^2 > 3.8$, $p < .05$; $\chi^2 > 6.6$, $p < .01$; $\chi^2 > 10.8$, $p < .001$.

above. This indicates that on most measures Beethoven's own judgments account for at least about two-thirds of the variance (ϕ^2) in posterity's judgments. Indeed, when comments were categorized by composition date and ϕ^2 averaged across the three measures, ϕ^2 rose from .032 to .329 to .786 across Beethoven's three periods. The most discrepant results in Table 5 are the weak associations in Beethoven's middle period when comment dates were analyzed. This seems at least partly due to diminished statistical power, since undatable self-criticisms reduced the number of quotations available for analysis below the 20 meeting the assumptions of the chi-square test (Delucchi, 1983).

In general, these results reinforce Beethoven's self-critical acumen and reveal that this ability increased substantially throughout his career. By his late period, Beethoven's positive or negative appraisals of his compositions predicted those of posterity very closely.

FINER DISTINCTIONS AMONG MASTERPIECES: RECORDING COUNTS AND HALSEY RATINGS

While the preceding results suggest Beethoven was a reliable self-critic overall, and that he improved in this capacity with age, binary comparisons between masterworks and non-masterworks represent a relatively basic metric. In practice, masterworks can range from short recital favorites like *Für Elise* to 'the great affirmative works of our culture' (Solomon, 1977: 315)

like the late quartets and the *Ninth Symphony*. Moreover, since all three citation measures include most sonata form works (symphonies, concerti, sonatas, and quartets), Beethoven could give the illusion of perspicacity by praising any of these, if he suspected they would form the core of his later reputation.

Are Beethoven's appraisals comparing several similar pieces, all of which could be considered masterworks, also consistent with posterity's verdicts? To investigate this, Halsey's 5-point rating scale and recording counts (Ford, 2003) were examined. As with his first three piano concerti, discussed earlier, many of Beethoven's relative judgments anticipate later criticism and relative popularity. For instance, of the op. 1 piano trios, Beethoven preferred the third, which is the most recorded (23 vs 19 and 14) and at least as highly rated by Halsey (see Appendix). Likewise, Beethoven judged his last set of *Bagatelles*, op. 126, his best. Compared to his other sets, opp. 33 and 119, op. 126 is rated higher by Halsey (1 vs 3 and 3) and is more recorded (31 vs 14 or 16, respectively). Finally, Beethoven made superlative remarks about several late string quartets (see Appendix), all of which outrank his earlier efforts in Halsey ratings (1 for each late quartet, compared to 3 or 4 for early, and 2 for middle, quartets). The five late quartets (M = 40.8, SD = 3.6) are also significantly more recorded than the 11 earlier ones (M = 33.4, SD = 4.6), $t(14) = 3.18$, $p < .01$.

While many comparisons of recording counts and Halsey ratings are largely consistent with the earlier findings, there are several exceptions. For instance, Beethoven once remarked that his Piano sonata No. 24 in F-sharp, op. 78, was superior to the higher-rated (2 vs 3) and more recorded (122 vs 37) *Moonlight* sonata. Likewise, he commented that his *Eighth Symphony* was better than the *Seventh*. The *Seventh* is more recorded than the *Eighth* (134 vs 87), and the two have equal Halsey ratings (2). However, it is unclear how to interpret this enigmatic single statement, compared to multiple letters explicitly praising the *Seventh* and not mentioning the *Eighth*. Finally, Beethoven implied that the *Grosse Fuge*, op. 133, was superior to the rest of his op. 130 quartet. While Halsey rates op. 130 higher (1 vs 3), op. 133 is more often recorded (47 vs 44). Note, however, that these three remarks are all conversational asides apparently made when Beethoven was upset or angry, compared to numerous letters documenting most examples in the previous paragraph.

LISTENER-HOURS

Expert ratings and recording counts permit some analyses, but listener-hours allow more sensitive and consistent comparisons of Beethoven's most popular works. They can also test the Price law and Halsey's (1976) distinction between popular and esoteric masterpieces. To examine their historical robustness, computed listener-hours were compared to entries on an earlier list (Moles, 1966[1958]: 29). The correlation was positive and

TABLE 6 *Beethoven's 30 most popular compositions, ordered by listener-hours*

Composition	Complete recordings	Listener-hours	Total percent
Symphony No.9 in D minor, op. 125 'Choral'	132	146.8	4.83
Symphony No.3 in E-flat, op. 55 'Eroica'	142	112.5	3.70
Piano Concerto No.5 in E-flat, op. 73 'Emperor'	147	98.7	3.25
Symphony No.5 in C minor, op. 67	157	95.0	3.12
Symphony No.7 in A, op. 92	134	89.0	2.93
Violin Concerto in D, op. 61	118	85.9	2.83
Symphony No.6 in F, op. 68 'Pastoral'	112	81.9	2.69
Piano Concerto No.4 in G, op. 58	115	67.2	2.21
Piano Concerto No.3 in C minor, op. 37	103	64.3	2.11
Missa Solemnis in D, op. 123	44	58.8	1.93
Piano Sonata No.29 in B-flat, op. 106 'Hammerklavier'	70	58.5	1.92
Fidelio, op. 72	31	58.1	1.91
Symphony No.4 in B-flat, op. 60	97	54.8	1.80
Diabelli Variations, op. 120	51	51.9	1.71
Piano Sonata No.23 in F minor, op. 57 'Appassionata'	118	51.8	1.70
Piano Concerto No.1 in C, op. 15	89	51.7	1.70
Violin Sonata No.9 in A, op. 47 'Kreutzer'	79	49.0	1.61
Symphony No.2 in D, op. 36	88	48.8	1.61
Piano Sonata No.32 in C minor, op. 111	95	48.2	1.58
Piano Concerto No.2 in B-flat, op. 19	82	41.3	1.36
Symphony No.1 in C, op. 21	93	41.0	1.35
Piano Sonata No.23 in C, op. 53 'Waldstein'	87	39.8	1.31
Symphony No.8 in F, op. 93	87	38.4	1.26
Piano Sonata No.31 in A-flat, op. 110	95	34.7	1.14
Piano Sonata No.14 in C-sharp minor, op. 27.2 'Moonlight'	122	33.5	1.10
Piano Sonata No.30 in E, op. 109	84	31.2	1.02
String Quartet No.15 in A minor, op. 132	38	30.7	1.01
Piano Sonata No.8 in C minor, op. 13 'Pathétique'	93	30.3	0.99
String Quartet No.14 in C-sharp minor, op. 131	45	29.7	0.98
String Quartet No.13 in B-flat, op. 130	44	28.7	0.94

Note: Listener-hours equal the performance time of each work multiplied by the number of complete recordings (Ford, 2003). Total listener-hours: 3041.3.

significant, $r\,(25) = .53$, $p < .01$, suggesting that Beethoven's most popular works remain fairly stable over time.

The resulting distribution was also compared to predictions of the Price law. The square root of total compositions (here, 421) should account for half of the listener-hours (here, 3041.3). The square root of 421 is approximately 20.5. The top 21 compositions account for 1455.2 listener-hours, 47.8% of the total, close to the predicted 50 percent. The upper tail is also consistent with the Price law, as can be seen in Table 6. Specifically, differences in listener-hours between adjacent works decrease lower on the list. Thus, a

clear rank order exists for top works; further down, differences become small and not meaningful. Indeed, among the 30 bottom compositions, the average listener-hour difference between adjacent pieces is less than 2 seconds.

Beethoven never provided an overall ranking of his compositions, but his opinions about his best works in several genres reflect the ordering in Table 6 remarkably well. For instance, the *Ninth Symphony* has the most listener-hours, followed by the *Third*. These were Beethoven's two most valued symphonic creations (see Appendix), preferred even over the *Fifth*, which ranks fourth overall. Next is the *Seventh*, repeatedly noted by Beethoven as one of his best works.

The piano sonatas and string quartets show similar patterns. Beethoven's most esteemed sonata, op. 106, dominates by a wide margin, followed by op. 57, the one he had previously thought best. Notably, his last two sonatas, also self-rated very highly, as well as op. 53, outrank the *Moonlight* in listener-hours. Beethoven's self-proclaimed greatest quartet, op. 131, ranks a very close second after op. 132. The difference is about one listener-hour, equivalent to one or two recordings of either work. Indeed, the close ranking of several late quartets reflects Beethoven's opinion that each was great in its own way. Thus, the claim that Beethoven's personal favorite symphonies, sonatas, and quartets are not necessarily the most performed or recorded (Simonton, 1977, 1984, 1988) does not seem entirely accurate, at least when performances are measured in terms of listener-hours.

One might object that the *Missa Solemnis*, Beethoven's choice as his greatest composition overall, ranks only 10th in listener-hours. However, it meets Simonton's (1991) operationalization of best work, i.e. the highest-rated work by Halsey (1976) with the longest performance time. Its recording count is probably handicapped by the large performance forces required, and it does not benefit from inclusion in complete sets of symphonies, as does the similarly scored *Ninth*.

Finally, the listener-hour measure does not neatly subdivide into Halsey's (1976: 43) two categories, popular favorites versus esoteric masterworks. The categories greatly overlap and virtually all of the works mentioned by Halsey appear in Table 6. 'Esoteric' works like the *Missa Solemnis, Third Symphony*, and *Fidelio* all occupy high places. The case for a dichotomy is somewhat stronger when raw recording counts are examined, but inter-genre comparisons are problematic (e.g. one expects fewer recordings of masses or quartets than symphonies). However, the piano sonatas in Halsey's two groups are comparable in popularity. Indeed, the last three sonatas, supposedly appreciated mainly by connoisseurs, have even more complete recordings than the 'popular' *Pathétique*.

In sum, these analyses quantitatively support Cooper's (1991: 160) assertion that Beethoven 'was well aware of the disparity between his best and his more mediocre works, but it is significant that the works he considered his best were usually the same as the most highly regarded today'.

Discussion

These findings quantitatively validate Beethoven's reputation as an astute self-critic. Posthumous aesthetic success criteria are strongly associated with Beethoven's positive or negative assessments, especially in the latter part of his career. Beethoven's comments comparing several similar masterpieces are likewise largely consistent with expert ratings and recording counts. Finer-grained analyses show that his preferences closely parallel works' rank ordering by listener-hours. Apparent lapses in judgment often reflect the trajectory of Beethoven's artistic development. False alarms typically represent contemporary assessments of early works. Misses sometimes involve comparisons of earlier with later works or early works whose popularity annoyed Beethoven. In 1823, a visitor noted that Beethoven 'cannot bear to hear his own earlier works praised . . . His latest productions . . . are his favorites' (Thayer, 1967: 871). The logistic regression analyses quantitatively demonstrate that Beethoven's self-critical ability improved considerably with age.

One might object that Beethoven may have been biased in which works he chose to critique. While this is difficult to know for sure, this study explicitly attempted to overcome any bias by examining *all* extant self-criticisms. The analyzed comments span Beethoven's entire career and most musical forms and encompass 34 percent of his output by performance time, and there is no compelling evidence indicating any idiosyncrasies in the choice of works. Alternatively, one might argue that many remarks' contexts (e.g. comments to associates in moments of anger, letters persuading publishers to accept a new work, or offhand remarks downplaying occasional or early compositions) preclude their usefulness. If anecdotal comments truly are worthless as data, they should probably be used extremely sparingly, even as illustrations. However, this would be counterproductive to the enterprise of understanding the nature of creativity. As Nobel laureate Peter Medawar (1991: 85) noted about creativity research, investigators should 'take evidence from all quarters likely to be informative, not excluding introspection, for no good will come of self-righteously abjuring such an important source of evidence'. If self-criticism data are to be used, the present methodology provides a model and framework for their analysis. Here a concerted effort was made to analyze only explicit, documented self-criticisms. Criticisms merely implied by other authors were excluded from the investigation. Certainly, some important lacunae remain, e.g. Beethoven's silence on the very popular *Fifth* and *Sixth Symphonies* and *Fifth Piano Concerto*. However, Beethoven's sheer effort on these works implies a high opinion of them. He worked on the symphonies for several years, even sketching 14 variations of the theme of the *Fifth*'s slow movement (Bernstein, 1959: 74).

The results show that qualitative self-criticisms by eminent creators can be

used to test quantitative statistical hypotheses about creativity, provided the comments are analyzed comprehensively. While explicit self-criticisms provide some information about creators' evaluative abilities, they are probably most useful as a complement to quantitative behavioral indices, such as hit ratio (Kozbelt, 2004, 2005; Simonton, 1977), which may better reflect creators' opinions about the relative value of their works. (See Simonton, 2003b, for a discussion of qualitative versus quantitative approaches to historical data.) For instance, Beethoven's hit ratio increased significantly throughout his career, even when the analysis was restricted to his compositional maturity, which onset was defined by his first Halsey (1976) masterpiece (Kozbelt, 2004). Qualitative comments may particularly help understand fallow periods in a career: hit ratio or work quality may drop when a composer writes profitable occasional pieces, as Beethoven did in his mid-40s. Beethoven's remarks on such works, like *Wellingtons Sieg*, suggest he had no illusions about their quality (Solomon, 1977: 222–4).

How did Beethoven evaluate so well? One possibility is that contrary to the stereotype of the isolated genius, he was not above heeding the advice of trusted associates, e.g. on the op. 1 trios (Wegeler and Ries, 1987: 58), *Waldstein* sonata and *Andante favori* (Thayer, 1967: 351), *Fidelio* (Wegeler and Ries, 1987: 92), and *Grosse Fuge* and op. 130 quartet (Solomon, 1977: 323). A more provocative possibility stems from a recent typology of creators (Galenson, 2001, 2004). This typology distinguishes 'finders' (or conceptual innovators) from 'seekers' (or experimentalists). In music, finders typically work deductively from some aesthetic theory or extra-musical consideration, revise relatively little as they work, and often make conceptual innovations rather early in their careers. In contrast, 'seekers' are likely more concerned with musical (as opposed to extra-musical or programmatic) considerations, are likely to constantly revise and rework their ideas, and create their greatest works later in their careers. In many respects, Beethoven's working methods and career trajectory resemble the latter type. The present results suggest that the working methods of experimentalists provide them with a rather reliable context that is associated with or enables sounder self-criticism.

One cannot make this point better than Thayer (1967: 380–1), citing Otto Jahn:

> One is amazed at this everlasting experimentation and cannot conceive how it will be possible to create an organic whole out of such musical scraps. But if one compares the completed art-work with the chaos of sketches one is overwhelmed with wonder at the creative mind which surveyed its task so clearly, grasped the foundation and the outlines of the execution so firmly and surely that with all the sketches and attempts in details the whole grows naturally from its roots and develops. And though the sketches frequently create the impression of uncertainty and groping, admiration comes again for the marvelously keen self-criticism, which, after everything has been tested with sovereign certainty, retains the best.

The psychological mechanisms of self-criticism remain unexplained. However, the present results show that such a mechanism can exist, and that, at least in Beethoven's case, it improved with age. Thus, to the best of our documented historical knowledge, and consistent with quantitative analyses of hit ratios (Kozbelt, 2004) and qualitative assessments by music critics (Cooper, 1991), Beethoven was an astute self-critic. Establishing the more general, if not universal, existence of very reliable self-evaluation processes would have important implications for theories of musical creativity. First, it would create complications for a Darwinian view (Simonton, 1977, 1988, 1997, 1999, 2003a), which postulates that even eminent creators cannot evaluate the quality of their own ideas and works very reliably (at least no more reliably than contemporary audiences), and especially that they cannot improve in this capacity. Even if other aspects of the theory are sound, the present data suggest that Beethoven may be an inappropriate choice as an illustration of limited self-critical ability (Simonton, 1977, 1984, 1988, 1994; cf. Weisberg, 2004).

Establishing the existence of quite reliable self-critical processes also presents a challenge to an expertise view of creativity (e.g. Ericsson, 1999; Hayes, 1989; Weisberg, 1999). Namely, what kinds of learning mechanisms account for the development of sound self-criticism? What is lacking when eminent creators show poor self-criticism? As noted earlier, Galenson's (2001, 2004) taxonomy may provide the start of an explanation. Since classical musicians are among the best-documented creators (Simonton, 1977), analyzing other eminent composers should be reasonably straightforward, and this study provides a methodology for doing so. It is an open question how far the present results would generalize.

ACKNOWLEDGEMENTS

Two anonymous referees provided very helpful comments that greatly improved the quality of this article.

REFERENCES

Anderson, E. (ed.) (1961) *The Letters of Beethoven*, 3 vols. London: Macmillan.
Bamossy, G., Johnston, M. and Parsons, M. (1985) 'The Assessment of Aesthetic Judgment Ability', *Empirical Studies of the Arts* 3: 63–79.
Barlow, H. and Morgenstern, S. (1948) *A Dictionary of Musical Themes*. New York: Crown.
Barlow, H. and Morgenstern, S. (1950) *A Dictionary of Opera and Song Themes.* New York: Crown.
Bernstein, L. (1959) *The Joy of Music*. New York: Simon and Schuster.
Budden, J. (1985) *Verdi*. New York: Vintage Books.
Campbell, D.T. (1960) 'Blind Generation and Selective Retention in Creative Thought as in Other Thought Processes', *Psychological Review* 67: 380–400.

Cooper, B. (ed.) (1991) *The Beethoven Compendium.* Ann Arbor, MI: Borders Press.
Csikszentmihalyi, M. (1996) *Creativity.* New York: HarperCollins.
Delucchi, K.L. (1983) 'The Use and Misuse of Chi-Square: Lewis and Burke Revisited', *Psychological Bulletin* 94: 166–76.
Ericsson, K.A. (1999) 'Creative Expertise as Superior Reproducible Performance: Innovative and Flexible Aspects of Expert Performance', *Psychological Inquiry* 10*:* 329–33.
Ewen, D. (ed.) (1965) *The Complete Book of Classical Music*. Englewood Cliffs, NJ: Prentice-Hall.
Ford, G. (ed.) (2003) *The RED Classical Catalog*. London: RED Publishing.
Galenson, D.W. (2001) *Painting Outside the Lines: Patterns of Creativity in Modern Art.* Cambridge, MA: Harvard University Press.
Galenson, D.W. (2004) 'The Life Cycles of Modern Artists: Theory and Implications', *Historical Methods* 37: 123–36.
Garden, E. (1973) *Tchaikovsky*. London: J.M. Dent.
Halsey, R.S. (1976) *Classical Music Recordings for Home and Library*. Chicago: American Library Association.
Hayes, J.R. (1989) 'Cognitive Processes in Creativity', in J.A. Glover, R.R. Roning and C.R. Reynolds (eds) *Handbook of Creativity*, pp. 202–19. New York: Plenum.
Hildesheimer, W. (1982) *Mozart*, trans. M. Faber. New York: Vintage Books.
Kalischer, A.C. (1969[1909]) *Beethoven's Letters: A Critical Edition with Explanatory Notes*, 2 vols, trans. J.S. Shedlock, with preface). Freeport, NY: Books for Libraries Press.
Kozbelt, A. (2004) 'Reexamining the Equal-Odds Rule in Classical Composers', in J.P. Frois, P. Andrade and J.F. Marques (eds) *Art and Science: Proceedings of the XVIII Congress of the International Association of Empirical Aesthetics*, pp. 540–3. Lisbon: IAEA.
Kozbelt, A. (2005) 'Factors Affecting Aesthetic Success and Improvement in Creativity: A Case Study of the Musical Genres of Mozart', *Psychology of Music* 33: 235–55.
Martner, K. (ed.) (1979) *Selected Letters of Gustav Mahler*. New York: Faber & Faber.
Medawar, P.B. (1991) 'Creativity – Especially in Science', in P.B. Medawar and David Pyke (eds) *The Threat and the Glory*, pp. 83–90. Oxford: Oxford University Press.
Moles, A. (1966[1958]) *Information Theory and Esthetic Perception*, trans. J.E. Cohen. Urbana: University of Illinois Press.
Osborne, C. (1988) *The Complete Operas of Richard Strauss*. New York: Da Capo.
Perkins, D.N. (1981) *The Mind's Best Work*. Cambridge, MA: Harvard University Press.
Price, D. (1963) *Big Science, Little Science*. New York: Columbia University Press.
Runco, M.A. and Chand, I. (1994) 'Problem Finding, Evaluative Thinking, and Creativity', in M.A. Runco (ed.) *Problem Finding, Problem Solving, and Creativity*. Norwood, NJ: Ablex.
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: 791–804.
Simonton, D.K. (1984) *Genius, Creativity, and Leadership*. Cambridge, MA: Harvard University Press.
Simonton, D.K. (1986) 'Aesthetic Success in Classical Music: A Computer Analysis of 1,935 Compositions', *Empirical Studies of the Arts* 4*:* 1–17.
Simonton, D.K. (1987) 'Musical Aesthetics and Creativity in Beethoven: A Computer Analysis of 105 Compositions', *Empirical Studies of the Arts* 5: 87–104.

Simonton, D.K. (1988) *Scientific Genius.* Cambridge: Cambridge University Press.
Simonton, D.K. (1991) 'Emergence and Realization of Genius: The Lives and Works of 120 Classical Composers', *Journal of Personality and Social Psychology* 61: 829–40.
Simonton, D.K. (1994) *Greatness: Who Makes History and Why*. New York: Guilford Press.
Simonton, D.K. (1995) 'Drawing Inferences from Symphonic Programs: Musical Attributes Versus Listener Attributions', *Music Perception* 12: 307–22.
Simonton, D.K. (1997) 'Creative Productivity: A Predictive and Explanatory Model of Career Landmarks and Trajectories', *Psychological Review* 104: 66–89.
Simonton, D.K. (1998) 'Fickle Fashion versus Immortal Fame: Transhistorical Assessments of Creative Products in the Opera House', *Journal of Personality and Social Psychology* 75: 198–210.
Simonton, D.K. (1999) *Origins of Genius: Darwinian Perspectives on Creativity.* New York: Oxford University Press.
Simonton, D.K. (2003a) 'Scientific Creativity as Constrained Stochastic Behavior: The Integration of Person, Product, and Process Perspectives', *Psychological Bulletin* 129: 475–94.
Simonton, D.K. (2003b) 'Qualitative and Quantitative Analyses of Historical Data', *Annual Review of Psychology* 54: 617–40.
Solomon, M. (1977) *Beethoven.* New York: Macmillan.
Sonneck, O.G. (1967[1926]) *Beethoven: Impressions by his Contemporaries*. New York: Dover.
Swafford, J. (1997) *Johannes Brahms*. New York: Random House.
Thayer, A.W. (1967) *Life of Beethoven*, rev. and ed. by E. Forbes. Princeton, NJ: Princeton University Press.
Tovey, D. (1935–9) *Essays in Musical Analysis*, 6 vols. London: Oxford University Press.
Wegeler, F. and Ries, F. (1987) *Beethoven Remembered: The Biographical Notes of Franz Wegeler and Ferdinand Ries*, trans. F. Noonan. Arlington, VA: Great Ocean Publishers.
Weisberg, R.W. (1999) 'Creativity and Knowledge: A Challenge to Theories', In R.J. Sternberg (ed.) *Handbook of Creativity*, pp. 226–50. New York: Cambridge University Press.
Weisberg, R.W. (2004) 'On Structure in the Creative Process: A Quantitative Case-Study of the Creation of Picasso's *Guernica*', *Empirical Studies of the Arts* 22: 23–54.

AARON KOZBELT is an Assistant Professor in the Psychology Department of Brooklyn College and received his PhD from the University of Chicago in 2002. His main research interests are in creativity and cognition, especially in visual art and music; his research on the psychology of music deals mainly with issues of creativity over the lifespan and the nature of creative and evaluative processes in classical composers.
Address: Department of Psychology, Brooklyn College of the City University of New York, 2900 Bedford Avenue, Brooklyn, NY 11210–2889, USA. [email: aaronk@brooklyn.cuny.edu]

Appendix: Source quotations

Work (year)	Representative quotation(s)	Source(s)	Halsey rating
Piano variations on *Venni amore*, WoO 65 (1790–1)	[Czerny:] 'Beethoven had a good opinion of them ... he ... used them to "introduce" himself.'	Thayer, 125	–
Piano Trio in E-flat, WoO 38 (1791?)	[Graeffer:] 'originally intended for the three trios, Op. 1, but omitted as too weak by Beethoven'	Thayer, 123	–
Variations on an original theme for piano trio, op. 44 (1792?) Lied with 6 variations for piano duet: *Ich denke dein*, WoO 74 (1799)	[1803] 'I ... consider [WoO 74] ... better than [Op. 44]'	Letter 82	3, –
Variations on *Se vuol ballare* for violin and piano, WoO 40 (1792–3)	[1793] 'I only wish that the work were greater and more worthy of you ... accept this trifle'	Letter 7	–
Rondo in G minor for violin and piano, WoO 41 (1793–4)	[1794?] 'trifle'	Letter 9	–
Piano Concerto No.2 in B-flat, op. 19 (c. 1794); Piano Concerto No.1 in C, op. 15 (c. 1797?)	[1800] 'not ... one of my best' (op. 19) [1801] 'neither do I reckon among the best of the kind.'	Letters 41, 48	3,2
3 Piano Trios, op. 1 (1794–5)	[1795] 'important works' [Ries:] 'Beethoven ... considered [the third] the best.'	Letter 10; Wegeler and Ries, 74	3,4,3
Lied: *Adelaide*, op. 46 (1794–5)	[1800] 'a work which welled forth so warmly from my heart'	Letter 40	3
Wind variations on *La ci darem*, WoO 28 (1795?)	[1822] 'smaller work'	Letter 1111	–
Wind Sextet in E-flat, op. 71 (1796?)	[1809] 'one of my early works ... composed in one night ... written by a composer who has produced ... better works.'	Letter 224	–

3 String Trios, op. 9 (1797–8)	[1798] 'la meilleure de ses oeuvres.'	Anderson, 1411	–,–,–
Septet, op. 20 (1799–1800)	[1802] 'The rabble is waiting for it' [1817] 'Beethoven replied ... that when he wrote the piece he did not know how to compose.'	Letter 57; Thayer, 683–4	3
Symphony No.1 in C, op. 21 (1799–1800)	[1801] 'a septet or a symphony does not sell as well as a sonata ... although a symphony should undoubtedly be worth more'	Letter 44	2
Piano Sonata No.11 in B-flat, op. 22 (1800)	[1801] 'first-rate'	Letter 44	3
Piano Concerto No.3 in C minor, op. 37 (1800)	[1800] 'I am still keeping back the better'	Letter 41	2
Piano Sonata No.14 in C-sharp minor, op. 27.2 (1801)	[to Czerny:] 'Everybody is always talking about the C-sharp minor Sonata! Surely I have written better things. There is the Sonata in F-sharp minor [sic] – that is something very different'	Thayer, 297	2,3
Piano Sonata No.24 in F-sharp, op. 78 (1809)			
Piano Sonata No.15 in D, op. 28 (1801)	[Czerny:] 'The Andante ... was long his favorite and he played it often for his own pleasure.'	Thayer, 296	3
Die Geschöpfe des Prometheus, op. 43 (1800–1)	[1801, Fuchs:] 'Haydn ... said ... "your ballet ... pleased me very much!" Beethoven replied: "O dear Papa, you are very kind; but it is far from being a *Creation*!" Haydn, surprised ... and almost offended, said ... "That is true; it is not yet a creation and I can scarcely believe that it will ever become one." Whereupon the men said their adieus, both somewhat embarrassed.'	Thayer, 273	3
Variations for Piano, opp. 34 and 35 (1802)	[both 1802] 'included ... [among] my greater musical works, the more so as the themes have been composed by me.'	Letter 67	4,3
"Kakadu" variations for piano trio, op. 121a (1803?)	[1816] 'These belong to my early works, but ... not poor stuff.'	Letter 642	3

Symphony No.3 in E-flat, op. 55 (*Eroica*) (c. 1803)	[1804] 'it will interest the musical public.' [1817, Kuffner:] 'Tell me frankly, which is your favorite among your symphonies? B. – [in great good humor] ... The "Eroica."'	Letter 96; Thayer, 674	1
3 Marches for Piano (4 hands), op. 45 (1803)	[1803] 'easy and yet not too trivial.' [Ries:] 'Beethoven composed part of the second march while giving me a lesson on a sonata ... which I had to play that evening ... I was also to play the marches with him on the same occasion.'	Letter 81; Wegeler and Ries, 80	–
Christus am Ölberge, op. 85 (1803–4)	[1804] 'I wrote the whole work in a few weeks and ... naturally some passages did not satisfy me later on.' [1811] 'my first work in that style ... written in a fortnight and during all kinds of disturbances'	Letters 96, 325	4
Piano Sonata No.23 in F minor, op. 57 (1804–5)	[Czerny:] 'Beethoven himself considered [it] his greatest [sonata].'	Thayer, 407	3
Fidelio, op. 72 (1804–5, revised 1814)	[1814] 'The whole has been thoroughly revised and its changed form is much more effective on the stage, and more than half written afresh.'	Kalischer, vol. I, 331; Wegeler and Ries, 58	4
32 Variations on an Original Theme in C minor, WoO 80 (1806)	[Jahn:] 'Such nonsense by me? Oh, Beethoven, what an ass you were!'	Thayer, 410	2
Piano Concerto No.4 in G, op. 58; 3 String Quartets, op. 59; Symphony No.4 in B-flat, op. 60; Violin Concerto in D, op. 61; *Coriolan* Overture, op. 62 (all 1805–6)	[1808] 'seven major works' [1806?] 'not for you, but for a later age!' (op. 59)	Letter 167; Thayer, 451–2, 409	1 2,2,2 2 1 2
Leonore Overture No.1, op. 138 (1807)	[Schindler:] 'the composer himself was not thoroughly satisfied ... it was laid aside'	Thayer, 385	3

Mass in C, op. 86 (1807)	[1808] 'I have treated the text in a manner in which it has rarely been treated.' [1808] 'especially close to my heart'	Letters 167, 169	3
Lied: *Andenken*, WoO 136 (c. 1809)	[1809] 'if you had not reminded me, I should have entirely forgotten [it]'	Letter 220	–
Lied aus der Ferne, WoO 137 (1809)	[1810] 'I composed the aria purely out of kindness'	Letter 245	–
4 marches for wind band, WoO 18, 19, 20, 24 (1809–10, 1810, 1810?, 1816)	[1825] 'minor works'	Letter 1368	–,–,–,–
Overtures: *Die Ruinen von Athen, Zur Namensfeier, König Stephan*, opp. 113, 115, 117 (1811, 1814–15, 1811)	[1816] 'the three overtures do not belong to my best and greatest works'	Letter 664	–,–,4
Symphony No.7 in A, op. 92 (1811–2)	[1815] 'one of my most excellent works'	Letter 544	2
Symphony No.8 in F, op. 93 (1811–12)	[Czerny:] 'The Eighth Symphony "by no means pleased" and Beethoven was angry thereat, "because it is much better" [than No.7]'	Thayer, 576	2
Entr'acte from *Tarpeja*, WoO 2 (1813)	[1826] 'trifle'	Letter 1487	–
Wellingtons Sieg, op. 91 (1813)	[Tomaschek:] '[Beethoven] declared the work to be a folly.' [1814] 'I was plunged into the most dreadful financial embarrassment'	Thayer, 565; Letter 485	5
Piano Sonata No.28 in A, op. 101 (1816)	[1817] 'difficult to perform … what is difficult, is also beautiful, good, great, and so forth … this is the most lavish praise'	Letter 749	2
Piano Sonata No.29 in B-flat, op. 106 (1817–18)	[1818?] 'Beethoven told Czerny it was to be his greatest.' [Potter:] 'he knew now, he thought [how to compose], and … said, "I am writing something better now."'	Thayer, 714, 684	2
Piano Sonatas Nos 30–2, opp. 109–11 (1820, 1821–2, 1821–2)	[Holz:] 'Beethoven … called [them] his last but also the best that he had written'	Thayer, 984	2,1,1

Missa Solemnis in D, op. 123 (1819–23)	[1824] 'my greatest work'	Letter 1270	1
Gratulations-Minuet, WoO 3 (1822)	[1825] 'minor work'	Letter 1368	–
Symphony No.9 in D minor, op. 125 (1818–24)	[1824] 'the grandest [symphony] I have written'	Letter 1305	1
6 Bagatelles, op. 126 (1823–4)	[1824] 'probably the best of this kind which I have composed'	Letter 1321	1
String Quartet No.12 in E-flat, op. 127 (1823–4)	[1825] 'the best and finest I have ever written.' [1825] 'as good a work as I could give to my best friend.'	Letters 1355, 1420	1
String Quartet No.15 in A minor, op. 132 (1825)	[1826?] 'Holz said that when he once remarked to Beethoven that the one in B-flat was the greatest of the three quartets (opp. 127, 130, 132) the composer replied: "Each in its own way."'	Thayer, 982	1
String Quartet No.13 in B-flat, op. 130 (1825–6)	[1826?] '[the] high opinion of the Cavatina was shared by Beethoven himself. Holz said that it cost the composer tears in the writing and brought out the confession that nothing that he had written had so moved him.'	Thayer, 975	1
Grosse Fuge in B-flat, op. 133 (1826)	[1826, Holz:] 'On hearing that the Alla danza tedesca and the Cavatina had received such thunderous applause that they had to be encored, he snapped in exasperation: "Yes, these delicacies! Why not the Fugue?" and then gave his opinion of the audience: "Cattle! Asses!"'	Solomon, 323	3
String Quartet No.14 in C-sharp minor, op. 131 (1826)	[1826?] 'He declared the C-sharp minor Quartet to be his greatest.'	Thayer, 982	1

Note: Dates of composition follow titles of works. Date of remarks (if known) or sources (if not direct Beethoven quotations) are bracketed before each quotation. Letter numbers follow Anderson (1961).