Psychology of Music,
2001, **29,** 108–127

The Relationship Between Musical Complexity and Liking in Jazz and Bluegrass

MARK G. ORR and STELLAN OHLSSON

Department of Psychology, University of Illinois at Chicago, 1009 Behavioural Sciences Building, 1007 West Harrison Street, Chicago, Illinois 60607-7137, U.S.A.

e-mail: morr@uic.edu

Abstract

Berlyne (1971) and others have proposed that liking of music is an inverted-U shaped function of its complexity. Past research has been inconclusive concerning the generality of the inverted-U relationship. The present studies tested the generality of this relationship across related but distinct musical styles. Our experimental participants judged the complexity and liking of short pieces of music created by professional musicians who where instructed to improvise at different levels of complexity. Experiment 1 tested whether the inverted-U relationship was present between perceived complexity and liking for jazz and bluegrass when each style was presented separately. Experiment 2 replicated the findings from Experiment 1 under interleaved presentation conditions. The results support the inverted-U shaped relationship between perceived complexity and liking for bluegrass but not for jazz. It is possible that the relationship between liking and perceived musical complexity varies across musical styles.

Why do people listen to music? One approach to this question is to investigate what people like about music and which factors determine how much they like any particular piece. As Berlyne (1971) and others (*e.g.* Davies, 1978; Heyduk, 1975; Walker, 1980) have pointed out, it is plausible that people find music more or less appealing as a function of its complexity. Overly simple compositions are likely to be boring, while overly complex ones might strain the listener's aesthetic sensibilities. These observations suggest that listeners prefer music of intermediate levels of complexity.

Empirical research has provided some support for this *inverted-U hypothesis* in music (Crozier, 1974; Heyduk, 1975; North and Hargreaves, 1995; North and Hargreaves, 1996a; North and Hargreaves, 1996b; Vitz, 1966). However, the support is qualified by some contradictions. Russell (1982) and Smith and Melara (1990) found a negative relationship between liking and complexity; others have found that the nature of the relationship differs across musical pieces (Simon and Wohlwill, 1968).

One issue left undecided by prior research is whether the inverted-U relationship generalises across musical styles. North and Hargreaves (1996a) found an inverted-U-shaped relationship for new-age music. In contrast, mechanical organ music of moderate complexity did not fit the inverted-U prediction. However, the complexity of the organ music was not varied. Another study by North and Hargreaves (1997) found that liking for a piece of music was more associated with liking for the style than for the piece itself. Finally, Hargreaves (1984) tested the prediction that if the complexity of a piece is too high for it to be liked, repeated exposure decreases subjective complexity to a level that is likeable. Although some musical styles, *e.g.* popular music, conformed to this prediction, jazz did not. Liking of jazz pieces did not increase after repeated exposure. These three studies suggest that the relationship between perceived complexity and liking might interact with musical style. To our knowledge there are no studies of liking that systematically vary complexity within several musical styles.

The hypothesis that the relationship between complexity and liking varies across musical styles is contrary to both the arousal explanation (Berlyne, 1971) and alternative explanations (Martindale, 1984; 1988) for the inverted-U relationship, because these explanations implicitly assume that the relationship between complexity and liking is the same in different musical styles. The purpose of the research reported in this article is to investigate the generality of the inverted-U relationship across musical styles.

Studies of the relationship between liking and complexity must address the following three methodological issues. First, there is the question of ecological validity. Many studies have used either artificially generated auditory stimuli that do not sound like music or stimuli that only approximate music. For example, Vitz (1966) used computer-generated tone sequences and Smith and Melara (1990) used variants of a simple nine-chord progression. Stimuli of this sort might lack expressive timing or other characteristics that are important in music, so results obtained with such stimuli might not generalise to music. The problem of ecological validity was partly overcome by North and Hargreaves (1995), who used 30-second excerpts from popular music. The fact that such excerpts do not have natural-sounding beginnings or endings lowers their ecological validity. To increase ecological validity, we used short (30-second) improvisations that were produced by professional musicians who were skilled improvisers. This resulted in improvisations with natural-sounding beginnings and endings.

Second, to investigate the relationship between complexity and liking, there must be a way to quantify complexity. A common approach is to postulate a theoretical definition of complexity and then to generate auditory stimuli that vary in complexity, so defined. For example, Crozier (1974) and Vitz (1966) defined complexity in terms of information theory, while Heyduk (1975), Simon and Wohlwill (1968) and Smith and Melara (1990) defined it by varying specific musical dimensions (*e.g.* number of different chord types). This approach requires independent verification that the theoretically defined dimensions are the dimensions that determine the level of complexity a listener perceives in a piece of music. We prefer the alternative approach of measuring the perceived complexity *via* subjective ratings (North and Hargreaves, 1995). However, this method provides no experimental control over the complexity levels. To overcome this problem, we instructed the professional musicians who produced our stimuli to vary the complexity across five conceptually defined complexity levels.

Third, the relationship between complexity and liking might be difficult to assess due to other determinants of liking. For example, Martindale and Moore (1989) have argued that prototypicality is a stronger determinant of liking than complexity both in music and in other domains (Martindale, 1984). Common sense and research suggest yet other determinants, *e.g.*, association with a particular style or artist (Zillmann and Gan, 1997), effects of training (Orr and Ohlsson, 1999) and situational influences (Konecni, 1982). If such alternative determinants of liking are orthogonal to complexity, the relationship between complexity and liking might be attenuated to the point where it is impossible to determine its mathematical shape. It turns out that this is not the case for the data presented in this article. Another possibility is that one or more determinants of liking are correlated and hence confounded with complexity. This latter possibility is discussed in the General Discussion.

The present series of experiments tested the generality of the inverted-U hypothesis with respect to improvisations in jazz and bluegrass, two musical styles which are underrepresented in music psychology. We chose jazz and bluegrass because improvisation is common in both styles. The use of improvisations allowed us to create short, naturalistic stimuli and to easily control each improvisation's complexity level. The jazz improvisations were created by professional jazz musicians and the bluegrass improvisations were created by professional bluegrass musicians. To separate generality across styles from generality across samples, we presented both musical styles to each experimental participant.

If the inverted U- hypothesis does generalise across musical styles, the following two characteristics should be present for both jazz and bluegrass. First, the relationship between complexity and liking should have a reliable quadratic component that is reliably larger in magnitude than the linear component. Second, there should be support for both the ascending and the descending legs of the curve.

Experiment 1 tested whether the inverted-U relationship is present between perceived complexity and liking for jazz and bluegrass when the improvisations are blocked by musical style. Experiment 2 replicated the findings from Experiment 1 with an interleaved (as opposed to blocked) presentation format.

Experiment 1

The purpose of Experiment 1 was to investigate the relationship between complexity and liking in jazz and bluegrass. To this end, we needed a method for generating stimuli that varied in complexity. One method is to select various excerpts that vary in complexity from pre-existing musical recordings. For example, Russell (1982) varied complexity by selecting 1-minute excerpts from jazz records. However, these excerpts featured an ensemble, soloist or both. The fact that ensemble sections were included in these excerpts obscures what each subject was attending to within the piece (*e.g.*, the piano, drums or both). In an attempt to control for this, our procedure used solo improvisations, *i.e.* improvisations without other instruments in the background, created especially for the purpose of this study.

The procedure was simple. Non-musicians judged the stimuli with respect to both complexity and liking in two sessions. Our first hypothesis was that liking is an inverted-U shaped function of perceived complexity. A second hypothesis was that this relationship holds for both musical styles.

Method

Participants

Sixty-four (36 females, 28 males) undergraduate psychology students from the University of Illinois at Chicago participated as listeners in return for course credit. Their mean years of musical training was 3·83 (*SD* = 4·75).

Materials

Two professional jazz musicians each created 20 short (30 to 40 s) jazz improvisations and two professional bluegrass musicians each created 20 short bluegrass improvisations, for a total of 80 improvisations. One of the jazz musicians performed on the Hammond-B3 organ; the other on the soprano saxophone and flute. One of the bluegrass musicians performed on the guitar; the other on the mandolin. The musicians improvised while listening to chord progressions over headphones. The use of headphones allowed us to record the improvisations without also recording the chord progressions.

Each musician was instructed to make each improvisation match one of five complexity levels; henceforth this dimension is called *performer complexity*. Level 1 – low complexity – was defined as "predictable, simple and uniform", while Level 5 – high complexity – was defined as "unpredictable, surprising, and erratic" (these terms were borrowed from North and Hargreaves 1995).

Each musician created 20 improvisations in two separate sessions. The first session served as practice. Only the improvisations recorded during the second session were used as stimuli. The 40 jazz improvisations were placed in two random sequences, referred to as Jazz A and Jazz B, with a 10-second pause after each improvisation. Similarly, the 40 bluegrass improvisations were placed in two random sequences, Bluegrass A and Bluegrass B.

Procedure

As practice, participants rated liking or complexity for three improvisations at complexity levels 1, 5, and 3 (in that order) without feedback concerning the complexity levels. All participants heard two sequences of improvisations, one from each style (either Jazz A or Jazz B and either Bluegrass A or Bluegrass B), in each of two sessions one week apart. The sequences were counter-balanced across participants, so that approximately half the participants listened to either sequence within each musical style. Specifically, 39 subjects listened to sequence A and 25 listened to sequence B of the bluegrass improvisations, while 28 subjects listened to sequence A and 36 listened to sequence B of the jazz improvisations. Each participant encountered the same two sequences, in the same order, in both sessions.

Approximately one-half (31) of the participants rated jazz improvisations in the first half of each session and bluegrass improvisations in the second half; the other half (33) rated these styles in the opposite order. The participants were allowed a three- to five-minute break between musical styles (*e.g.*, between Jazz A and Bluegrass B).

In each session, the listeners rated each improvisation for either complexity or liking during the 10-second pause between successive improvisations. The listeners rated complexity on a 7-point scale where the end-points were identical to those used by the performers (Level 1 = low complexity; Level 7 = high complexity). Liking was also rated on a 7-point scale. Level 1 – low liking – was defined as "do not like at all", and Level 7 – high liking – was defined as "like very much". The effects of the participants' assumptions about the relationship between complexity and liking (Sluckin, Hargreaves and Colman, 1983) were controlled for by counter-balancing the order of the rating tasks. Approximately one-half (31) of the participants rated complexity in their first session and liking in their second session; the other half (33) performed these rating tasks in the opposite order.

At the end of the second session the participants completed a musical background questionnaire. Finally, they wrote their answers to the following two questions: *What makes one piece of music more or less likeable than another?*, and *What makes one piece of music more or less complex than another?* These data will be reported elsewhere.

Results

Complexity Manipulation Verification

Before examining the main results, we needed to verify that our complexity manipulation worked. In other words, did our instructions to the musicians result in musical pieces that were, in fact, experienced as varying in complexity by non-musician listeners.

The main question was whether listeners perceived differences among the five complexity levels. For the jazz improvisations, an ANOVA indicated a main effect for performer complexity, F (4, 252) = 148·37, $p < {\cdot}001$, MSE = 0·26 (see Figure 1). In addition, pair-wise comparisons indicated that each complexity level was perceived as distinct from the complexity levels adjacent to it (Performer Complexity Level 1 *v.* Level 2, F (1, 63) = 8·32, $p < {\cdot}01$, MSE = 0·32; Level 2 *v.* Level 3, F (1, 63) = 110·64, $p < {\cdot}001$, MSE = 0·36; Level 4 *v.* Level 5, F (1, 63) = 102·49, $p < {\cdot}001$, MSE = 0·43); except for Performer Complexity Level 3 *v.* Level 4, F (1, 63) = 1·45, *ns*, MSE = 0·42.

For the bluegrass improvisations, an ANOVA indicated a main effect for performer complexity, F (4, 252) = 130·19, $p < {\cdot}001$, MSE = 0·30 (see Figure 1). In addition, pair-wise comparisons indicated that each complexity level was perceived as distinct from the complexity levels adjacent to it (Performer Complexity Level 1 *v.* Level 2, F (1, 63) = 59·29, $p < {\cdot}001$, MSE = 0·47; Level 2 *v.* Level 3, F (1, 63) = 44·79, $p < {\cdot}001$, MSE = 0·43; Level 3 *v.* Level 4, F (1, 63) = 24·76, $p < {\cdot}001$, MSE = 0·43; Level 4 *v.* Level 5, F (1, 63) = 20·07, $p < {\cdot}001$, MSE = 0·40).

Regression Analyses

The main purpose of Experiment 1 was to determine the shape of the relationship between complexity and liking. For each regression analysis, the linear and quadratic equations were tested. The linear equation was interpreted as the correct interpretation when (*a*) it was statistically reliable, and (*b*) the quadratic equation did not explain reliably more variance compared to the linear equation, as tested by an increment to R^2 test. The quadratic equation was interpreted as

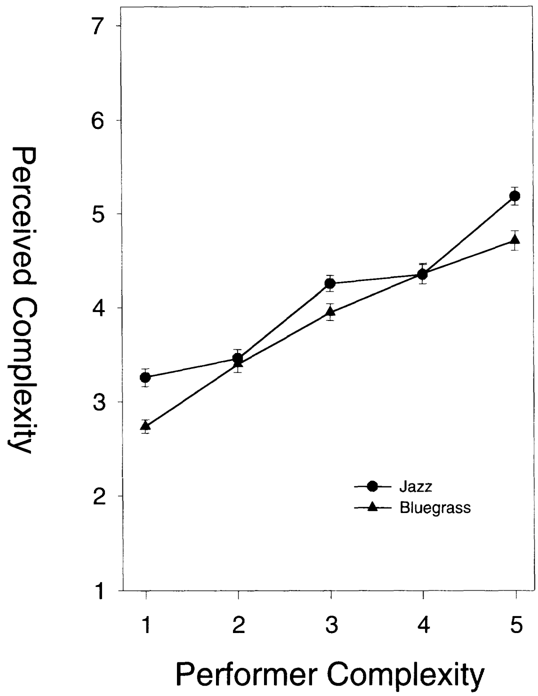

FIG. 1

Mean perceived complexity as a function of performer complexity for the jazz and bluegrass improvisations in Experiment 1.

the correct interpretation when (*a*) it was statistically reliable, and (*b*) the quadratic equation explained reliably more variance compared to the linear equation. When the quadratic equation is reliable, the linear equation might also be reliable. This does not imply that both interpretations are supported. The increment to R^2 test determines which one should be interpreted as correct (Stimson, Carmines and Zeller, 1981).

A figure accompanies each regression analyses. In each figure, the regression line represents the best fitting line according to the above criteria.

All improvisations. The mean liking rating for each improvisation was regressed onto the mean complexity rating for that improvisation. The results are shown in Figure 2. The linear component was statistically reliable, $F(1, 78) = 66{\cdot}79$, $p < {\cdot}001$, $MSE = 0{\cdot}19$, as was the quadratic component, $F(2, 77) = 48{\cdot}98$, $p < {\cdot}001$, $MSE = 0{\cdot}15$. Furthermore, the quadratic component accounted for reliably more variance in liking than the linear component, $R_q^2 = {\cdot}56$ as compared to $R_l^2 = {\cdot}46$, $F(1, 77) = 17{\cdot}25$, $p < {\cdot}001$.

Jazz improvisations. The same analyses were computed for the jazz improvisations as for all improvisations. The results are shown in Figure 3. The

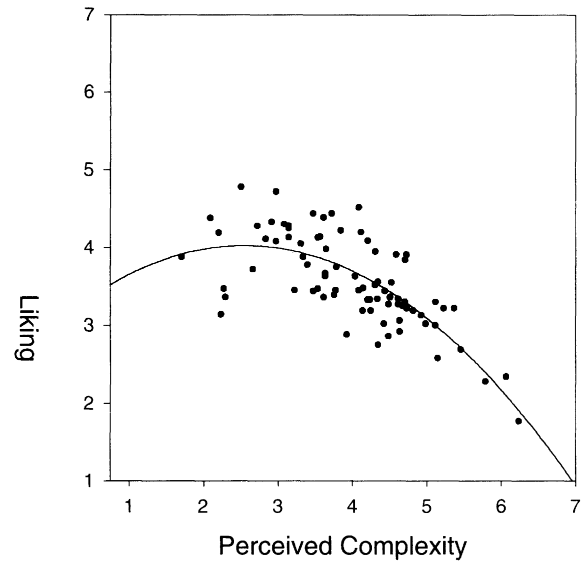

FIG. 2
Mean liking plotted over mean perceived complexity for all 80 improvisations in Experiment 1.

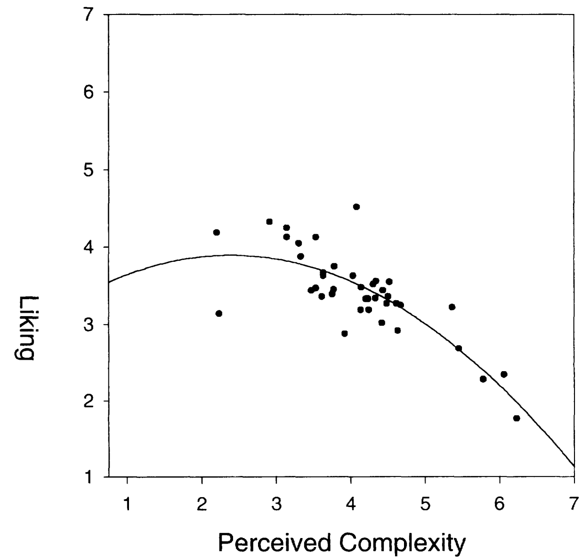

FIG. 3
Mean liking plotted over mean perceived complexity for the 40 jazz improvisations in Experiment 1.

linear component was statistically reliable, F (1, 38) = 52·04, $p < \cdot 001$, MSE = 0·13, as was the quadratic component, F (2, 37) = 35·02, $p < \cdot 001$, MSE = 0·11. Furthermore, the quadratic component accounted for reliably more variance in liking than the linear component, $R_q^2 = \cdot 65$ as compared to $R_l^2 = \cdot 58$, F (1, 37) = 8·17, $p < \cdot 01$.

Bluegrass improvisations. The same analyses were computed for the bluegrass improvisations as for the jazz improvisations. These results are shown in Figure 4. The linear and quadratic components were statistically reliable, F (1, 38) = 20·99, $p < \cdot 01$, MSE = 0·22 and F (2, 37) = 19·93, $p < \cdot 001$, MSE = 0·17, respectively. Furthermore, the quadratic component explained reliably more variance in liking than the linear component, $R_q^2 = \cdot 52$ as compared to $R_l^2 = \cdot 36$, F (1, 37) = 12·52, $p < \cdot 01$.

Order Effects

For each musical style, we used two random sequences of the improvisations. For the bluegrass improvisations, 39 subjects listened to sequence A and 25 listened to sequence B. For the jazz improvisations, 28 subjects listened to sequence A and 36 listened to sequence B. Hence, the improvisations preceding any one improvisation in a sequence differed between sequences. This might have affected the complexity or liking ratings.

In order to rule out such ordering effects, it is necessary to find a reliable positive product–moment correlation between each sequence (for liking or complexity ratings) and a non-reliable mean difference, as evidenced by the t-test. Either of

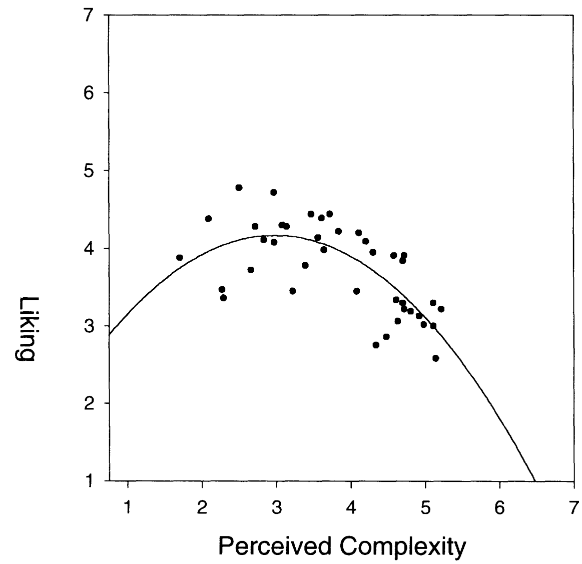

FIG. 4

Mean liking plotted over mean perceived complexity for the 40 bluegrass improvisations in Experiment 1.

these statistics alone will not reveal the ordering effects because both the mean difference and the direction of the relationship are important.

The correlations used the mean ratings (complexity or liking) for each improvisation (40 for each style) as data points in order to test for a positive linear relationship between the two sequences. The *t*-tests used the mean rating (complexity or liking) between the two sequences to test whether the difference is reliable. These effects were tested separately for each rating type (complexity or liking) and for each musical style (jazz or bluegrass).

Complexity rating order effects. For both jazz and bluegrass, the correlations for the complexity ratings between Sequence A and Sequence B were reliable, $r = \cdot 92$ ($p < \cdot 05$, $n = 40$) and $r = \cdot 88$ ($p < \cdot 05$, $n = 39$), respectively. In addition, no reliable mean differences existed between Sequence A and Sequence B for the jazz improvisations ($M_a = 4\cdot 13$, $SD = 0\cdot 63$; $M_b = 4\cdot 08$, $SD = 0\cdot 57$) $t\ (62) = -0\cdot 38$, *ns*, or for the bluegrass improvisations, ($M_a = 3\cdot 80$, $SD = 0\cdot 57$; $M_b = 3\cdot 88$, $SD = 0\cdot 54$), $t\ (62) = 0\cdot 60$, *ns*.

Liking rating order effects. For both jazz and bluegrass, the correlations for the liking ratings between Sequence A and Sequence B were reliable, $r = \cdot 76$ ($p < \cdot 05$, $n = 40$) and $r = \cdot 85$ ($p < \cdot 05$, $n = 39$), respectively. In addition, no reliable mean difference in liking existed between Sequence A and Sequence B for the jazz improvisations ($M_a = 3\cdot 54$, $SD = 0\cdot 74$; $M_b = 3\cdot 31$, $SD = 0\cdot 79$) $t\ (62) = -1\cdot 21$, *ns*. However, there was a reliable mean difference in liking between Sequence A and Sequence B for the bluegrass improvisations ($M_a = 4\cdot 03$, $SD = 0\cdot 88$; $M_b = 3\cdot 27$, $SD = 0\cdot 93$), $t\ (62) = -3\cdot 27$, $p < \cdot 01$.

In summary, there was only one ordering effect. The mean liking rating between subjects that rated bluegrass improvisations in one random sequence was higher than in the other random sequence. This will be discussed further in the Discussion section.

Discussion

The results of the regression analyses are consistent with the inverted-U hypothesis. When the average liking ratings are plotted as a function of the average complexity ratings for each improvisation, the curvilinear fit explains more of the variance than the linear fit. This is true for the data as a whole and for each musical style.

However, the statistical characteristics revealed by the regression analyses are necessary but not sufficient to identify the relationship as an inverted-U function. In addition, there should be support for both legs of the curve.

The application of this latter criterion reveals a difference between the two musical styles. Visual inspection of Figure 4 suggests strong support for the inverted-U relation for the bluegrass improvisations. This impression can be documented by (*a*) locating the apex on the abscissa, (*b*) counting the number of data points to the left and to the right of the apex of the inverted-U curve, and (*c*) calculating the linear correlation coefficient for the data on either side of the apex. The apex is at 2·99 on the complexity scale. Ten improvisations lie to the left and 30 to the right of this point. The correlation for the ten improvisations to the left of the apex is ·36 (*ns*, $n = 10$); for the 30 to the right it is $-\cdot 72$ ($p < \cdot 05$, $n = 30$).

With the exception of the non-reliability of the correlation for the data points to the left of the data points, these results support an inverted-U relationship between liking and complexity for the bluegrass improvisations.

However, visual inspection of Figure 3 suggests that the left, ascending leg of the inverted-U has less support than the right, descending leg for the jazz improvisations. The apex of the curve in Figure 3 is found at a perceived complexity value of 2·38. Only two data points are to the left of this value; the other 38 are to the right. A correlation with only two data points necessarily equals 1, so the data to the left of the apex do not reveal the strength or direction of the relationship. The correlation for the 38 data points to the right was $-{\cdot}84$ ($p < {\cdot}05$, $n = 38$). Hence, the superiority of the inverted-U interpretation rests on a slender basis for this style. The hypothesis that the relationship is negative linear, *i.e.*, that non-musicians prefer simple jazz, is as good an explanation for the data in Figure 3 as the inverted-U hypothesis.

There was one ordering effect. The average liking of the bluegrass pieces when presented in the one random sequence (Bluegrass A) was reliably higher than the liking of those same stimuli when presented in the other random sequence (Bluegrass B; see the Results section). As Figure 5 shows, when the relationship between liking and complexity is plotted separately for these two presentation sequences, the inverted-U relationship appears for Bluegrass B (see panel b). The linear component was statistically reliable, F (1, 37) = 13·72, $p < {\cdot}01$, MSE = 0·29, as was the quadratic component, F (2, 36) = 17·93, $p < {\cdot}001$, MSE = 0·20. The quadratic component accounted for reliably more variance in liking than the linear component, $R_q^2 = {\cdot}50$ as compared to $R_l^2 = {\cdot}27$, F (1, 36) = 16·42, $p < {\cdot}001$.

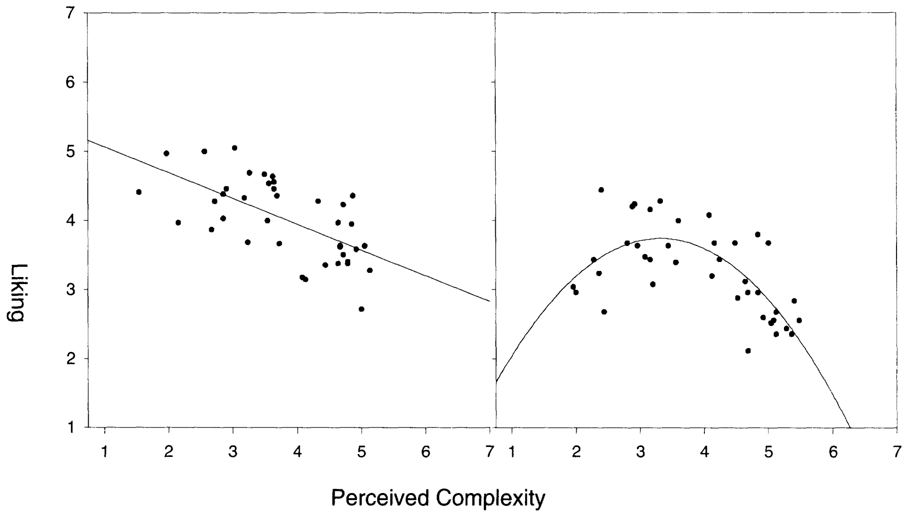

FIG. 5

Mean liking plotted over mean perceived complexity for both random sequences of the 40 bluegrass improvisations in Experiment 1.

In contrast, the inverted-U relationship did not appear for Bluegrass A; see panel a in Figure 5. The linear component was statistically reliable, $F\ (1, 38) = 24{\cdot}59$, $p < {\cdot}001$, $MSE = 0{\cdot}20$, as was the quadratic component, $F\ (2, 37) = 14{\cdot}02$, $p < {\cdot}001$, $MSE = 0{\cdot}19$, but the quadratic component did not account for reliably more variance in liking than the linear component, $R_q^2 = {\cdot}43$ as compared to $R_l^2 = {\cdot}39$, $F\ (1, 37) = 2{\cdot}49$, *ns*. We have no explanation for this effect. A similar breakdown of the data for the jazz pieces exhibits the inverted-U relationship in both plots.

In short, the results of the regression analyses are consistent with the inverted-U hypothesis for both jazz and bluegrass. The amount of support for the two legs of the curve is also consistent with the inverted-U relationship, but only for the bluegrass improvisations. Due to the lack of support for the left, ascending leg of the curve, a negative linear relationship cannot be ruled out for the jazz improvisations. The possibility that the relationship between complexity and liking varies across musical styles invites replication of these findings.

Experiment 2

The purpose of Experiment 2 was to replicate and generalise the results from Experiment 1. If the inverted-U relationship is robust and generalises across musical styles, it should not matter whether pieces of music are presented blocked by style or interleaved. Experiment 1 used the blocked presentation format; Experiment 2 used an interleaved format. To eliminate interactions between the two types of ratings, we followed Sluckin, Hargreaves and Colman (1983) by using a between-subjects design in which each participant rated either complexity or liking, instead of the within-subjects design used in Experiment 1.

Method

Participants

One hundred and fifty-one undergraduate psychology students (83 females, 68 males) from the University of Illinois at Chicago participated as listeners in return for course credit. Their mean years of musical training was 3·99 (SD = 6·58).

Materials

We used the same improvisations created for Experiment 1 with the following changes. First, 20 of the 80 improvisations were deleted from the set to reduce the listening fatigue of the participants. We decided to remove two improvisations from each of the five performer complexity levels (see Experiment 1) for each musical style. The two improvisations that were removed had the highest standard deviation in rated complexity in Experiment 1. We adopted this criterion in order to reduce the error variance associated with the complexity judgements. One of the removed improvisations was inadvertently duplicated in the original stimulus set. It was thus removed twice, leaving 59 improvisations in this set instead of 60.

Second, the jazz and bluegrass improvisations were combined into one set. We constructed two random sequences (A and B) of the 59 combined jazz and bluegrass improvisations. The musical experience questionnaire used in Experiment 1 was also given to the participants in Experiment 2.

Procedure

Eighty-one participants listened to Sequence A and 71 to Sequence B. Furthermore, 71 participants rated the complexity of the improvisations and 81 rated their likeability. (It is accidental that the number of participants for each random sequence coincides with the number of participants for each rating type; these two variables were independent.) For subjects who made complexity ratings, 32 rated Sequence A and 39 rated Sequence B. For subjects who made liking ratings, 39 rated Sequence A and 42 rated Sequence B. The instructions were identical to those used in Experiment 1.

Results

Complexity Manipulation Verification

Because Experiment 2 used an interleaved presentation format as opposed to the blocked format used in Experiment 1, we needed to verify that our complexity manipulation still worked in this presentation format. As in Experiment 1, the

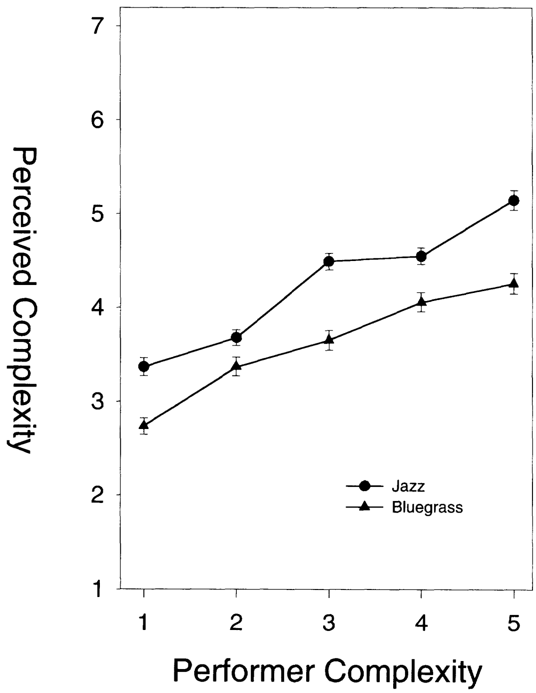

FIG. 6
Mean perceived complexity as a function of performer complexity for the jazz and bluegrass improvisations in Experiment 2.

question was whether listeners perceived differences among the five complexity levels. For the jazz improvisations, an ANOVA indicated a main effect for performer complexity, F (4, 280) = 110·21, p < ·001, MSE = 0·33 (see Figure 6). In addition, pair-wise comparisons indicated that each complexity level was perceived as distinct from the complexity levels adjacent to it (Performer Complexity Level 1 *v.* Level 2, F (1, 70) = 14·83, p < ·001, MSE = 0·46; Level 2 *v.* Level 3, F (1, 70) = 84·13, p < ·001, MSE = 0·56; Level 4 *v.* Level 5, F (1, 70) = 58·72, p < ·001, MSE = 0·43); except for Performer Complexity Level 3 *v.* Level 4, F (1, 70) = 0·51, *ns*, MSE = 0·44.

For the bluegrass improvisations, an ANOVA indicated a main effect for performer complexity, F (4, 280) = 73·25, p < ·001, MSE = 0·35 (see Figure 6). In addition, pair-wise comparisons indicated that each complexity level was perceived as distinct from the complexity levels adjacent to it (Performer Complexity Level 1 *v.* Level 2, F (1, 70) = 51·93, p < ·001, MSE = 0·55; Level 2 *v.* Level 3, F (1, 70) = 9·09, p < ·01, MSE = ·62; Level 3 *v.* Level 4, F (1, 70) = 23·93, p < ·001, MSE = 0·50; Level 4 *v.* Level 5, F (1, 70) = 4·89, p < ·05, MSE = 0·56).

Regression Analyses

The main purpose of Experiment 2 was to determine whether the support for the inverted-U relationship obtained in Experiment 1 could be replicated in a second sample of participants and with a different presentation format. The statistical criteria for determining the presence of an inverted-U relationship were described in the Results section of Experiment 1.

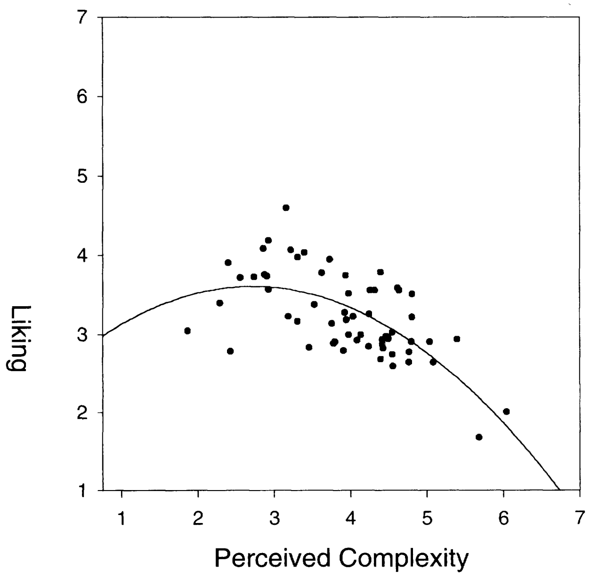

FIG. 7
Mean liking plotted over mean perceived complexity for all 59 improvisations in Experiment 2.

All Improvisations. The mean liking rating for each improvisation was regressed onto the mean complexity rating for each improvisation. The results are shown in Figure 7. The linear component was statistically reliable, $F(1, 57) = 31{\cdot}56$, $p < 0{\cdot}001$, $MSE = 0{\cdot}19$, as was the quadratic component, $F(2, 56) = 22{\cdot}37$, $p < {\cdot}001$, $MSE = 0{\cdot}17$. Furthermore, the quadratic component accounted for reliably more variance in liking than the linear component, $R_q^2 = {\cdot}44$ as compared to $R_l^2 = {\cdot}36$, $F(1, 56) = 8{\cdot}83$, $p < {\cdot}01$.

Jazz Improvisations. The same analyses were computed for the jazz improvisations as for all improvisations. The results are shown in Figure 8. The linear component was statistically reliable, $F(1, 28) = 16{\cdot}38$, $p < {\cdot}001$, $MSE = 0{\cdot}11$, as was the quadratic component, $F(2, 27) = 12{\cdot}72$, $p < {\cdot}001$, $MSE = 0{\cdot}09$. Moreover, the quadratic component accounted for reliably more variance in liking than the linear component, $R_q^2 = {\cdot}49$ as compared to $R_l^2 = {\cdot}37$, $F(1, 27) = 6{\cdot}08$, $p < {\cdot}03$.

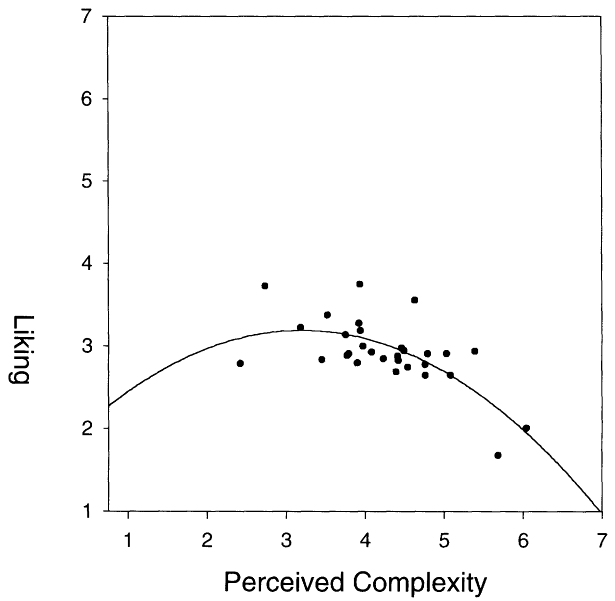

FIG. 8

Mean liking plotted over mean perceived complexity for the 30 jazz improvisations in Experiment 2.

Bluegrass Improvisations. The same analyses were computed for the bluegrass improvisations as for the jazz improvisations. The results are shown in Figure 9. The linear component was reliable, $F(1, 27) = 5{\cdot}10$, $p < {\cdot}05$, $MSE = {\cdot}17$, as was the quadratic component, $F(2, 26) = 7{\cdot}94$, $p < {\cdot}01$, $MSE = {\cdot}13$. Furthermore, the quadratic component accounted for reliably more variance in liking than the linear component, $R_q^2 = {\cdot}38$ as compared to $R_l^2 = {\cdot}16$, $F(1, 26) = 9{\cdot}22$, $p < {\cdot}01$.

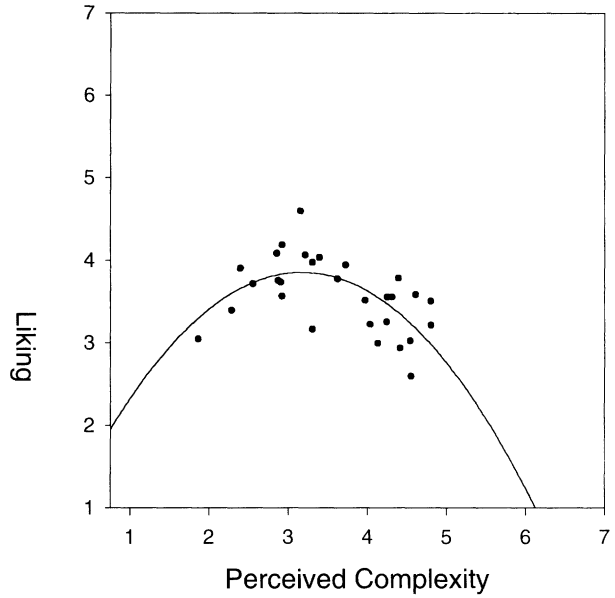

FIG. 9
Mean liking plotted over mean perceived complexity for the 29 bluegrass improvisations in Experiment 2.

Order Effects

As explained in the Results sections of Experiment 1, because the improvisations were presented in two different random sequences we used product–moment correlations and *t*-tests to test for ordering effects. The tests were applied separately for each type of rating (complexity or liking).

The correlations between the two random sequences for the complexity and liking ratings were reliable, $r = \cdot 91$ ($p < \cdot 05$, $n = 59$) and $r = \cdot 79$ ($p < \cdot 05$, $n = 59$), respectively. In addition, the mean rating was not reliably different between the two random sequences for the complexity ratings ($M_a = 3{\cdot}90$, $SD = 0{\cdot}56$; $M_b = 3{\cdot}96$, $SD = 0{\cdot}43$), $t\ (69) = -0{\cdot}58$, *ns* or for the liking ratings ($M_a = 3{\cdot}23$, $SD = 0{\cdot}91$; $M_b = 3{\cdot}27$, $SD = 0{\cdot}97$), $t\ (79) = -0{\cdot}16$, *ns*. In summary, no order effects were found in this experiment.

Discussion

The results from Experiment 2 replicate the main findings from Experiment 1. When the mean liking is plotted as a function of the mean complexity for each piece, the curvilinear fit explains more of the variance than the linear fit. This is true of the data set as a whole as well as for each music style.

Once again, visual inspection of the scatter plots suggests a difference between the two musical styles. The inverted-U relationship is strongly supported for the bluegrass improvisations; see Figure 9. The apex is at 3·16 on the perceived complexity scale. Ten improvisations lie to the left and 19 to the right of this point. The correlation for the ten improvisations to the left of the apex was ·79 ($p <$ ·05, $n = 10$); for the 19 to the right it was –·56 ($p <$ ·05, $n = 19$). Hence, all of our criteria for an inverted-U relationship are satisfied for bluegrass.

The graph for the jazz improvisations (Figure 8) suggests that the left, ascending leg of the inverted-U has less support than the right, descending leg. The apex of the curve is at 3·19 on the perceived complexity scale. Only three data points are to the left of this point; the other 27 are to the right. The correlation for the three improvisations to the left of the apex was ·38 (*ns*, $n = 3$); for the 27 to the right it was –·64 ($p <$ ·05, $n = 27$). In short, the jazz data do not clearly differentiate between an inverted-U relationship and a negative linear relationship.

The order effect found in Experiment 1 did not reappear in Experiment 2. That effect might have been a function of presentation format (blocked *versus* interleaved). However, the fact that it appeared only for bluegrass but not for jazz in Experiment 1 and for neither style in Experiment 2 suggest that it might have been due to sample variation.

General Discussion

The purpose of the work presented in this article was not to distinguish between the alternative theoretical explanations for the inverted-U relationship (*e.g.*, Berlyne, 1971 *v.* Martindale, 1988) but to address the question of whether this relationship is as general and robust as these explanations presuppose.

We found strong support for the inverted-U relationship between liking and complexity for bluegrass. The two criteria that were outlined in the introduction were fulfilled in both Experiments 1 and 2. First, the quadratic equation explained reliably more variance compared to the linear equation. Second, both the ascending and descending legs of the curve were supported by (*a*) the number of improvisations to the left and to the right of the apex, and (*b*) the strength and reliability of the correlations between the data points on both sides of the apex. The correlation for the data to the right of the apex was reliable in both experiments, while the correlation for the data to the left of the apex was reliable for Experiment 2 but not for Experiment 1.

By contrast, we found only weak support for the inverted-U relationship for jazz. Of the two criteria, only one was satisfied. The quadratic equation explained reliably more variance than the linear equation. However, the support for the left, ascending leg of the inverted-U relationship was minimal. In Experiment 1 there were only two data points to the left of the apex; in Experiment 2, only three. Moreover, the linear correlations to the left of the apex were not reliable. In short, we found strong support for the inverted-U relationship for bluegrass but a negative linear relationship cannot be ruled out for jazz.

In principle, the inverted-U and the negative linear interpretations of our data are empirically distinguishable. All of the stimuli in this experiment were pieces of music, created by professional musicians who knew that they were being recorded. Even when instructed to produce something very simple, they nevertheless played music. They did not, for example, play the same tone over and over again. The full range of possible complexity levels of auditory stimuli thus extend below the level of the least complex stimuli used in this experiment. This is clearly visible in the data. The lowest mean perceived complexity of the jazz improvisations was 2·20 in Experiment 1 and 2·42 in Experiment 2 as compared to 1·70 in Experiment 1 and 1·86 in Experiment 2 for the bluegrass improvisations.

If we would have included simpler stimuli among the jazz improvisations, we might have found more support for the left, ascending leg of the inverted-U relationship. The lower bound on the complexity for the jazz improvisations might have cut off the left leg of the inverted-U relationship, leaving us with a close approximation to a negative linear relationship; this might be what we see in Figures 3 and 8. This implies that extending the range of complexity below the lowest level included in our experiments would produce stimuli that would be judged as less likable than the least liked stimuli in the present set. The negative linear interpretation, on the other hand, implies that extending the range of complexity towards the simple end would produce stimuli that are more likeable than the most liked stimuli in the present set.

However, this way of resolving the issue would encounter the difficulty that simpler stimuli might not be perceived as music. Figure 10 provides an example of a very simple jazz improvisation used in Experiments 1 and 2. It is not clear how music could be any simpler and still be music. Hence, the lower bound on complexity observed in our experiment might reflect the nature of music rather than a methodological difficulty.

There are two additional observations relevant to the range of complexity that favour the negative linear over the inverted-U interpretation of the jazz data. First, Steck and Machotka's (1975) research suggested that the inverted-U relationship between complexity and liking might not depend upon the range of complexity

FIG. 10
A very simple jazz improvisation used in Experiments 1 and 2.

that is sampled. They found that the inverted-U relationship between complexity and liking was independent of the range of complexity of artificially generated tone sequences. Second, our data also suggest that the range of complexity sampled is not important. The range of perceived complexity for the bluegrass improvisations was about one-half a point less than for the jazz improvisations for Experiments 1 and 2 (0·51 and 0·68, respectively). The restriction of range argument thus implies that there ought to be less support for the inverted-U relationship for bluegrass than for jazz. However, we found the opposite.

These three observations – that the complexity range cannot, in fact, be extended towards the simple end without losing ecological validity, that other studies have found that the shape of the inverted-U relation is insensitive to the range of complexity, and that our data show a stronger inverted-U relation for the style that exhibited the narrower complexity range – suggest that the relationship between liking and complexity is a negative linear relationship for jazz, *i.e.,* that listeners prefer simplicity in jazz. Taken in conjunction with the strong empirical support for the inverted-U relation in bluegrass, this implies that the relationship between complexity and liking has different mathematical shapes for different musical styles.

If future studies replicate this difference in the complexity–liking relationship across musical styles, we have to inquire into the causes of this difference. Russell (1982) suggested that the negative linear relationship he found between liking and complexity might have been due to the participants' lack of familiarity with modern jazz. This idea is supported by Smith and Melara (1990) who also found a negative relationship between liking and complexity. Their novices preferred a simple I-V-I-IV-V-vi-IV-V-I progression to more complex progressions; *e.g.,* I-iii-III-flatII-III-VI-flatII-V-I, as the most complex. In their study, rated unusualness increased reliably with complexity. These two studies suggest that the relationship between liking and complexity might be moderated by familiarity with the musical style. The more familiar the music style, the closer the fit to an inverted-U function; the less familiar, the closer the fit to a negative linear relationship. Applied to our data, this explanation presupposes that bluegrass was more familiar to our participants than jazz. Our data do not allow us to evaluate this presupposition. Empirical evaluation requires a measure of familiarity such as subjective estimates of exposure to a particular musical style (North and Hargreaves, 1995).

Another possible reason why the complexity–liking relationship might vary with musical style is that complexity is confounded with prototypicality. Martindale (1984; 1988) suggested that liking is a function of prototypicality, not complexity, and that the relationship is positive linear. Martindale and Moore (1989) further suggested that in some studies the observed inverted-U relationship was due to the fact that those stimuli that were either very simple or very complex were also less prototypical than those of moderate complexity. Applied to the observed differences between jazz and bluegrass in our data, this principle implies that the simple jazz pieces and the moderately complex bluegrass pieces were the most prototypical. Empirical resolution of this question requires an operational definition of prototypicality. This might be difficult to achieve with respect to real music (but see Hekkert and van Wieringen, 1990).

Finally, specific sonic dimensions might underlie why the complexity–liking relationship varies with musical style. One possible dimension is the distance between successive or simultaneous pitches (*e.g.*, in terms of intervallic distance). Another possible dimension is the degree of variance in rhythmic phrasing (*i.e.*, how much does a corpus of phrases in an improvisation vary from one another). Understanding how the differences among musical styles along such dimensions is related to the complexity–liking relationship would require strict experimental or statistical control of such dimensions. However, the interpretation of any results using this method would be complicated by the fact that other factors – such as familiarity and prototypicality – might be correlated with these dimensions. In other words, the pattern of sonic dimensional values for any type of music might determine its familiarity or prototypicality. Barring this difficulty, such results would allow for a concrete explanation of why the complexity–liking relationship varies across musical styles. We will refrain from speculating about how differences along these and other dimensions explain why the inverted-U relationship between complexity and liking exists for our bluegrass improvisations but not for our jazz improvisations.

Another type of sonic dimension that was indisputably different across the improvisations used in our studies was musical instrument timbre. The bluegrass improvisations consisted of the guitar and mandolin; the jazz improvisations consisted of the soprano saxophone, flute and Hammond B-3 organ. An important consideration is whether our present results are affected by this dimension. In other words, to what degree does instrument type affect the shape of the complexity–liking relationship? Further research employing only one instrument across the musical styles would control for this dimension. However, this method would reduce ecological validity in some cases because musical styles naturally differ concerning their respective instrumentation. As an extreme case imagine a comparison between rock music (*e.g.*, Jimi Hendrix) and Gregorian chants when using only a distorted electric guitar as the instrument. Fortunately, many instruments are shared across musical styles.

In summary, although the inverted-U relationship between complexity and liking has empirical support both in the literature and in our own bluegrass data, it might not be general and robust. Our results suggest that the relationship might vary with musical style. If this finding is replicated in future studies, it throws doubt on any theoretical explanation that derives the inverted-U relationship from general psychological mechanisms. Any general explanation would need to be augmented with some auxiliary principle to explain variation across styles.

References

Berlyne, D. E. (1971). *Aesthetics and psychobiology.* New York: Appleton–Century–Crofts.

Crozier, J. B. (1974). Verbal and exploratory responses to sound sequences varying in uncertainty level. In: D. E. Berlyne (Ed.), *Studies in the new experimental aesthetics,* pp. 27–90. Washington, D.C.: Hemisphere.

Cupchik, G. C. and Winston, A. S. (1996). Confluence and divergence in empirical aesthetics, philosophy, and mainstream psychology. In: M. P. Friedman and E. C. Carterette (Eds.), *Cognitive Ecology,* pp. 61–85. New York: Academic Press.

Davies, J. B. (1978). *The Psychology of Music.* London: Hutchinson.

Hargreaves, D. J. (1984). The effects of repetition on liking for music. *Journal of Research in Music Education,* **32(1),** 35–47.

Hekkert, P. and van Wieringen, P. C. W. (1990). Complexity and prototypicality as determinants of the appraisal of cubist paintings. *British Journal of Psychology,* **81,** 483–495.

Heyduk, R. G. (1975). Rated preference for musical composition as it relates to complexity and exposure frequency. *Perception and Psychophysics,* **17,** 84–91.

Konecni, V. J. (1982). Social interaction and musical preference. In: D. Deutsch (Ed.), *The Psychology of Music.* New York: Academic Press.

Martindale, C. (1984). The pleasures of thought: A theory of cognitive hedonics. *Journal of Mind and Behavior,* **5,** 49–80.

Martindale, C. (1988). Cognition, psychobiology, and aesthetics. In: F. Farley and R. Neperud (Eds.), *The foundation of aesthetics, art, and art education,* pp. 7–42. New York: Praeger.

Martindale, C. and Moore, K. (1989). Relationship of musical preference to collative, ecological, and psychophysical variables. *Music Perception,* **6(4),** 431–445.

North, A. C. and Hargreaves, D. J. (1995). Subjective complexity, familiarity, and liking for popular music. *Psychomusicology,* **14,** 77–93.

North, A. C. and Hargreaves, D. J. (1996a). Responses to music in a dining area. *Journal of Applied Social Psychology,* **26(6),** 491–501.

North, A. C. and Hargreaves, D. J. (1996b). Responses to music in aerobic exercise and yogic relaxation classes. *British Journal of Psychology,* **87,** 535–547.

North, A. C. and Hargreaves, D. J. (1997). Liking for musical styles. *Musicae Scientiae,* **1(1).**

Orr, M. and Ohlsson, S. (1999). *The inverted-U hypothesis: Can it predict expert behavior?* Paper presented at the Society for Music Perception and Cognition Bi-Annual Conference, Evanston, Illinois, U.S.A.

Russell, P. A. (1982). Relationships between judgements of the complexity, pleasingness and interestingness of music. *Current Psychological Research,* **2,** 195–202.

Simon, C. R. and Wohlwill, J. F. (1968). An experimental study of the role of expectation and variation in music. *Journal of Research in Music Eduaction,* **16,** 227–238.

Sluckin, W., Hargreaves, D. J. and Colman, A. M. (1983). Novelty and human aesthetic preferences. In: J. Archer and L. Birke (Eds.), *Exploration in animals and humans.* London: Van Nostrand.

Smith, J. D. and Melara, R. J. (1990). Aesthetic preference and syntactic prototypicality in music: 'Tis the gift to be simple. *Cognition,* **34(3),** 279–298.

Steck, L. and Machotka, P. (1975). Preference for musical complexity: Effect of context. *Journal of Experimental Psychology,* **54,** 54–61.

Stimson, J. A., Carmines, E. G. and Zeller, R. A. (1981). Interpreting Polynomial Regression. In: P. V. Marsden (Ed.), *Linear models in social research,* pp. 88–96. Sage.

Vitz, P. C. (1966). Affect as a function of stimulus variation. *Journal of Experimental Psychology,* **71(1),** 74–79.

Walker, E. L. (1980). *Psychological complexity and preference: A hedgehog theory of behavior.* Monterey, California: Brooks/Cole.

Zillmann, D. and Gan, S. (1997). Musical taste in adolescence. In: D. J. Hargreaves and A. C. North (Eds.), *The social psychology of music,* pp. 161–187). Oxford: Oxford University Press.